

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE CIÊNCIAS ECONÔMICAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA**

**LUIZ FELIPE DE VASCONCELLOS SILVEIRA**

**A AVALIAÇÃO DO IMPACTO DE UM TREINAMENTO UTILIZANDO  
*PROPENSITY SCORE MATCHING*: UMA ABORDAGEM NÃO-PARAMÉTRICA E  
SEMIPARAMÉTRICA**

**Porto Alegre**

**2015**

**LUIZ FELIPE DE VASCONCELLOS SILVEIRA**

**A AVALIAÇÃO DO IMPACTO DE UM TREINAMENTO UTILIZANDO  
*PROPENSITY SCORE MATCHING*: UMA ABORDAGEM NÃO-PARAMÉTRICA E  
SEMIPARAMÉTRICA**

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título de Mestre em Economia.

Orientador: Prof. Dr. Hudson da Silva Torrent

**Porto Alegre**

**2015**

CIP - Catalogação na Publicação

Silveira, Luiz Felipe de Vasconcelos

A avaliação do impacto de um treinamento utilizando propensity score matching: uma abordagem não-paramétrica e semiparamétrica / Luiz Felipe de Vasconcelos Silveira. -- 2015.

69 f.

Orientador: Hudson da Silva Torrent.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Faculdade de Ciências Econômicas, Programa de Pós-Graduação em Economia, Porto Alegre, BR-RS, 2015.

1. Avaliação de impacto. 2. Propensity score matching. 3. Econometria não-paramétrica. 4. Econometria semi-paramétrica. I. Torrent, Hudson da Silva, orient. II. Título.

**LUIZ FELIPE DE VASCONCELLOS SILVEIRA**

**A AVALIAÇÃO DO IMPACTO DE UM TREINAMENTO UTILIZANDO  
*PROPENSITY SCORE MATCHING*: UMA ABORDAGEM NÃO-PARAMÉTRICA E  
SEMIPARAMÉTRICA**

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título de Mestre em Economia.

Aprovada em: Porto Alegre, 16 de dezembro de 2015.

BANCA EXAMINADORA:

---

Prof. Dr. Hudson da Silva Torrent – Orientador  
UFRGS

---

Prof. Dr. Daniel de Abreu Pereira Uhr  
UFPEL

---

Prof. Dr. Marco Tulio Aniceto Franca  
PUCRS

---

Prof. Dr. Sabino da Silva Porto Junior  
UFRGS

## RESUMO

O objetivo dessa dissertação é avaliar o impacto de um programa de treinamento voltado para trabalhadores, utilizando o *propensity score matching*, mas com dois tipos de abordagem, uma não-paramétrica e a outra semi-paramétrica. Para estimação não paramétrica foi utilizado um método proposto por Li, Racine e Wooldridge (2009) e para estimação semi-paramétrica, o modelo utilizado foi o *Generalized Additive Model* proposto por Hastie e Tibshirani (1990). Os resultados obtidos indicam que os dois métodos utilizados apresentam estimativas tão boas ou melhores do que quando estimadas parametricamente.

**Palavras-chave:** Avaliação de impacto. *Propensity score matching*. Econometria Não-paramétrica. Econometria semi-paramétrica.

## ABSTRACT

The goal of this thesis is to evaluate the impact of a job training program using propensity score matching methods with two types of approaches: a nonparametric e another semiparametric. For non-parametric estimation was used a method proposed by Li, Racine and Wooldridge (2009) and for the semiparametric model the Generalized Additive Model proposed by Hastie and Tibshirani (1990). The results indicate that both methods provide estimates as good or better than when parametrically estimated.

**Keywords:** Impact evaluation. Propensity score matching. Nonparametric econometrics. Semiparametric econometrics.

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	7
<b>2</b>	<b>REFERENCIAL TEÓRICO</b> .....	9
2.1	AVALIAÇÃO DE IMPACTO .....	9
2.2	<i>JOB TRAINING</i> .....	10
2.3	MODELO CAUSAL DE RUBIN (RCM).....	11
<b>2.3.1</b>	<b>Resultados Potenciais</b> .....	12
<b>2.3.2</b>	<b>Mecanismo de Atribuição</b> .....	14
<b>2.3.3</b>	<b>Interações e Equilíbrio Geral</b> .....	15
2.4	EFEITOS MÉDIO DO TRATAMENTO .....	15
2.5	O PROBLEMA DO VIÉS DE SELEÇÃO.....	16
<b>3</b>	<b>METODOLOGIA</b> .....	19
3.1	EXPERIMENTOS ALEATÓRIOS.....	19
3.2	<i>PROPENSITY SCORE MATCHING</i> .....	22
<b>3.2.1</b>	<b>O Propensity Score</b> .....	23
<b>3.2.2</b>	<b>Matching (Pareamento)</b> .....	27
3.3	ABORDAGEM NÃO -PARAMÉTRICA.....	29
<b>3.3.1</b>	<b>Estimador Kernel de Densidade</b> .....	29
<b>3.3.2</b>	<b>Least Squares Cross Validation</b> .....	30
<b>3.3.3</b>	<b>Regressão Não-Paramétrica</b> .....	32
<b>3.3.4</b>	<b>Propensity Score Não-Paramétrico</b> .....	35
3.4	ABORDAGEM SEMI-PARAMÉTRICA .....	38
<b>3.4.1</b>	<b>GAM</b> .....	38
<b>3.4.2</b>	<b>Estimadores Spline</b> .....	39
<b>3.4.3</b>	<b>Propensity Score Semi-paramétrico</b> .....	40
<b>4</b>	<b>APLICAÇÃO ECONÔMICA</b> .....	41
4.1	INTRODUÇÃO .....	41

