

SESSÃO REVISÃO DE LITERATURA

METODOLOGIAS DE AVALIAÇÃO DE POLÍTICAS PÚBLICAS: ANÁLISE E APLICAÇÃO DOS PRINCIPAIS MÉTODOS NO CONTEXTO DO BANCO DO NORDESTE

Public policy evaluation methodologies: analysis and application of the main methods in the Banco do Nordeste context

Jaqueline Nogueira Cambota

Economista. Doutorado em Economia pela Universidade de São Paulo (2012). Coordenadora de Estudos e Pesquisas do Escritório Técnico de Estudos Econômicos do Nordeste - ETENE. Banco do Nordeste do Brasil S/A. jacquelinenc@bnb.gov.br

Diego Rafael Fonseca Carneiro

Economista. Doutorado em Economia pela Universidade Federal do Ceará (2018). Bolsista de Inovação SET-CNPq/ETENE. Banco do Nordeste do Brasil S/A. dr.carn@gmail.com

Resumo: O objetivo deste artigo é apresentar os principais parâmetros de interesse e métodos de avaliação em programas sociais. Também se resgata algumas das principais aplicações desses métodos aos Programas operacionalizados pelo Banco do Nordeste. A estimativa de impacto de um Programa dependerá da forma que ocorreu a designação ao tratamento. Quando a designação não acontece de forma aleatória, os métodos de avaliação baseiam-se na seleção de observáveis e não observáveis para construção de contrafactual válido. As evidências empíricas encontradas nas avaliações dos Programas operacionalizados pelo BNB reforçam que uma das maiores dificuldades na estimação do impacto de um Programa está na construção de um grupo comparação válido. As avaliações, tanto internas quanto externas, desses Programas têm abordado essa limitação, bem como buscado forma de contorná-la por meio dessas metodologias.

Palavras-chave: Avaliação; Políticas Públicas; Banco do Nordeste.

Abstract: The aim of this article is to present the main parameters of interest and evaluation methods in social programs. It also rescues some of the main applications of these methods to the Programs operated by Banco do Nordeste. The estimated impact of a Program will depend on how the treatment designation happened. When the designation does not occur randomly, the assessment methods are based on the selection of observable and unobservable for valid counterfactual construction. The empirical evidence found in the evaluations of the Programs operationalized by the BNB reinforces that one of the greatest difficulties in estimating the impact of a Program is in the construction of a valid comparison group. Evaluations, both internal and external, of these Programs have addressed this limitation, as well as sought ways to circumvent it through these methodologies.

Keywords: Evaluation; Public Police; Banco do Nordeste.

1 INTRODUÇÃO

Os métodos econométricos de avaliação de impacto baseiam-se na ideia de resultados potenciais. Essa abordagem preconiza que para se obter o efeito causal de determinado tratamento, a princípio, seria necessário comparar o mesmo indivíduo nas situações em que ele recebe e não recebe o tratamento. Contudo, apenas uma dessas situações é observável de cada vez. Desse modo, todo o esforço nos processos de avaliação se dá pela obtenção de um grupo de comparação válido, ou seja, que simule o contrafactual das unidades tratadas.

A construção do grupo de comparação é cercada de potenciais problemas, mas que podem ser circunscritos ao chamado viés de seleção. Esse decorre fundamentalmente da forma como o tratamento foi distribuído, ou de outra maneira, como foram selecionados para participar do programa. Um caso bastante comum, principalmente, nos programas de crédito, é a auto-seleção¹, em que as unidades escolhem participar ou não de acordo com análise custo-benefício do recebimento do tratamento. A auto-seleção dificulta, portanto, a construção de um grupo de comparação adequado para estimação do impacto do tratamento, uma vez que os indivíduos que podem tirar maior proveito do tratamento são também aqueles com maior interesse na participação.

Os métodos para lidar com esse problema podem ser divididos em duas abordagens: experimental e não experimental. O método experimental consiste na aleatorização da distribuição do tratamento². Quando bem conduzida, a aleatorização proporciona um grupo de comparação adequado, com as mesmas características do grupo de tratamento, permitindo estimar consistentemente o efeito do programa. A aleatorização do tratamento, apesar de ser considerado o “padrão ouro” das avaliações de impacto, é pouco frequente por exigir uma preparação *ex ante* a implementação da política, o que a torna mais complexa e dispendiosa, além das questões éticas e/ou políticas inerentes a distribuição do tratamento por critérios fortuitos.

1 Para maiores detalhes ver Roy (1956)

2 Nos casos em que a aleatorização não foi bem realizada ou quando o grupo de comparação não é um contrafactual válido, os métodos baseados em seleção sob observáveis ou não observáveis fornecem estimativas mais confiáveis do que as obtidas por experimento natural. Isso porque a avaliação experimental só pode ser realizada em programas em que o tratamento foi aleatorizado.

Desse modo, quando não é possível a aleatorização do tratamento, os métodos não experimentais oferecem alternativa por meio do controle das características observadas ou não dos indivíduos que afetam a decisão de participação do tratamento. O problema desses métodos é que muitas vezes se baseiam em suposições fortes acerca da forma como o grupo de tratamento foi selecionado, a exemplo disso, alguns deles assumem que a participação do indivíduo depende exclusivamente de suas características observadas e independem de seu resultado potencial, a chamada ignorabilidade (ROSENBAUM; RUBIN, 1983). Portanto, “um avaliador deve escolher um estimador de maneira cuidadosa considerando a natureza dos dados avaliados, a natureza institucional do programa e o parâmetro de interesse” (SMITH, 2004, p. 9).

Diante disso, o objetivo deste artigo é apresentar os principais parâmetros de interesse das avaliações, bem como discutir os métodos de avaliação utilizados em suas estimações. Também são recuperadas as principais experiências de avaliação aplicadas aos Programas operacionalizados pelo BNB.

Além desta introdução, o artigo está dividido em mais cinco seções. Na segunda seção, é feita a descrição dos quatro principais parâmetros de interesse em avaliação de programas. Na terceira seção, são discutidos os principais métodos de estimação dos efeitos dos tratamentos, expondo suas aplicações e limitações. Na quarta, também são discutidas, na medida do possível, aplicações desses métodos nas avaliações dos Programas operacionalizados pelo BNB. Por fim, são realizadas as considerações finais.

2 RESULTADOS POTENCIAIS E OS PARÂMETROS DE INTERESSE NA AVALIAÇÃO DE PROGRAMAS

A definição de impacto de um programa pode tomar diferentes conotações a depender do desenho da política e do método aplicado em sua avaliação. Contudo, o ponto de partida para todas as abordagens é o arcabouço de resultados potenciais, em que cada unidade i é observada em apenas um estado, tratado (Y_{i1}) ou não tratado (Y_{i0}), mas o par (Y_{i1}, Y_{i0}) nunca é observado simultaneamente para uma mesma unidade, o que é definido por Holland (1986) como o problema fundamental da inferência causal. Portanto, como Y_{i0} não é observado, o avaliador precisa encontrar

um grupo contrafactual válido, o qual deve possuir características indistinguíveis do grupo dos tratados, diferindo apenas por não serem expostos ao tratamento. Dessa forma, Heckman et al. (2001) definiram os principais parâmetros de interesse como ATE, ATT³, MATE e LATE, que serão detalhados nas próximas seções.