4.2	BASE DE DADOS .....	41
4.3	RESULTADOS EXPERIMENTAIS.....	44
4.4	RESULTADOS NÃO-EXPERIMENTAIS.....	45
4.5	ESTIMAÇÃO NÃO-PARAMÉTRICA .....	46
4.5.1	<b>Regressão com <i>Propensity Score</i> Não-Paramétrico .....</b>	<b>47</b>
4.5.2	<b>PSM Não-Paramétrico .....</b>	<b>50</b>
4.6	ESTIMAÇÃO SEMI-PARAMÉTRICA .....	57
4.6.1	<b>Regressão com <i>Propensity Score</i> Semi-Paramétrico .....</b>	<b>57</b>
4.6.2	<b>PSM Semi-paramétrico .....</b>	<b>60</b>
5	CONCLUSÕES.....	63
	REFERÊNCIAS.....	65



## 1 INTRODUÇÃO

Esse trabalho consiste em pesquisar alguns dos métodos mais modernos em econometria não-paramétrica e semi-paramétrica e aplicá-los num campo muito importante da economia, o de avaliação de impacto. Para avaliarmos o impacto que uma política ou programa teve sobre um certo grupo de indivíduos utilizaremos o Efeito Médio do Tratamento sobre os Tratados (ATT). A utilização dessa técnica, inicialmente confinado ao ambiente médico, para avaliação das respostas dos pacientes à doses de remédios, hoje em dia, é usada em diversos campos da ciência (LI; RACINE; WOOLDRIDGE, 2009). O efeito médio do tratamento pode ser estimado de diferentes maneiras, entretanto, nesse trabalho focaremos na utilização do *propensity score* (ROSENBAUM; RUBIN, 1983, 1984) para estimá-lo.

O termo não-paramétrico é usado para se referir às técnicas estatísticas que não requerem a especificação da forma funcional do objeto sendo estimado. Em vez de presumir que a forma funcional do objeto é conhecida através de alguns parâmetros desconhecidos, propõe-se a utilização de hipóteses menos restritivas tais qual, diferenciabilidade e restrições de momento (WAND; JONES, 1995).

Como técnicas não-paramétricas fazem menos hipóteses acerca do objeto sendo estimado do que técnicas paramétricas, os estimadores não-paramétricos tendem a convergir mais devagar para os objetos sendo estudados do que estimadores paramétricos bem especificados. Além disso, a taxa de convergência é inversamente relacionada com os números de variáveis envolvidas o que gera o problema conhecido como “Curse of Dimensionality”, ou maldição da dimensionalidade. Os modelos não-paramétricos, mesmo através de dados mais modestos, podem revelar estruturas nos dados que não seriam possíveis caso fossem usadas especificações paramétricas, portanto são mais adequados em situações onde pouco se sabe sobre a forma do objeto sendo estudado, o número de variáveis é pequeno, e quando a base de dados é relativamente grande (LI; RACINE, 2007). Uma justificativa dada por Yatchew (1998) para utilização de métodos não-paramétricos é de que tais ferramentas nos apresentam inferências mais robustas, o que torna os resultados empíricos mais valiosos. Do ponto de vista da análise dos dados, o ganho de se utilizar essa técnicas advém do fato dos seus estimadores e procedimentos de inferência serem menos dependentes de algum tipo de forma funcional.

Já os modelos semi-paramétricos conseguem unir modelagem paramétrica e modelagem não-paramétrica, oferecendo uma certa flexibilidade que os modelos não-paramétricos não possuem, de tal forma que as hipóteses feitas acerca da forma funcional são

mais fortes do que as feitas em modelos não-paramétricos, mas mais fracas do que as feitas pelos modelos paramétricos, sendo assim, possibilitam a redução de um erro na especificação do modelo (HOROWITZ, 2009). São conhecidos na literatura estatística como Redutores de Dimensão.

Na literatura econômica encontramos muitos trabalhos acerca da utilização de *propensity score matching* como Dehejia e Wahba (1999, 2002); Heckman, Ichimura e Todd (1997, 1998), Smith e Todd (2005); Abadie e Imbens (2006) e outros. O *matching* envolve o pareamento de indivíduos que receberam algum tipo de tratamento (participação em algum programa) com indivíduos que não receberam (ou não participaram), que são similares em relação às características observáveis, e quando as diferenças relevantes entre dois indivíduos quaisquer é capturada por essas características, ou seja, os resultados potenciais independem da participação no programa condicional às covariáveis pré-tratamento, métodos de pareamento podem retornar estimativas não viesadas para o impacto desse tratamento (DEHEJIA; WAHBA, 2002).

Na maioria das vezes o método utilizado para a estimação do *propensity score* (probabilidade condicional de receber o tratamento originalmente) são modelos paramétricos como *logit* ou *probit*. A vantagem que podemos obter estimando-os não-parametricamente ou semi-parametricamente é de que podemos obter os impactos de alguma política ou programa, de forma relativamente mais fácil, pois não precisaremos fazer hipóteses tão restritivas.

Diante do que foi exposto, o presente trabalho tem o intuito de responder à seguinte pergunta: Os métodos não-paramétricos e semi-paramétricos são bons substitutos para o método tradicional de se estimar o impacto de um programa? Se sim, devemos incorporá-los aos métodos já existente na literatura econômica.

Esse trabalho está organizado em cinco capítulos. Além da introdução, no capítulo 2 abordamos o referencial teórico do trabalho, no capítulo 3, apresentamos os métodos utilizados para estimação do impacto de um programa. Já no capítulo 4, utilizamos uma aplicação econômica para discutirmos e analisarmos a estimação do impacto de um programa, utilizando os métodos propostos. Por fim, no capítulo 5, apresentamos as conclusões acerca do nosso trabalho.

## 2 REFERENCIAL TEÓRICO

Nesse capítulo iremos abordar o referencial teórico utilizado nesse trabalho, ele está subdividido em cinco seções.

### 2.1 AVALIAÇÃO DE IMPACTO

Programas públicos desenhados pelos governos, sejam eles na esfera municipal, estadual ou federal, são desenvolvidos com o intuito de atingir um público alvo e também alcançar certos objetivos. Eles podem parecer bastante promissores antes da implementação, porém podem não obter os impactos ou benefícios esperados. Nesse cenário, a avaliação de impacto se torna extremamente necessária, pois ajuda os formuladores de políticas a decidirem se os programas tiveram o efeito pretendido ou não, a alocar os recursos entre o programas existentes de forma ótima, a compreender o que funciona ou não, e por último, a atribuir se as mudanças nas variáveis escolhidas são, de fato, resultado da intervenção ou não (KHANDKER; KOOLWAL; SAMAD, 2010). De acordo com Gertler et al. (2011), diversas abordagens podem ser utilizadas para se avaliar programas, as três principais são monitoramento (para mais detalhes sobre essa abordagem, ver Kusek e Rist, 2004), avaliação operacional e avaliação de impacto. Nesse trabalho, utilizaremos somente a terceira abordagem.

A avaliação de impacto nos ajuda a compreender se as mudanças no bem-estar dos tratados são de fato devido à intervenção ou outros fatores. Uma avaliação efetiva deve ser capaz de analisar os mecanismos pelos quais os beneficiários estão respondendo à intervenção. Como veremos nas seções seguintes, na ausência de dados contrafactuais, avaliações de impacto podem ser bastante eficientes em identificar os efeitos do programa através da aplicação de diferentes métodos na construção de grupos de comparação para os participantes, o principal objetivo é isolar o efeito do programa de outros fatores e também do potencial viés de seleção.

Essa abordagem abrange métodos qualitativos e quantitativos, assim como métodos *ex ante* e *ex post*. A análise qualitativa busca aferir possíveis impactos que o programa pode gerar, os mecanismos desses impactos, e também a extensão dos benefícios para destinatários, através de pesquisas realizadas com os participantes da intervenção. Os resultados quantitativos podem ser generalizados, já os resultados qualitativos não, porém, os métodos

qualitativos geram informações importantes para a compreensão dos mecanismos do programas.

Uma avaliação *ex ante* tenta medir o impacto de políticas futuras, dado a situação atual da área que receberá a intervenção, e pode utilizar simulações baseadas em hipóteses de como a economia funciona. Muitas vezes essas avaliações se baseiam em modelos estruturais econômicos, em contraste, avaliações *ex post* utilizam dados recolhidos e estimam os impactos acumulados pelos beneficiários, atribuídos pela intervenção do programa (GERTLER et al, 2011). Possuem benefícios imediatos e refletem a realidade, porém, podem custar muito mais do que uma avaliação *ex ante* pois necessitam de coleta de dados dos resultados finais do programa, tanto dos participantes quanto dos não-participantes. Um custo adicional que pode ocorrer é a falha na intervenção, algo que seria previsto com uma avaliação *ex ante*.

O principal problema do pesquisador é medir os efeitos dessa intervenção, para que os *policy makers* possam decidir se o programa produz resultados válidos, e assim decidir se ele é mantido, expandido ou desmantelado. Para esse trabalho será utilizado o efeito médio do tratamento, que é um método de avaliação de impacto quantitativo e *ex post*. Na próxima seção será feita uma breve revisão da literatura acerca da utilização de avaliações de impacto em programas de treinamento voltados para trabalhadores, conhecido como *Job Training*.

## 2.2 JOB TRAINING

O *job training* é um programa de treinamento voltado para o aumento da capacidade produtiva dos indivíduos participantes. Muitos governos utilizam esses programas como políticas públicas. No passado por exemplo, nas décadas de sessenta e setenta, a utilização de treinamento para trabalhadores se tornou uma obsessão por parte do governo dos Estados Unidos. O governo defendia esse tipo de programa por três razões principais:

- a) redução da inflação através do provimento de trabalhadores mais capazes para auxiliar nos problemas de curto prazo da oferta;
- b) para reduzir o desemprego em certos grupos demográficos e
- c) para reduzir a pobreza através do aumento da capacidade dos trabalhadores contemplados (ASHENFELTER, 1974, 1978).

Para Ashenfelter (1978), avaliar o impacto dos programas de treinamento é no final das contas, uma avaliação quantitativa do efeito que o programa teve no salário dos

participantes, e esse é um processo muito importante por duas razões, a primeira é porque ajuda a disseminar a discussão sobre esses tipos de políticas públicas (se são necessárias, o quanto alocar de recursos públicos, etc.) e a segunda, porque proporciona aos formuladores de políticas (*policy makers*) meios de se testar a capacidade que esses programas possuem de aumentar o capital humano dos trabalhadores. Para ver mais sobre avaliação de impacto voltado para o treinamento de indivíduos, ler Ashenfelter e Card (1985), Heckman e Robb (1985) e Card e Sullivan (1988).