2.1 O Efeito Médio do Tratamento (ATE)

Esse parâmetro mede o efeito do tratamento sobre todas as unidades em uma população, incluindo participantes e não-participantes do programa. Sua obtenção se dá a partir da comparação do resultado médio entre os dois status de tratamento, o que empiricamente se traduz na diferença média entre tratados e não tratados. De maneira formal é possível definir o ATE como:

$$ATE = E(Y_{i1} - Y_{i0}) = E(Y_{i1}) - E(Y_{i0}) = \frac{1}{N} \sum (y_i(1) - y_i(0)) \quad (1)$$

A Equação 1 pressupõe a uniformidade das características, observáveis e não observáveis, entre tratados e não tratados, ou seja, ausência do viés de seleção. Mas quando essa condição não é razoável, em função do desenho do programa, suposições adicionais fazem-se necessárias para a obtenção do ATE. Nesses casos, a hipótese de identificação mais frequente é a de ignorabilidade do tratamento, ou seja, (Y_{i1}, Y_{i0}) são independentes do tratamento quando condicionados às covariadas relevantes.

Em outras palavras, assume-se que a escolha dos indivíduos tratados baseou-se apenas nas características observáveis, e não no seu resultado

$$E(Y_{i1}|d=1) - E(Y_{i0}|d=0) = [E(Y_{i1}|d=1) - E(Y_{i0}|d=1)] + [E(Y_{i0}|d=1) - E(Y_{i0}|d=0)] = ATT + Viés de seleção \quad (4)$$

Assim, na ausência de viés de seleção, é possível estimar o ATT pela simples diferença dos resultados médios dos grupos de tratamento e controle.

Alternativamente, quando não é possível afastar a possibilidade de ausência de viés, mais uma vez recorre-se a hipótese de ignorabilidade para obter um estimador condicionado as características observáveis na forma:

$$ATT(x) = E(Y_{i1} | x, d=1) - E(Y_{i0} | x, d=0) \quad (5)$$

³ Alguns avaliadores podem ainda estar interessados no efeito do tratamento sobre os não tratados, $ATTNT = E(Y_{i1}|d=0) - E(Y_{i1}|d=1)$.

potencial, ou seja, o tratamento não foi atribuído aos que mais se beneficiariam dele, de modo que o diferencial de resultados entre os grupos se deve, além do próprio tratamento, as diferenças nessas características. Apesar de ser uma hipótese forte, a maior parte dos métodos de avaliação depende de sua validade.

Assim, sob ignorabilidade consegue-se estimar consistentemente o efeito médio do tratamento, condicional as características observáveis, por:

$$ATE(x) = E(Y_{i1} - Y_{i0} | x) = E(Y_{i1} | x) - E(Y_{i0} | x) \quad (2)$$

2.2 O Efeito Médio do Tratamento sobre os Tratados (ATT)

Diferente do ATE, que mede o efeito sobre toda a população, o ATT mede o efeito do programa apenas sobre as unidades que efetivamente participaram do tratamento. Conceitualmente, o ATT pode ser entendido como:

$$ATT = E(Y_{i1} - Y_{i0} | d=1) = E(Y_{i1} | d=1) - E(Y_{i0} | d=1) \quad (3)$$

Onde $d=1$ quando há participação no programa, $d=0$ caso contrário.

Mas, como já discutido, não é possível visualizar o resultado dos indivíduos tratados na ausência do tratamento, de modo que o termo $E(Y_{i0} | d=1)$ não pode ser observado. Assim, desde que exista um grupo de comparação fiel ao contrafactual das unidades tratadas, é possível utilizar o seu resultado médio, $E(Y_{i0} | d=0)$ para estimar o ATT. Mas note que:

2.3 O Efeito Médio do Tratamento Local (LATE)

Conforme Imbens e Angrist (1994), o LATE pode ser entendido como efeito médio do tratamento para as unidades, cujo status do tratamento é influenciado por mudanças em uma variável instrumental, Z , satisfazendo uma restrição de exclusão. Nesse caso, a VI determina, em parte, a participação no tratamento, mas não é correlacionada com o seu resultado potencial, de modo que é possível utilizá-la para aferir o efeito médio do tratamento. Contudo, esse efeito médio refere-se apenas ao subconjunto

dos indivíduos cujo comportamento foi alterado em função de Z, por isso o efeito é local, sendo, via de regra, não expansível a população como um todo.

Uma forma mais simples de compreender o problema é tomando o exemplo de uma variável instrumental binária que quando assume valor igual a um aumenta a chance de participação no programa (monotonicidade). Assim, podem-se delimitar três casos: (i) os *always-takers*, que participariam do programa independentemente da variável instrumental; (ii) os *never-takers*, que nunca participariam do programa; e, finalmente, (iii) os *compliers*, que seguem exatamente a orientação da variável Z. Esse último

grupo será aquele cujo resultado será afetado, e consequentemente integrará o LATE.

Quadro 1 – Ilustração do caso com Z binário.

	Z=0	Z=1
Participou do programa	<i>complier/never-taker</i>	<i>never-taker</i>
Não participou do programa	<i>always-taker</i>	<i>complier/always-taker</i>

Fonte: Elaborado pelos autores.

Adicionando a hipótese de restrição de exclusão, de que o resultado depende apenas do status do tratamento, o efeito médio do tratamento para o grupo dos *compliers* é dado por:

$$LATE(Z) = E(Y_i|Z_i=1, T_i=1) - E(Y_i|Z_i=0, T_i=0) = \frac{E(Y_i|Z_i=1) - E(Y_i|Z_i=0)}{P(T_i=1|Z_i=1) - P(T_i=1|Z_i=0)} \quad (6)$$

2.4 O Efeito Marginal do Tratamento (MTE)

Se o custo de participação for definido como $C=Z\theta+U_D$, em que Z é a variável instrumental definida anteriormente e U_D é o componente não observado deste custo, o MTE avalia o efeito do tratamento para indivíduos, cujo componente não observado do custo de participação, U_D , é igual a um determinado valor.

$$MTE(x) = E(Y_{i1}-Y_{i0}|x, U_d = u_d) \quad (7)$$

$$E(Y_{i1}|x, U_d = u_d) - E(Y_{i0}|x, U_d = u_d)$$

A avaliação do MTE para um valor baixo de u_d mensura o ganho no resultado para indivíduos, cujo fator não observado os tornam menos prováveis a participar, enquanto que a avaliação do MTE para valores altos de u_d é o ganho para os indivíduos, cujo fator não observado os tornam mais prováveis a participar. O MTE também pode ser expresso em termos do LATE, sendo o efeito marginal do tratamento avaliado em termos de ganhos dos indivíduos que estão indiferentes entre participar ou não do tratamento. Esse caso ocorre quando o custo de participação é igual a zero, ou seja, $U_d = (-Z\theta)$.

$$MTE(x) = E(Y_{i1}-Y_{i0}|x, U_d = -Z\theta) \quad (8)$$

Heckman et al. (2001) mostraram que sob independência entre U_D e $U_1 - U_0$,⁴ todos os parâme-

tros são iguais, portanto, se a $Cov(U_1 - U_0, U_D)=0$ todos os quatro parâmetros se reduzem ao ATE. Neste caso, não existe nada não observado pelo avaliador que influencie o indivíduo na decisão de participação do tratamento. Por outro lado, se a $Cov(U_1 - U_0, U_D)>0$, o ATT será maior que o ATE. Portanto, as pessoas se selecionam para o tratamento com base em seu ganho idiossincrático, sendo o ganho dos indivíduos observados no estado de tratamento maior do que de um indivíduo médio.