### 2.3 MODELO CAUSAL DE RUBIN (RCM)

Um dos maiores obstáculos para os modelos de causalidade foi chamado por Holland (1986) como “problema fundamental da inferência causal”. Formalmente, seja  $X$  a hipotética causa e  $Y$  o resultado. Manipulando o valor de  $X$  podemos obter mudanças no valor de  $Y$ , e supondo que o valor de  $X$  mude de  $x_1$  para  $x_2$ , então a medida do impacto em  $Y$  é medido comparando dois valores:  $y_2 = Y(x_2)$ , que resulta da mudança, e  $y_1 = Y(x_1)$ , que teria ocorrido se não houvesse mudança em  $x$ . Entretanto, se houve mudança em  $X$ , então o valor de  $Y$  na ausência da mudança,  $y_1$ , não pode ser observado, logo nada mais pode ser dito sobre o impacto causal sem que algumas hipóteses sejam estabelecidas acerca do valor de  $Y$ , na ausência da mudança de  $X$ . Ele atesta que é impossível observar o valor de  $y_2$  e  $y_1$  num indivíduo (ou unidade), portanto é impossível observar o efeito do tratamento.

No caso do *job training* queremos observar qual o aumento do salário advindo do treinamento recebido, mas só observamos o indivíduo em um estado, ou ele participou do programa e recebeu o treinamento ou não participou, nunca os dois. Para contornar esse problema utilizaremos métodos estatísticos, que implica em utilizar a população  $N$  para se estimar o efeito causal médio. O valor  $y_1$  é conhecido como contrafactual, que representa um valor hipotético que não pode ser observado.

Uma abordagem relativamente moderna da microeconometria fornece um bom arcabouço teórico para estimação de parâmetros causais. Ele é conhecido como o Modelo Causal de Rubin, ele recebe esse nome devido aos trabalhos seminais desenvolvidos pelo estatístico Donald Rubin (1974, 1976, 1978). Além de Rubin, no começo do século XX, Neyman (1923) também propôs um modelo similar assim como Fisher (1925) e depois Roy (1951). Nessa seção são descritos os principais elementos acerca da avaliação de impacto, baseados no trabalho de Rubin.

### 2.3.1 Resultados Potenciais

O primeiro elemento importante do modelo RCM é a abordagem dos resultados potenciais. O efeito causal de um tratamento num indivíduo ou numa unidade de observação (famílias, empresas, etc.) é a comparação entre o valor do resultado caso ele seja tratado e o valor do resultado caso não seja tratado. Como veremos mais a frente o alvo da estimação geralmente é o efeito causal médio, definido como a diferença média entre os resultados dos tratados e não-tratados em toda amostra ou população.

Formalmente, suponha que um governo queira analisar a participação num programa utilizando observações de  $N$  indivíduos,  $i = 1, 2, \dots, N$ . Temos que  $t$  é a variável binária que indica se houve participação, onde  $t_i = 1$  implica que o indivíduo participou do programa e  $t_i = 0$  implica que não houve participação. Além disso, cada unidade possui um vetor de características  $X_i$ , chamado de covariáveis.

Existem dois resultados possíveis, são eles  $y_i(1)$  e  $y_i(0)$ . O primeiro indica o resultado que será realizado caso o indivíduo  $i$  participe do programa, o segundo denota o resultado que será realizado caso o indivíduo não participe do programa. Só existem duas situações possíveis para o indivíduo  $i$ , participar ou não participar, logo somente um dos resultados pode ser realizado.

Se houver participação no programa, por parte do indivíduo  $i$ ,  $y_i(1)$  será o resultado realizado e  $y_i(0)$  será o resultado contrafactual. Caso contrário, os papéis se invertem, como o indivíduo  $i$  não participou do programa,  $y_i(0)$  será o resultado realizado e  $y_i(1)$  será o resultado contrafactual. E como só podemos observar  $y_i(1)$  ou  $y_i(0)$ , não ambos para cada pessoa, então, juntamente com  $t_i$ , o resultado observável será:

$$y_i = (1 - t)y_i(0) + ty_i(1) = y_i(0) + t(y_i(1) - y_i(0)) \quad (2.1)$$

Então, teremos

$$\begin{aligned} y_i &= y_i(0) \text{ se } t = 0 \\ y_i &= y_i(1) \text{ se } t = 1 \end{aligned} \quad (2.2)$$

Podemos especificar  $y_i(0)$  e  $y_i(1)$  na forma de equações lineares, então teremos:

$$y_i(1) = X_i' \alpha + \beta_i + \mu_i \quad (2.3)$$

$$y_i(0) = X_i' \alpha + \mu_i \quad (2.4)$$

Onde  $\mu_i$  é o termo não observável que influencia os resultados potenciais do indivíduo  $i$ , e  $\beta_i$  fornece o impacto da intervenção, ou seja,  $\beta_i = y_i(1) - y_i(0)$ .

Substituindo as equações (2.3) e (2.4) na equação (2.1) temos<sup>1</sup>:

$$y_i = X_i' \alpha + \beta t_i + \mu_i \quad (2.5)$$

Mais à frente na seção 2.4 desse capítulo, iremos mostrar o problema que ocorre ao estimarmos essa equação.

Segundo Imbens e Wooldridge (2009), a abordagem dos resultados potenciais possui diversas vantagens sobre a abordagem baseada diretamente nos resultados realizados (modelos estruturais<sup>2</sup>). A primeira vantagem é que ela permite os pesquisadores definir os efeitos causais antes de especificar o mecanismo de atribuição (mecanismo que decide se o indivíduo vai participar ou não) e sem fazer hipóteses acerca da forma funcional ou distribuições. Como visto acima, a definição mais comum do efeito causal sobre a unidade é a diferença  $y_i(1) - y_i(0)$ , o que torna desnecessário fazer afirmações sobre os efeitos acerca da sua variabilidade, ou seja, se eles são constantes ou não na população. Além disso, usando essa a abordagem dos resultados potenciais para definir o efeito de tratamento individual não precisamos assumir exogeneidade ou endogeneidade dos mecanismo de atribuição.

A segunda vantagem é que essa abordagem vincula análise dos efeitos-causais à manipulações explícitas, em outras palavras, considerando os dois resultados potenciais o pesquisador é forçado a pensar em cenários sob o qual cada resultado pode ser observado, ou seja, deve considerar quais tipos de experimentos que podem revelar os efeitos causais.

A terceira vantagem é que separa as duas modelagens, a dos resultados potenciais e do mecanismo de atribuição. Modelar o resultado realizado é complicado pois devemos combinar os resultados potenciais e o mecanismo de atribuição.

A quarta vantagem é que permite que se formule hipóteses probabilísticas sobre as variáveis potencialmente observáveis ao invés dos componente não observáveis. Nessa

<sup>1</sup> Nesse caso, assumimos a hipótese de que o efeito da intervenção é homogêneo para qualquer indivíduo  $i$  da população, ou seja,  $\beta_i = \beta$ .

<sup>2</sup> Para uma leitura mais completa sobre a utilização de modelos estruturais na avaliação de programas ver Heckman e Vytlačil (2005).

abordagem, a maioria das hipóteses críticas serão formuladas como hipóteses de independência condicional envolvendo resultados potenciais.

A quinta vantagem é que nos mostra de onde vem a incerteza dos estimadores. Mesmo se pudéssemos observar toda a população finita para estimarmos a média da população sem incerteza, ainda assim os efeitos causais seriam incertos, pois para cada unidade, somente observamos um resultado.

### 2.3.2 Mecanismo de Atribuição

Como dito anteriormente, é o mecanismo que atribui se o indivíduo participa ou não do programa. É definido como a probabilidade condicional de receber o tratamento, em função dos resultados potenciais e das covariáveis observáveis. Existem três classes distintas de mecanismos de atribuição (IMBENS, 2004)

A primeira classe é a de experimentos aleatórios. Nos experimentos aleatórios, a probabilidade de participação no tratamento não varia com os resultados potenciais e pode ou não ser função das covariáveis. Como será visto no capítulo 3, o processo de aleatorização pode ser feita puramente, quando não é condicionada à nenhuma covariável, ou parcialmente, quando condicionamos à covariáveis (como renda, local onde mora, etc.), sendo a segunda forma predominante na literatura econômica. Além disso, o caso mais utilizado é quando é realizado um experimento aleatório completo, ou seja, numa população de  $N$  unidades,  $N_1 < N$  unidades são escolhidas aleatoriamente para receber o tratamento  $t$ , e o restante das unidades  $N_0 = N - N_1$  não recebem o tratamento, ou seja, estão no grupo de controle.

A segunda classe de mecanismos impõe uma restrição de que as probabilidades de participação não dependem dos resultados potenciais, ou seja, as probabilidades de participação não dependem mais das covariáveis. Mais à frente veremos que algumas hipóteses devem ser feitas para chegarmos a esse resultado. A terceira classe de mecanismos de atribuição contém o restante dos mecanismos com alguma dependência condicional com os resultados potenciais.

Como veremos mais a frente, os mecanismos de atribuição desempenham papel fundamental na estimação do impacto de um programa, pois conforme o seu tipo, diferentes abordagens devem ser consideradas



### 2.3.3 Interações e Equilíbrio Geral

Na maior parte da literatura assume-se que o tratamento recebido pela unidade  $i$  não possui efeito nenhum no resultado de outra unidade  $i'$ , ou seja, o tratamento recebido, potencialmente, afeta somente a unidade  $i$ . Essa hipótese é conhecida como SUTVA<sup>3</sup> (IMBENS, 2004, COLIN; TRIVEDI, 2005, WOOLDRIDGE, 2010).

Nas aplicações médicas, essa hipótese é bastante encontrada, entretanto, nas aplicações econômicas, as interações entre os agentes é muito importante. Esses efeitos de equilíbrio geral podem ou não ser bastante problemáticos. O efeito indireto, num indivíduo exposto ao tratamento, provavelmente será muito menor do que o efeito direto, entretanto, deve se observar em qual área está sendo aplicada a intervenção (IMBENS; WOOLDRIDGE, 2009).

## 2.4 EFEITOS MÉDIO DO TRATAMENTO

Utilizando o modelo causal de Rubin (RCM), iremos definir alguns dos principais efeitos médios na literatura econométrica (IMBENS, 2004, COLIN; TRIVEDI, 2005, WOOLDRIDGE, 2010). O mais popular dele é o efeito médio de tratamento da população (ATE). Então teremos

$$\tau_{ate} \equiv E(y(1) - y(0)) \quad (2.6)$$

Uma segunda quantidade de interesse é o efeito médio do tratamento sobre o tratado da população (ATT), que denotamos por  $\tau_{att}$ .

$$\tau_{att} \equiv E(y(1) - y(0)|t = 1) \quad (2.7)$$

Heckman e Robb (1985) e Heckman, Ichimura e Todd (1997) afirmam que em determinados contextos é mais interessante utilizar a subpopulação de tratados do que toda a população total, é o nosso caso nesse trabalho, veremos na aplicação econômica proposta, que estamos interessados no efeito do tratamento sobre os tratados.