3 MÉTODOS DE AVALIAÇÃO

3.1 Experimentos Aleatórios⁵

A necessidade de aleatorização surge do problema da inferência causal, uma vez que, em muitos casos, torna-se difícil estimar o impacto do tratamento sobre um grupo de indivíduos, dado que outros fatores que afetam o resultado podem ter sofrido mudança desde a introdução do tratamento. Dessa forma, o resultado pode, em parte, ser decorrente da mudança nessas variáveis e, não devido ao tratamento.

A aleatorização também é justificável em programas já existentes como, por exemplo, quando a demanda pelo tratamento excede a oferta ou quando o tratamento é dado de forma progressiva. Nesse último caso, o grupo comparação pode ser afetado por sua participação futura no tratamento

4 Termo do erro aleatório para as equações de resultados potenciais, definidas na seção 3.2.1.

5 Duflo et al. (2006) forneceram um guia prático para avaliação de experimentos aleatórios.

deixando, portanto, de ser um contrafactual válido. Outros exemplos em que, mesmo com a aleatorização, o grupo comparação pode não ser um contrafactual válido surgem quando há *compliance* imperfeito (parcial), externalidades ou atrito.

Se apenas parte das unidades do grupo de tratamento é tratada ou, inversamente, quando algumas unidades do grupo comparação recebem o tratamento, se diz que o *compliance* é imperfeito. Uma das explicações para existência de *compliance* imperfeito é a dificuldade em controlar as escolhas do grupo comparação.

As externalidades surgem quando os resultados das unidades não tratadas são afetados pelo tratamento. A literatura tem encontrado diversos efeitos *spillovers* como efeitos físicos, mudanças de preços ou na forma de aprendizagem e efeitos imitação. O atrito se refere à perda de dados que fizeram parte da amostra original. Se o atrito for aleatório ele somente reduzirá o poder do teste. Entretanto, se ele estiver relacionado ao tratamento pode resultar em estimativas viesadas. Ou seja, se o atrito for maior entre as unidades beneficiadas pelo programa e o pesquisador ignorar essa informação, as estimativas do efeito do programa podem ser sobreestimadas.

O estimador de mínimos quadrados ordinários providencia uma estimativa não viesada do impacto do programa sobre a população de tratados, quando a avaliação aleatória foi corretamente desenhada e implantada. Neste caso, o ATE pode ser obtido por uma regressão de mínimos quadrados ordinários

$$Y_i = \alpha_i + \beta_d d_i + U_i \quad (9)$$

Em que d_i é a uma *dummy* para participação no grupo tratamento, e $\beta_{ols} = \hat{E}(Y_{i1} | d=1) - \hat{E}(Y_{i0} | d=0)$ mensura o efeito do tratamento.

O controle sob covariadas que, provavelmente, alterem o resultado pode não afetar o valor esperado do estimador β_{ols} , mas reduzem sua variância. Assim, o controle por variáveis que possuem um extenso impacto sobre os resultados podem ajudar a reduzir o erro-padrão das estimativas. Isso é especialmente útil quando se está trabalhando com amostras pequenas. Neste caso, deve-se considerar a equação:

$$Y_{ij} = \alpha_i + \beta_d d_i + X_{ij} \gamma + \tilde{v}_j + \tilde{w}_{ij} \quad (10)$$

Em que: X_{ij} é um conjunto de variáveis de controle e, \tilde{v}_j e \tilde{w}_{ij} representam a variância não explicada após controlar por X_{ij} .

Por sua vez, se uma determinada covariada exerce maior influência sobre a variável resultado, a estratificação da amostra dentro de grupos que possuem valores similares de certas características observadas pode aumentar a precisão das estimativas. Assim, a aleatorização dentro dos estratos pode ser usada tanto quando se deseja reduzir a variância, como quando se deseja estudar o efeito tratamento sobre um subgrupo específico. Uma estimativa do efeito tratamento é obtida ao estimar:

$$Y_{ij} = \alpha_i + \beta_d d_i + E_{ij} + \tilde{v}_j + \tilde{w}_{ij} \quad (11)$$

Por OLS, em E_{ij} é um conjunto de variáveis *dummies* indicando o estrato de cada observação.

A avaliação experimental, entretanto, é difícil de ser implantada. Muitas vezes requer parcerias com os responsáveis pela implantação do programa, estando a aleatorização inserida dentro de um projeto piloto. No entanto, avaliação com parceria governamental é rara, pois requer um alto nível de cooperação política, o que faz muitos avaliadores buscarem parcerias com ONGs e empresas privadas para sua realização.

3.2 Métodos Não experimentais

Se o tratamento não foi aleatório, as unidades não podem ser diretamente comparadas, pois aquelas expostas ao tratamento diferem em algumas características das que não foram expostas, o que poderia afetar a variável resultado. Com base nessas diferenças alguns métodos fornecem estimativas de impacto de programas baseadas em hipóteses envolvendo características observáveis ou não.

3.2.1 Seleção sob Observáveis

Os métodos de seleção sob observáveis baseiam-se na hipótese de que, dado o vetor de características X , os resultados potenciais são independentes do tratamento, ou seja, $(Y_1, Y_0) \perp V | X$. Portanto, a probabilidade de ser tratado dependeria somente de X , podendo ser expressa com uma função probabilidade condicional $p(X)$. Neste sentido, admite-se a existência do viés de seleção, mas o avaliador observa todas as características relevantes para a seleção.

Métodos de Regressão

Como discutido anteriormente, o efeito médio do tratamento, condicional sob as covariadas relevantes, é dado por:

$$ATE(x) = Y_i = d_i Y_{i1} + (1 - d_i) Y_{i0} \quad (12)$$

Supondo que as equações de resultados potenciais tenham a forma linear e a seguinte especificação:

$$\begin{aligned} Y_{i1} &= \alpha_1 + \delta d_i + X_i \beta_1 + U_{i1}, \text{ se } d_i > 0 \\ Y_{i0} &= \alpha_0 + \delta d_i + X_i \beta_0 + U_{i0}, \text{ se } d_i \leq 0 \end{aligned} \quad (13)$$

Desse modo, a formação do grupo comparação requer uma média de unidades desconhecidas, que sob ignorabilidade pode ser obtida pela média das observações do grupo controle que possuam X iguais aos do grupo tratamento, portanto, a equação para a variável resultado observada poderá então ser escrita na forma:

$$Y_i = \alpha_0 + \delta d_i + X_i \beta_1 + \varepsilon_i \quad (14)$$

que sob ignorabilidade estima o efeito tratamento por mínimos quadrados ordinários:

$$E(Y_i \mid d_i, X_i) = \alpha_0 + \delta d_i + X_i \beta_1 \quad (15)$$

Método de Pareamento (Matching)