Analogamente, utilizando uma amostra aleatória  $N$  (i.i.d) dessa população teremos o efeito medio do tratamento amostral. Logo

---

<sup>3</sup> Foi proposta por Rubin (1978).

$$\hat{\tau}_{ate} = \frac{1}{N} \sum_{i=1}^N (y_i(1) - y_i(0)) \quad (2.8)$$

E o efeito médio do tratamento sobre o tratado amostral é

$$\hat{\tau}_{att} = \frac{1}{N_t} \sum_{i: t_i=1} (y_i(1) - y_i(0)) \quad (2.9)$$

Onde  $N_t = \sum_{i=1}^N t_i$  é o número de unidades de tratados.

Podemos expandir a definição dos dois tratamentos condicionando-os as covariáveis.

Se  $X$  é uma covariável observada, o ATE condicionado a  $X$  é simplesmente

$$\bar{\tau}_{ate} = E(y(1) - y(0)|X) \quad (2.10)$$

Similarmente, teremos

$$\bar{\tau}_{att} = E(y(1) - y(0)|X, t = 1) \quad (2.11)$$

Como dito anteriormente, o problema fundamental da inferência causal só nos permite observar o resultado  $y(0)$  ou  $y(1)$ , não ambos para cada pessoa. A partir da equação (2.1), o resultado observável na população é

$$y = (1 - t)y(0) + ty(1) = y(0) + t(y(1) - y(0)) \quad (2.12)$$

## 2.5) O Problema do Viés de Seleção

Utilizando a abordagem dos resultados potenciais podemos definir o impacto do tratamento como  $d$ , através da equação abaixo

$$d = E(y(1)|t = 1) - E(y(0)|t = 0) \quad (2.13)$$

O problema é que provavelmente os grupos de tratamento e controle não são iguais antes da intervenção, portanto, a diferença média entre eles talvez não seja devido à intervenção. Utilizando a equação (2.13), podemos somar e subtrair  $E(y(0)|t = 1)$ , que

corresponde à esperança do resultado do grupo de controle caso eles tivessem participado. Então

$$d = E(y(1)|t = 1) - E(y(0)|t = 0) \pm E(y(0)|t = 1)$$

$$d = [E(y(1)|t = 1) - E(y(0)|t = 1)] + [E(y(0)|t = 1) - E(y(0)|t = 0)]$$

O primeiro termo  $[E(y(1)|t = 1) - E(y(0)|t = 1)]$  é o já conhecido Efeito Médio de Tratamento sobre o Tratado,  $\tau_{att}$ . Então

$$d = \tau_{att} + [E(y(0)|t = 1) - E(y(0)|t = 0)]$$

$$d = \tau_{att} + B \tag{2.14}$$

O termo  $B = E(y(0)|t = 1) - E(y(0)|t = 0)$  é conhecido como viés de seleção. O que esse termo nos diz é que, caso haja uma diferença nos resultados, entre o grupo de tratados e de controle, quando não há intervenção, ou seja,

$$E(y(0)|t = 1) \neq E(y(0)|t = 0) \rightarrow B \neq 0 \tag{2.15}$$

Então, a estimativa do impacto será viesada. Em outras palavras, para que possamos remover o viés de seleção devemos encontrar um contrafactual factível, de tal forma que

$$E(y(0)|t = 1) = E(y(0)|t = 0) \rightarrow B = 0 \tag{2.16}$$

Para Ravallion (2001), este viés pode ser dividido em dois tipos, um que decorre das diferenças nas características observáveis e o outro resultante das diferenças nos fatores que não conseguimos observar. O primeiro viés pode ocorrer por duas maneiras, a primeira, pode não existir um suporte<sup>4</sup> comum entre a amostra dos tratados e o grupo de comparação, a segunda, é de que mesmo que haja um suporte comum, a distribuição das características observáveis podem ser distintas nessa região comum, ou seja, os dados do grupo de controle não estão ponderados corretamente. Se os pesquisadores conseguirem escolher um grupo de controle de tal forma que a distribuição nas características observáveis seja igual no grupo de tratamento, esse tipo de viés pode ser eliminado. Já o segundo tipo surge quando, dado os valores de X, existe uma relação sistemática entre a participação no programa e os resultados,

---

<sup>4</sup> É o conjunto de variáveis de controle que são observadas para os resultados e a intervenção. Mais à frente veremos que  $\mathcal{X}$  é definido como suporte das covariáveis.

na ausência da intervenção. Ou seja, existem variáveis não observáveis que influenciam conjuntamente o resultado e o tratamento, condicionais às variáveis observáveis nos dados.

Para observarmos a questão do viés de seleção econometricamente, vamos utilizar a equação linear (2.5) descrita acima. Então,

$$y_i = X_i' \alpha + \beta t_i + \mu_i \quad (2.5)$$

Para estimarmos a equação (2.5) temos um problema. Segundo Khandker, Koolwal e Samad (2010), a participação no programa  $t_i$  geralmente não é aleatória, isso ocorre por dois motivos: o primeiro é a colocação proposital no tratamento e a segunda é a que os indivíduos decidem participar do programa. Isto é, as intervenções são posicionadas de acordo com as necessidades das comunidades e dos indivíduos, que em contrapartida, decidem se vão participar ou não, mediante programa proposto. Esta auto seleção pode ser baseada tanto em características observadas quanto fatores não observados ou ambos.