No *matching*, o grupo comparação é construído por meio de um pareamento, ou seja, busca-se formar pares de comparação formados entre cada indivíduo do grupo tratamento. Para cada indivíduo i , se $Y_{i1}(1)$ é observado ($T_i=1$), então $Y_{i1}(0)$ é um *missing* ($T_i=0$). Desse modo, buscam-se indivíduos j , tais que $T_j=0$, e que satisfaçam:

$$\|X_i - X_j\| \leq \|X_i - X_i\|, \text{ tal que } T_j=0 \quad (16)$$

Em geral, $\|X_i - X_j\|$ será diferente de zero tornando as estimativas do ATE e ATT viesadas assintoticamente em decorrência da imperfeição do *matching*. O estimador de *matching* também não é eficiente, uma vez que não utiliza toda a amostra, mas apenas aquelas unidades com o X mais próximo do grupo tratamento. Existe também *trade-off* entre viés e variância nesse estimador, visto que o aumento no número de observações a servir de par para uma mesma unidade observada aumenta o viés, pois aumenta a distância, enquanto diminui a variância por utilizar mais informações.

O *matching* também pode ser visto como um caso particular da regressão, que quando estimado pelo vizinho mais próximo e número de vizinhos fixo, é algebricamente idêntico à regressão.

No entanto, ao contrário do método de regressão o *matching* não é eficiente, pois utiliza apenas uma parte da amostra.

Pareamento por Escores de Propensão (Propensity Score Matching)

O problema do enfoque de regressão é que ele impõe uma suposição de normalidade sobre os termos do erro, $E(U_{i1} \mid X)$ e $E(U_{i0} \mid X)$, que são termos desconhecidos. No *matching*, o problema é a condicionalidade, que corresponde ao agrupamento das observações sobre as covariadas relevantes. Em casos de covariadas contínuas, esse agrupamento pode tornar-se impossível.

Alternativamente, o *propensity scores* se baseia na suposição de ignorabilidade do tratamento para propor um estimador para o ATE(x) que não impõe a restrição da forma funcional, e que também reduz a dimensionalidade dos condicionantes. Rosenbaum e Rubin (1983) a partir da suposição de ignorabilidade ofereceram uma solução potencial para os problemas dos outros enfoques. Seja $p(X_i)$ a probabilidade da unidade i ter sido destinada ao tratamento, definido como:

$$p(X_i) = P(T_i = 1 \mid X_i) = E(T_i \mid X_i), \text{ em que}$$

$$0 < p(X_i) < 1, \text{ portanto } (Y_{i1}, Y_{i0}) T_i / X_i \quad (17)$$

$$\text{implica } (Y_{i1}, Y_{i0}) T_i / p(X_i)$$

De acordo com Rosenbaum e Rubin (1983), se a participação no tratamento é fortemente ignorada, dado o vetor X, então as diferenças entre as médias dos tratados e comparação, naquele escore de balanceamento, é uma estimativa confiável do efeito do tratamento para aquele valor, conseqüentemente, o *propensity scores* produz estimativas não viesadas do efeito médio do tratamento. Portanto, a probabilidade de participar do tratamento condicional as covariadas pode ser denotada por:

$$e(x) = pr(d=1 \mid x)$$

$$pr(d_1 \dots d_n \mid X_1 \dots X_n) = \quad (18)$$

$$\prod_{i=1}^n e(X_i)^{d_i} \{1 - e(X_i)\}^{1-d_i}$$

Neste caso, a função *propensity scores*, $e(x)$, é sempre desconhecida e não existe nenhuma especificação previamente aceita, podendo $e(x)$ ser estimada por logit.

Propensity Score Generalizado

O método de pareamento por escore de propensão assume que o tratamento é atribuído de maneira uniforme entre os participantes do programa, contudo, existem situações em que cada unidade recebe diferentes “dosagens” do tratamento. Nesse intuito, Imbens (2000) e Hirano e Imbens (2004) propuseram uma generalização do PSM, que considerasse os diferentes níveis de tratamento.

Sobre a hipótese de ignorabilidade, assume que $X \perp\!\!\!\perp \{T=t\} \mid Vr(t, X)$, em que r é o escore de propensão generalizado, e ainda que $TVXN(\beta_0 + \beta_1 X, \sigma^2)$. Assim, é possível estimar r em um primeiro estágio pelo método de máxima verossimilhança. No segundo estágio, por sua vez, modela-se a expectativa condicional de Y dado r e T , a partir de uma forma flexível, com a forma quadrática:

$$E(Y|r, T) = \alpha_0 + \alpha_1 T + \alpha_2 T^2 + \alpha_3 \hat{r} + \alpha_4 \hat{r}^2 + \alpha_5 T \times \hat{r} \quad (19)$$

Dessa forma, é possível obter o efeito do tratamento para cada nível t , assim como obter o efeito marginal do tratamento tomando $\theta(t) = E[Y(t+\varepsilon)] - E[Y(t)]$.

3.2.2 Seleção sob não Observáveis

Em alguns casos, a escolha de participar do tratamento é determinada por características não observáveis, o que torna o controle das observáveis insuficiente para eliminar o viés de seleção. Nesse sentido, os métodos de seleção sob não observáveis fornecem alternativas para formação de um grupo comparação.

Métodos Longitudinais

Uma forma simples de avaliação consiste em examinar os resultados das unidades tratadas antes e após a implantação do programa. A suposição implícita deste método é que as unidades tratadas deveriam apresentar os mesmos resultados, caso não tivessem sido submetidos ao tratamento. Assim sendo, os resultados antes da implantação do programa fornecem um bom grupo comparação. Desse modo, os métodos longitudinais estimam o efeito do tratamento sobre as unidades tratadas, ou seja, o ATT. O estimador antes-depois se baseia nesta suposição para formar um grupo comparação:

$$BA = E(Y_{1t}|d=1) - E(Y_{0t}|d=1) \quad (20)$$

No entanto, este estimador só será consistente quando não houver tendências temporais ou mudanças macroeconômicas agregadas, ou seja, somente se $E(Y_{0t}|d=1) = E(Y_{0t}|d=0)$, onde t representa o período após a implantação do programa. Na presença de tais fontes de inconsistências, o estimador de diferença em diferenças será mais robusto. Tal método avalia o efeito do tratamento pela comparação entre as mudanças antes e depois das unidades tratadas com as mudanças antes e depois das unidades não tratadas.

$$DD = [E(Y_{1t}|d=1) - E(Y_{1t}|d=0)] - [E(Y_{0t}|d=1) - E(Y_{0t}|d=0)] \quad (21)$$

Modelos de dados em painel providenciam uma forma geral para estes estimadores:

$$Y_{it} = \alpha_0 + \mu_i + \mu_t + \beta_d d_{it} + X_{it} \beta_1 + U_{it} \quad (22)$$

Sendo que β_d é o estimador de impacto do programa para dados em painel, μ_i é o termo específico da unidade i , μ_t é o termo tempo invariante, β_1 são os impactos das outras covariadas e ε_{it} é o termo do erro.

Variável Instrumental

Imbens e Angrist (1994), seguindo os avaliadores que criticam os métodos de avaliação baseados em regressão, propuseram um método para estimação dos efeitos causais em termos de resultados potenciais ou contrafactuais, em vez de parâmetros dos modelos de regressão. Os estimadores de variável instrumental possuem a vantagem de requererem suposições mais fracas que os métodos de regressão linear. Contudo, seus resultados não podem ser generalizados para a população, sendo restritos a indivíduos que estão no limiar para receber o tratamento, sendo diferenciados apenas pelo vetor de variável instrumental, Z .

Assim, os autores adotam uma estratégia de variável instrumental para identificar o efeito do tratamento. Essas estimativas serão consistentes desde que a variável instrumental, z_i , seja independente das variáveis de resultados potenciais Y_{1i} e Y_{0i} , e correlacionada com o indicador de participação d_i . Quando os instrumentos são binários, o LATE pode ser estimado por Wald:

$$VI = \frac{E(Y \mid Z=1) - E(Y \mid Z=0)}{Pr(d=1 \mid Z=1) - Pr(d=0 \mid Z=0)} \quad (23)$$

Modelos de variável latente, por sua vez, são frequentemente empregados para instrumentos contínuos ou múltiplos instrumentos em que a decisão de participação é modelada por:

$$d_i^* = \gamma_0 + Z_{iy1} + v_i \quad (24)$$

Desse modo, a equação de resultado é relacionada com o tratamento pela decisão de participação:

$$Y_i = \beta_0 + d_i \beta_1 + U_i \quad (25)$$

Em que:

$$d_i = \begin{cases} 1 & \text{se } d_i^* > 0 \\ 0 & \text{se } d_i^* \leq 0 \end{cases} \quad (26)$$

3.3 Desenhos de Regressão com Descontinuidade

Os desenhos de RD⁶, originalmente introduzidos por Thistlethwaite e Campbell (1960), são outro tipo de estimador de LATE, cuja estimação se dá pela comparação entre aqueles que estão logo abaixo e logo acima do ponto de corte para o recebimento do tratamento.

Existem dois tipos de desenho de RD, *sharp* e *fuzzy*. O desenho *sharp* é um caso de seleção sob observáveis, enquanto o *fuzzy* é um tipo de estimador de variável instrumental. No desenho *sharp*, a participação no tratamento depende da “variável de designação” de forma determinística. No desenho *fuzzy*, a dependência é probabilística e determinada por um *propensity scores*. Mesmo que a elegibilidade dependa de uma regra de ponto de corte, nem todos os elegíveis são tratados, pois a regra nem sempre é satisfeita.

A suposição básica para formação do grupo comparação é que os agentes são incapazes de controlar precisamente a “variável de designação” ao tratamento próximo ao ponto de corte. Como resultado dessa suposição, os desenhos de RD são frequentemente definidos como um quase-experimento⁷, já que o tratamento é considerado como localmente aleatório. Portanto, a hipótese de identificação de RD é baseada na suposição de que a

regra de participação ao tratamento é arbitrária, por isso proporciona um bom experimento.

Em programas que se caracterizam como desenhos RD, a participação do tratamento é definida por uma variável de designação, sendo o tratamento uma função contínua dessa variável. No entanto, a existência do tratamento como uma função contínua da “variável de designação” não é suficiente para aplicação de RD, uma vez que esse desenho é inválido, se indivíduos são capazes de manipular precisamente essa variável.

Não existe um consenso na literatura sobre a forma funcional desses desenhos, sendo utilizadas estimações paramétricas ou não paramétricas. Porém, ambas podem viesar as estimativas do efeito tratamento. A estimação paramétrica precisa da suposição de linearidade da forma funcional, no entanto, não existe nenhuma razão a priori para acreditar que a verdadeira forma funcional seja linear.

Uma forma de relaxar essa suposição é incluir funções polinomiais na “variável de designação” do modelo de regressão, o que corresponde ao enfoque de regressão em séries. A desvantagem deste enfoque é que ele providencia estimativas globais, enquanto que o desenho de RD depende de estimativas locais ao redor do ponto de corte. A regressão de *kernel*, que também é um estimador não paramétrico, é um método local para estimação da função de regressão em um ponto particular. Entretanto, regressão de *kernel* pode não possuir um bom desempenho em desenhos de RD, em que o ponto de corte representa um ponto de fronteira.

Por fim, pode-se afirmar que o mais importante na aplicação dos desenhos de RD é analisar a robustez das estimativas à inclusão de termos polinomiais de ordem superior, na regressão em série, ou a mudanças na largura da janela ao redor do ponto de corte, na regressão *kernel*.

3.4 Controle Sintético

O método do controle sintético, desenvolvido por Abadie e Gardeazabal (2003), utiliza como estratégia para a construção do grupo comparação, uma média ponderada das outras regiões que não receberam intervenções, a qual deverá ser comparada com a região que efetivamente recebeu intervenção. Essa média é denominada de grupo de controle “sintético”.

6 Lee e Lemieux (2010) discutiram exaustivamente aplicações de RDD em economia.

7 Os experimentos aleatórios, assim como os desenhos de RD, devem ser considerados como um processo gerador de dados e, não como um método.

De acordo com Abadie (2015), pode-se assumir sem perda de generalidade, que o primeiro país ($j=1$) é afetado pela intervenção política de interesse e que $j=2, \dots, J$ é uma coleção de regiões não afetadas pela intervenção. Assume-se também que o conjunto de dados compreende T períodos e que T_0 ocorre antes da intervenção. Para cada região, j , e tempo, t , assume-se um resultado de interesse y_{jt} . Também, assume-se para cada região j , um conjunto de k preditores do resultado: X_{1j}, \dots, X_{kj} (que podem incluir valores pré-intervenção de Y_{jt}). Para cada região afetada pela intervenção, $j=1$ e período pós-intervenção, $t > T_0$, em que os resultados potenciais devem ser observados com e sem intervenção, Y_{1t}^I e Y_{1t}^N , respectivamente. Desse modo, o efeito da intervenção para a região afetada no período t (com $t > T_0$) é dado por:

$$\tau_{1t} = Y_{1t}^I - Y_{1t}^N \quad (27)$$

Como a região 1 é exposta ao tratamento após o período T_0 , tem-se que para $t > T_0$, $Y_{1t} - Y_{1t}^I$. Para a região afetada pela intervenção, observam-se no período pós-intervenção um resultado potencial sobre intervenção. A mudança na política é dada pela estimação do contrafactual, Y_{1t}^N para $t > T_0$, ou seja, como se comportaria Y_{1t}^N na localidade afetada na ausência de tratamento. Isso é um resultado contrafactual, uma vez que a região afetada foi, por definição, exposta a intervenção após $t > T_0$. Fica claro, portanto, que Y_{1t}^I é observado, o problema de estimação do efeito da intervenção é equivalente ao problema de estimar Y_{1t}^N .

A estimação do efeito da intervenção consiste, portanto, na estimação de Y_{1t}^N , o qual corresponde ao valor da variável resultado na ausência da intervenção, usando uma unidade não afetada ou um pequeno número de unidades não afetadas que tenham características similares com aquelas da unidade afetada no momento da intervenção.