No caso dos fatores não observáveis, o termo  $\mu_i$  vai conter variáveis que também são correlacionadas com o tratamento  $t_i$ , ou seja,  $Cov(t_i, \mu_i) \neq 0$ . Essa é uma violação da hipótese básica da estimação por OLS que diz que deve existir independência entre os regressores e o termo do erro, ou seja, as estimativas desse modelo serão viesadas.

O objetivo final de qualquer pesquisador disposto a desenvolver uma avaliação de impacto é fazer com que esse viés de seleção seja eliminado ou mensurado. Com dito nesse capítulo, o problema fundamental é que não conseguimos observar o resultado dos tratados, caso eles não tivessem participado do programa, logo, temos um problema de falta de dados, que vamos contornar construindo um grupo contrafactual, a partir dos dados disponíveis. Esse grupo deve ser desenvolvido de tal forma que seja muito similar ao grupo de tratados, caso eles não tivessem participado, pois só assim será possível eliminar o viés de seleção. Diversos métodos podem ser utilizados para elaboração do grupo de controle, no próximo capítulo serão apresentados alguns deles.

### 3 METODOLOGIA

Conforme vimos no capítulo anterior, para se estimar efetivamente o impacto de um programa é necessário a utilização de métodos que contornem o problema do viés de seleção, mediante à construção de um contrafactual. Esse grupo de controle pode ser projetado através de dois tipos de experimentos<sup>5</sup>: o primeiro, experimentos aleatórios, onde a atribuição do tratamento é aleatória, e o segundo, experimentos não-aleatórios, onde a atribuição do tratamento não é aleatória, no segundo caso, a participação do indivíduo pode ser condicionada à variáveis observadas (seleção em observáveis), ou à variáveis não-observadas (seleção em não-observáveis). Como veremos na seção abaixo, nos experimentos não-aleatórios, utilizaremos os métodos cuja estimação é feita condicional à variáveis observáveis.

O método proposto nesse trabalho vai estimar o efeito médio do tratamento de duas formas distintas:

- a) Regressão com *Propensity Score* e
- b) *Propensity Score Matching*.

Além disso, na nossa metodologia desse trabalho, englobamos duas abordagens de estimação distintas, a primeira é não-paramétrica e a segunda é semi-paramétrica. Mas antes começaremos falando de experimentos aleatórios.

#### 3.1 EXPERIMENTOS ALEATÓRIOS

Nesse método o viés de seleção pode ser completamente eliminado quando a participação dos indivíduos no tratamento é atribuída de forma aleatória. Estatisticamente, ela é feita em dois estágios, primeiro, uma amostra de N indivíduos é retirada aleatoriamente da população de interesse (participantes potenciais), de tal maneira que possa representar muito bem a população para que haja validade externa; e segundo, os indivíduos presentes nessa amostra retirada da população, devem ter sua participação no tratamento atribuída de forma aleatória, de modo a garantir que haja validade interna. Para alguns autores, esse método é

---

<sup>5</sup> Mais precisamente, Shadish, Cook e Campbell (2001), definem experimento como um estudo no qual uma intervenção é deliberadamente introduzida para se observar os efeitos. Os dois principais são: a) Experimento Aleatório: é um experimento no qual as unidades recebem o tratamento, ou não, através de um processo aleatório; b) Quasi-Experimento: é um experimento no qual a atribuição de se participar, das unidades, não é feita de forma aleatória.

considerado o “padrão-ouro” para se estimar os impactos de algum programa (LALONDE, 1986).

Para calcularmos o impacto do programa, através desse método, essa amostra  $N$  será dividida aleatoriamente em dois grupos, o grupo dos tratados  $N_t$  e o grupo de controle  $N_c$ , tal que  $N = N_t + N_c$ . Assim garantimos que

$$(y_i(0), y_i(1)) \perp t \quad (3.1)$$

Ou seja, o tratamento é estatisticamente independente de  $(y_i(0), y_i(1))$ . Logo, podemos mostrar que o viés de seleção é eliminado utilizando a equação (2.14), onde  $d$  representa o impacto do da intervenção

$$d = \tau_{att} + [E(y(0)|t = 1) - E(y(0)|t = 0)] \quad (3.2)$$

Como dito anteriormente, o viés  $B$  é dado pelo segundo termo à direita

$$B = [E(y(0)|t = 1) - E(y(0)|t = 0)]$$

Mas como temos que  $y_i(0)$  e  $t$  são independentes, então

$$B = [E(y_i(0)) - E(y_i(0)|t)] = 0$$

Portanto conseguimos estimar efetivamente o impacto do programa, de tal forma que

$$d = \tau_{att}$$

Outra implicação da independência entre status do tratamento e os resultados potenciais é de que  $\tau_{ate}$  e  $\tau_{att}$  são idênticos, isto é

$$\tau_{ate} \equiv E(y(1) - y(0)) = \tau_{att} \equiv E(y(1) - y(0)|t = 1) \quad (3.3)$$

Usando a equação abaixo, temos

$$E(y|t = 1) = E(y(1)|t = 1) = E(y(1))$$

Onde a última igualdade ocorre porque  $y(1)$  e  $t$  são independentes. Similarmente,

$$E(y|t = 0) = E(y(0)|t = 0) = E(y(0))$$

Então

$$\tau_{ate} = \tau_{att} = E(y|t = 1) - E(y|t = 0) \quad (3.4)$$

O lado direito é facilmente estimado pela diferença das médias amostrais: a média da amostra de  $y$ , para as unidades tratadas, menos a média da amostra de  $y$ , para as unidades que não foram tratadas, e dessa forma, conseguimos obter o impacto do programa.