O método de controle sintético baseia-se, portanto, na suposição, de que a combinação de unidades no grupo com características semelhantes da unidade tratada são um grupo comparação significativamente melhor do que qualquer unidade afetada sozinha. Formalmente, o controle sintético pode ser representado por um vetor ($J \times 1$) de pesos $W = (w_2, \dots, w_J)$, com $0 \leq w_j \leq 1$ para $j=2, \dots, J$ e $w_2 + \dots + w_J = 1$. A escolha de um valor particular para W é equivalente a escolher um controle sintético.

Portanto, os estimadores de controle sintético de \hat{Y}_{it}^N e τ_{1t} são, respectivamente:

$$\begin{aligned} \hat{Y}_{it}^N &= w_2 Y_{2t} + \dots + w_J Y_{Jt} \quad e: \\ \hat{\tau}_{1t} &= Y_{1t} - \hat{Y}_{it}^N \end{aligned} \quad (28)$$

Portanto, a distância entre a variável de interesse efetiva e a variável de interesse sintética fornece o efeito decorrente da intervenção.

As principais limitações para aplicação desse método se referem: (i) ao tamanho do impacto e a volatilidade da variável de interesse, uma vez que pequenos efeitos de intervenção podem não ser isolados de choques aleatórios; (ii) o grupo placebo deve ser formado apenas por variáveis que não tiveram intervenção; (iii) inexistência de efeitos não antecipados da política, ou seja, o efeito da política deve ocorrer após sua intervenção; (iv) ausência de efeitos transbordamentos, ou seja, os benefícios de uma unidade tratada ultrapassam sua fronteira geográfica e (v) a combinação de unidades não tratadas consegue aproximar as características pré-intervenção na unidade afetada pela política.

4 EVIDÊNCIAS EMPÍRICAS APLICADAS AO CONTEXTO DOS PROGRAMAS OPERACIONALIZADOS PELO BANCO DO NORDESTE

No contexto dos programas operacionalizados pelo Banco do Nordeste, a aplicação de métodos de avaliação de impacto foi, inicialmente, motivada pela dificuldade de expansão das ações de apoio ao crédito. Desse modo, com objetivo de subsidiar ações de incremento e qualidade de acesso ao crédito nas atividades microempresariais urbanas nordestinas, o Banco do Nordeste encomendou a Fundação Getúlio Vargas – FGV⁸ estudo para avaliação do Crediamigo, principal programa de microcrédito do País.

O estudo deu origem ao livro “Microcrédito, o mistério nordestino e o Grameen brasileiro: perfil e performance dos clientes do Crediamigo”, organizado pelo professor Marcelo Neri, reunindo onze artigos. Dentro dessa perspectiva, o artigo de Neri e Medrado (2008) utilizou o método de

⁸ O CrediAamigo (Programa de Microcrédito Produtivo Orientado) foi concebido, em 1997, como projeto piloto em cinco unidades no Nordeste. No ano seguinte, foi expandido com a criação de 45 unidades.

diferenças em diferenças para analisar o impacto do Crediamigo sobre o microcrédito. Após controlar por outras variáveis relevantes que afetam a variável de interesse, os autores estimam que, entre 1997 e 2003, aumentou a chance de crédito no Nordeste tanto no agregado quanto no fluxo, aumentou a chance de estar endividado e diminuiu a chance de se reclamar da falta de crédito, bem como diminuiu a chance de se reclamar da falta de crédito somada à falta de capital.

Em avaliação do CrediaAmigo para os empreendimentos dos municípios atendidos pelo Banco do Nordeste no norte de Minas Gerais, Capobiango, Silveira e Braga (2012) usaram a metodologia de *propensity scores matching*. Os autores a partir da base do programa identificaram a primeira e última operação para formar o grupo controle e o grupo comparação. Desse modo, formaram o grupo tratamento com os clientes cujo prazo entre a primeira e a última operação foi de pelo menos 180 dias, e que estavam ativos em 31/12/2010, enquanto, os clientes que fizeram sua primeira operação entre junho e dezembro de 2010 foram considerados como grupo controle. Encontrou-se como resultado geral impacto positivo e significativo nas variáveis de resultado dos empreendimentos atendidos.

Dentro da mesma motivação do que foi realizado para o CrediAmigo, o Banco do Nordeste, na condição de principal financiador do Pronaf na região Nordeste e norte de Minas Gerais e Espírito Santo, realizou em parceria com a Fundação Instituto de Pesquisas Econômicas (Fipe), estudo para avaliação do Agroamigo e do Pronaf B tradicional. O referido estudo foi publicado na forma de livro, intitulado “Cinco anos de Agroamigo: retrato do público e efeitos do Programa”.

Nesse estudo, Abramovay et al. (2012) utilizaram o método de diferenças em diferenças para estimar o impacto do Programa em indicadores relacionados à produção, vendas e acumulação de riqueza. A estimativa de impacto do Programa é realizada comparando a evolução das variáveis de interesse em dois grupos: os indivíduos com mais tempo de exposição ao programa (grupo de tratamento), e os indivíduos ingressantes no Agroamigo (grupo comparação). A metodologia sugere efeitos positivos entre os clientes expostos a mais tempo em comparação aos ingressantes para todas as variáveis estudadas, entretanto, a utilização dos

ingressantes como grupo de comparação pode enviesar as estimativas⁹.

A limitação da construção de um grupo comparação formada por indivíduos não beneficiados também foi retratada na avaliação do PRONAF realizado pela consultoria Datamétrica (2013) para o Banco do Nordeste. De acordo com Relatório, o ideal seria utilizar agricultores beneficiados pelo programa para constituírem a categoria tratamento, e agricultores não beneficiados para constituírem a categoria comparação.

Na impossibilidade dessa estratégia, dada a inexistência de cadastros com nomes e endereços de agricultores não beneficiários, optou-se por formar a categoria comparação pelos beneficiários mais recentes do programa, ou seja, aqueles que possuem um único contrato de crédito com o Banco. A hipótese implícita para adoção desse critério é a de que o impacto do crédito concedido aos agricultores é pequeno devido ao pouco tempo de exposição ao Programa.

A política de desenvolvimento regional (PNDR) também tem sido extensamente avaliada por essas metodologias. Silva, Resende, Silveira Neto (2006) realizaram avaliação econômica dos Fundos Constitucionais de Financiamento do Nordeste (FNE) e do Norte (FNO) a partir de estimativas de *propensity scores* das firmas beneficiadas por esses fundos e de um conjunto de firmas não beneficiadas. Os autores focaram a avaliação de impacto nas variáveis taxa de variação do emprego e a taxa de variação dos salários médios pagos pelas firmas, entre os anos de 1995 e 1998. Além disso, a análise se restringiu às empresas tomadoras dos recursos do fundo, identificadas na RAIS (Relação Anual de Informações Sociais) de 1995. O trabalho apresenta como resultado geral, impacto positivo sobre a taxa de variação do número de empregados e ausência de impacto sobre a taxa de variação do salário médio pago pelas firmas.

9 Os autores argumentam que a utilização dos ingressantes do Agroamigo como grupo comparação pode refletir apenas tendências comuns a toda população, podendo não ter relação com o Programa. Para tentar isolar o efeito dessas outras variáveis e diminuir o viés nas estimativas de impacto, os autores controlam por outras variáveis que podem afetar as variáveis de interesse. Os autores argumentam, ainda, que os resultados da análise de impacto devem ser interpretados com cautela, porque existem características não observadas entre clientes mais antigos e clientes ingressantes que são difíceis de mensurar, e que poderia estar influenciando um crescimento mais rápido em indicadores como produção, vendas e riqueza, o que tenderia a superestimar os efeitos do Programa.

Posteriormente, os mesmos pesquisadores (SILVA; RESENDE; SILVEIRA NETO, 2009) realizaram avaliação para os três Fundos Constitucionais - FNE, FNO e FCO – comparando a eficácia dos gastos desses Fundos na taxa de variação do número de empregados e na taxa de variação do salário médio pago pelas firmas por meio de estimativas de *propensity scores* das firmas beneficiadas por esses fundos e de um conjunto de firmas não beneficiadas, identificadas na RAIS, entre 2000 e 2003. Em relação ao FNE, tem-se como resultado geral que não foi possível encontrar impacto positivo da aplicação dos recursos para taxa de variação do salário médio; por outro lado, verificou-se impacto positivo para a taxa de variação do número de empregados.

Soares, Sousa e Pereira Neto (2009) expandiram a análise de Silva, Resende e Silveira Neto (2009) com a inclusão de mais períodos de avaliação para observar a evolução de impactos, ano a ano, após a concessão dos financiamentos nas variáveis, estoque de emprego, massa salarial e salário médio, entre 1999 e 2005. Os resultados indicam impacto considerável do FNE no estoque de emprego das empresas do Nordeste e da massa salarial, mas não se encontra impacto no salário médio. Ou seja, o aumento das contratações ocorreu sob os salários vigentes.

Análise semelhante foi realizada por Soares et al. (2014), os autores aplicaram a metodologia de *matching* com *propensity scores* para avaliar os impactos do FNE (Fundo Constitucional de Financiamento do Nordeste) em variáveis do mercado de trabalho formal. Foi avaliado o crescimento no estoque do emprego, do salário médio e da massa salarial ao longo de cinco anos, utilizando como base de dados a RAIS, a qual foi desagregada por setores de atividade e por áreas específicas do Nordeste (Semiárido fora do Semiárido, e regiões da Política Nacional de Desenvolvimento Regional - PNDR). Encontrou-se, como resultado, geral, que os diferenciais são favoráveis às empresas beneficiadas pelo FNE, e que esse diferencial cresce à medida que o período de tempo aumenta.

Além da abordagem usual de efeitos fixos e pareamento por Escore de Propensão, Oliveira et al. (2017) avaliaram o efeito dos Fundos Constitucionais sobre o volume de emprego e produtividade do trabalho (salário médio). A técnica de *propensity score* Generalizado é utilizada pelos

autores para tentar captar o efeito dose-resposta da política. Os resultados da estimação de efeitos fixos mostraram um efeito positivo de 2,5 p.p. sobre o salário médio pago pelas empresas beneficiadas pelos empréstimos do FNE, quando comparadas às não beneficiadas. Essa abordagem não mostrou diferença estatisticamente relevante no volume total de emprego entre as empresas. Da mesma forma, o modelo de *propensity score matching* também não encontrou qualquer efeito do Fundo sobre o salário ou o nível de emprego. Por fim, a função dose resposta resultante do método generalizado sugere um efeito positivo, mas não linear do FNE sobre os salários, mas não é conclusiva.

O PRODETUR/NE-II¹⁰ programa de fomento ao turismo na região Nordeste, cuja operacionalização foi realizada, em conjunto, pelo Banco do Nordeste, Banco Interamericano de Desenvolvimento (BID) e os governos federal, estadual e municipal, teve avaliação de impacto realizada pelo método de controle sintético. Para ser avaliado por esse método, o programa tem que satisfazer a dois requisitos: i) a formação de unidade de controle e tratamento e ii) definição de um marco lógico temporal de implementação de política.

Em relação ao primeiro requisito, observa-se que os polos não tratados são, realmente, potenciais componentes sintéticos dos polos beneficiados. Como características comuns observa-se que a maioria dos polos está localizada em regiões costeiras, sugerindo o mesmo tipo de estrutura turística, além de baixo efeito transbordamento dada a pequena área de fronteira comum entre os polos beneficiados.

O segundo requisito é cumprido com alguma limitação, dado que os recursos e serviços nos polos beneficiados foram sendo realizados ao longo dos anos, dificultando a definição de um marco lógico temporal, pois embora o programa tenha começado, em 2004, uma soma mais expressiva dos recursos apenas foi investida a partir de 2007, e uma soma superior a de 2007, foi liberada, em 2010 (Gráfico 1).

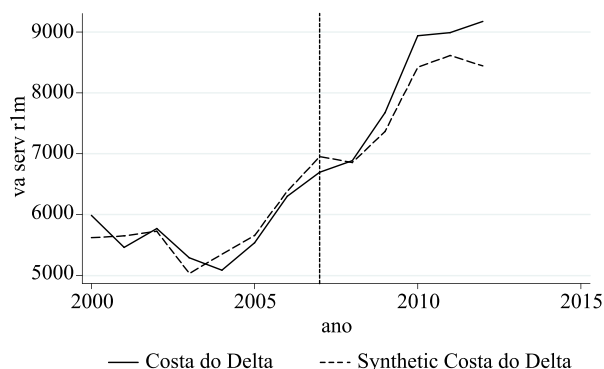
A despeito do PRODETUR/NE-II atender aos requisitos exigidos pelo método de controle sintético, mesmo que, parcialmente, apenas em 28 situa-

10 O PRODETUR-II teve como objetivo ampliar as ações voltadas para o Programa de Desenvolvimento do Turismo (PRODETUR-I), iniciado em 1995, com objetivo de incrementar a atividade turística nordestina na área de atuação da SUDENE, o que envolve os nove estados nordestinos e à região norte do Estado de Minas Gerais.

ções das setenta e duas possibilidades de utilização do método (dadas pelos nove polos de tratamento e as oito variáveis de resultados selecionadas), foi possível encontrar um grupo de controle sintético. Na maior parte dos casos em que não foi possível encontrar resultados, a condição de convexidade não foi atendida, ou seja, o polo tratado possuía valores muito mais alto ou muito mais baixo em alguma variável de interesse que todos aqueles não tratados, de tal forma que a formação de grupo de controle sintético não conseguiu ser dada pela condição de convexidade¹¹.

No geral, isso aconteceu para as variáveis que são influenciadas diretamente pelo tamanho dos pólos como PIB, PIB per *capita*, valor adicionado dos serviços, empresas, empregos e arrecadação de impostos todas em nível. Para as variáveis dadas em termos de participação, tal como a participação dos serviços no valor adicionado e o salário médio, foi possível estimar resultados.

Gráfico 1 – Evolução do Valor Adicionado dos Serviços Costa do Delta versus Sintético



Fonte: Relatório de avaliação de impacto do Programa de Desenvolvimento do Turismo do Nordeste (Banco do Nordeste do Brasil).

5 CONSIDERAÇÕES FINAIS

A literatura de avaliação mostra diversas alternativas para estimação do impacto de um Programa. A seleção do método dependerá de muitas questões, em especial, de como é definida a participação no tratamento, que pode ser aleatória ou baseada em características observadas ou não pelo avaliador. A inferência correta do parâmetro de interesse dependerá da validade da hipótese de identificação de cada método. Portanto, o avaliador deverá escolher o método com base em uma

¹¹ Principalmente, para aqueles pólos que incluem as grandes capitais do Nordeste.

observação cuidadosa do processo de participação no tratamento, a fim de fundamentar a hipótese de identificação do método adotado.

Quando a designação aleatória ao tratamento é bem realizada, o pesquisador não precisa se preocupar com o viés de seleção, uma vez que a aleatorização torna o tratamento independente dos resultados potenciais. Entretanto, quando as unidades participam de um programa de forma voluntária, existe um viés de seleção que precisa ser isolado.

Nesses casos, métodos não experimentais fornecem como alternativa o controle por características que influenciam na decisão de participação do tratamento. Se a participação do tratamento é definida por características observadas pelo avaliador, métodos de seleção sob observáveis fornecem uma boa solução para o problema de viés de seleção. Mas, se existirem fatores não observados, que são decisivos para a participação do tratamento, a estimação do efeito tratamento dependerá de hipóteses adicionais, identificadas por métodos de seleção sob não observáveis.

Por fim, as evidências empíricas encontradas nas avaliações dos Programas operacionalizados pelo Banco do Nordeste reforçam que uma das maiores dificuldades na estimação do impacto de um Programa está na construção de um grupo comparação válido. As avaliações, tanto internas quanto externas, desses programas têm abordado essa limitação, bem como buscado maneiras de contorná-la por meio dessas metodologias.

REFERÊNCIAS

- ABADIE, A; DIAMOND, A.; HAINMUELLER, J. Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, v. 59, n. 2, p. 495–510, 2015.
- ABADIE, A; GARDEAZABAL, J. The Economic Costs of Conflict: A Case Study of the Basque Country. *The American Economic Review*, v. 93, n. 1. p. 113-132, 2003.
- ABRAMOVAY, R et al. **Cinco Anos de Agro-migo**: retrato do público e efeitos do Programa. Fortaleza: Banco do Nordeste, 113 p., 2012.
- DATAMÉTRICA – CONSULTORIA, PESQUISA E TELEMARKEETING. **Avaliação do PRONAF** – Relatório Final. Datamé-

trica – Consultoria, Pesquisa e Telemarketing. Recife: Datamétrica, 216p., 2013.

BANCO DO NORDESTE DO BRASIL. **Relatório de avaliação de impacto do Programa de Desenvolvimento do Turismo do Nordeste** – Segunda fase (PRODETUR/NE-II): Condicionalidades metodológicas, resultados exploratórios e perspectivas. Fortaleza: Banco do Nordeste, 60p., 2016.

CAPOBIANGO, R. P.; SILVEIRA, S. de F. R.; BRAGA, M. J. Avaliação do Impacto do Programa CrediAmigo nos Empreendimentos de Municípios no Norte de Minas Gerais. In: **Encontro da ANPAD**, 36., Rio de Janeiro: 22 a 26 de setembro de 2012. Disponível em: < http://www.anpad.org.br/admin/pdf/2012_APB1936.pdf>. Acesso em: 13 de set. 2016.

DUFLO, E; GLENNESTER, R; KREMER, M. Using Randomization in Development Economics Research: A Toolkit. 2006.

IMBENS, G; ANGRIST, J. Identification and Estimation of Local Average Treatment Effects. In: **Econometrica**, v. 62, n. 2, p. 467-475, 1994.

HECKMAN, J; TOBIAS, J; VYTLACIL. Four Parameters of Interest in the Evaluation of Social Programs. In: **Southern Economic Journal**, v. 68, n. 2, p. 210-223, 2001.

HIRANO, K.; IMBENS, G. The propensity score with continuous treatments. In: GELMAN, A.; MENG, X.-L. (Eds.). **Applied Bayesian modeling and causal inference from incomplete-data perspectives**. Nova Iorque: Wiley, 2004.

HOLLAND, P. Statistics and Causal Inference. In: **Journal of the American Statistical Association**, n. 84, p. 862-880, 1986.

LEE, D. S.; LEMIEUX, T. Regression discontinuity designs in economics. **Journal of economic literature**, v. 48, n. 2, p. 281-355, 2010.

IMBENS, G. The role of the propensity score in estimating dose-response functions. **Biometrika**, v. 87, n. 3, p. 706-710, 2000.

NERI, M; MEDRADO, A. O Mistério Nordestino. In: **Microcrédito, o Mistério Nordestino e o Grameen Brasileiro**: Perfil e performance

dos clientes do Crediamigo. NERI, M. (orgs). Rio de Janeiro: Editora FGV, 2008. 376 p.

OLIVEIRA, G. R.; RESENDE, G. M.; GONÇALVES, C. N.; SILVA, D. F.; Efeitos Não Lineares dos Fundos Constitucionais de Financiamento entre 2000 e 2012. 45º Encontro Nacional de Economia, 2017. Disponível em: <https://www.anpec.org.br/encontro/2017/submissao/files_I/i10-88c58a83d6d4c2fe7e-83012477f3141c.pdf>. Acesso em: 16/07/2018.

ROY, A. D. Some Thoughts on the Distribution of Earnings. In: **Oxford Economic Papers**, v. 3, n. 2, p. 135-146, 1951.

ROSENBAUM, P; RUBIN, D. The Central Role of the Propensity Score in Observational Studies for Causal Effects. In: **Biometrika**, v. 70, n. 1, p. 41-55, 1983.

SMITH, J. Evaluating Local Economic Development Policies: Theory and Practice. In: **Alistair Nolan and Ging Wond** (eds). Paris: OECD, p. 287-332, 2004.

SILVA, A. M. A. DA; RESENDE, G. M.; SILVEIRA NETO, R. M. **Avaliação econômica dos Fundos Constitucionais de Financiamento do Nordeste (FNE) e do Norte (FNO)**. IPEA, Brasília, Texto para discussão, n. 1207, p. 1-37, 2006.

_____. **Eficácia do Gasto Público: Uma Avaliação do FNE, FNO e FCO**. Estudos Econômicos, São Paulo, v. 39, n. 1, p. 89-125, 2009. Disponível em:< <http://www.scielo.br/pdf/ee/v39n1/v39n1a04.pdf>>. Acesso em: 15 set. 2016.

SOARES, R. B. et al. **Avaliações de Impacto e Eficiência das Empresas Beneficiadas pelo FNE: Geração de emprego, massa salarial e salário médio 2000-2008**. Fortaleza: Banco do Nordeste do Brasil, 103 p., 2014.

SOARES, R.B; SOUSA, J. M. P; PEREIRA NETO, A. **Avaliação de Impacto do FNE no Emprego, na Massa Salarial e no Salário Médio em Empreendimentos Financiados**. Revista Econômica do Nordeste, v. 40, n. 1, p. 217-234, 17p., 2009.

THISTLETHWAITE, D; CAMPBELL, D. Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment. In: **Journal of Educational Psychology**, v. 54, p. 309-17, 1960.