O processo de aleatorização demonstrado acima é conhecido como aleatorização pura, porém ele é raramente utilizado, ao invés disso, utiliza-se um processo chamado aleatorização parcial. Nesse caso, os indivíduos do grupo de tratamento e do grupo de controle são escolhidos aleatoriamente, condicionados à  $X$  características observáveis. Segundo Ravallion (2005), a aleatorização incondicional é virtualmente inconcebível para programas voltados para o combate à pobreza, pois o alvo dos *policy makers* vai ser sempre baseado em características observadas, como renda, dependentes, etc.

Na literatura econômica encontramos diversos estudos<sup>6</sup> que utilizam experimentos aleatórios para se estimar o impacto de programas ou políticas, e que abrangem os mais variados ramos da economia, tais como, economia da educação, economia da pobreza, economia comportamental e etc. Um dos artigos mais importantes foi publicado por Lalonde (1986), nele, o autor demonstra que alguns procedimentos econométricos para se estimar o impacto de um programa de treinamento para trabalhadores, apresentam estimativas significativamente diferentes das apresentadas quando estimadas por experimentos aleatórios. Já em Krueger (1999), o autor faz um experimento aleatório para avaliar estudantes que são distribuídos aleatoriamente em salas de aula de diferentes tamanhos. No artigo de Angrist et al. (2002), a aleatorização é utilizada para avaliar um programa colombiano, que através de uma loteria, distribuía *vouchers* que cobriam parcialmente os custos do ensino fundamental numa escola particular. Três anos após a loteria, foi constatado que o grupo tratado (ganhadores da loteria) tinham uma incidência menor de repetição de série e notas melhores

---

<sup>6</sup> Ver também Miguel e Kremer (2004) e Schultz (2001).

nos testes. Em Duflo (2006), a autora analisa a crescente utilização de experimentos aleatórios para testar teorias econômicas e hipóteses sobre economia comportamental, se limitando à três domínios: incentivos, aprendizado social e preferências temporais inconsistentes.

Para uma leitura mais completa sobre experimentos aleatórios ver Duflo e Kremer (2003) e Duflo, Glennester e Kremer (2008), sendo esse último um guia prático para utilização de experimentos aleatórios, na avaliação de políticas, inclusive cobrindo os diversos tipos de métodos para se fazer o processo<sup>7</sup> de aleatorização.

### 3.2 PROPENSITY SCORE MATCHING

Quando o mecanismo de participação no tratamento não pode ser aleatório, o melhor a se fazer é tentar imitar a aleatorização, ou seja, produzir um experimento análogo à aleatorização. Com métodos de pareamento, um pesquisador tenta desenvolver um contrafactual o mais parecido possível ao grupo de tratados, baseado nas características observáveis. O objetivo é, a partir de um grupo numeroso de pessoas que não participaram, achar entre eles, indivíduos similares aos participantes em termos das características observáveis. Cada participante é pareado com um não-participante similar, e então a diferença média dos resultados ao longo dos dois grupos é comparado para obtermos o efeito do tratamento do programa. Caso o pesquisador assume que a diferença entre participar e não participar baseia se somente nas características observáveis, e se existe um grupo grande o suficiente para parear com os participantes, então o efeito do tratamento pode ser calculado.

O pareamento é um procedimento que busca parear unidades tratadas com unidades de controle baseado nas suas características observáveis  $X$ . São dois tipos, exato e inexato. O *pareamento exato* só é possível se todas as covariáveis forem discretas e se a amostra pesquisada possui um número muito elevado de observações para cada  $x_i$ . Se o vetor de covariáveis possui uma dimensão alta, ou se possui covariáveis contínuas, esse método se torna impraticável, então usamos o método de *pareamento inexato*. Esse método consegue diminuir a dimensão de  $\mathbf{x}$  transformando o num escalar  $f(\mathbf{x})$ , no qual o pareamento é baseado (COLIN; TRIVEDI, 2005). O método de pareamento utilizado nesse trabalho é inexato, e é conhecido como *Propensity Score Matching* (PSM), nele, constrói-se estatisticamente um grupo de controle baseado nas características observáveis.

---

<sup>7</sup> São quatro tipos de métodos descritos em Duflo, Glennester e Kremer (2008): *oversubscription, randomized phase-in, within-group randomization e encouragement design*.



Como existem muitas variáveis observáveis possíveis, conseguir identificar de forma eficiente grupos similares se torna um problema na hora de parear os indivíduos. No PSM, cada participante é pareado com um não-participante baseado num único *propensity score*, que reflete a probabilidade de participação condicionada às diferentes características observáveis, mais a frente o *propensity score* será definido formalmente.

Como dito, o impacto do programa é calculado através da diferença média entre esses dois grupos. Para que a estimação seja feita algumas hipóteses devem ser satisfeitas, ou seja, quando  $t$  e  $(y(0), y(1))$  podem ser relacionados, precisamos de hipóteses que nos permitam identificar os efeitos do tratamento. O PSM só é válido se duas hipóteses são satisfeitas: independência condicional e *overlap*, além disso diferentes abordagens são utilizadas para parear os dois grupos (tratado e controle) com base no *propensity score*, são elas: *nearest-neighbor matching*, *caliper* ou *radius matching*, *stratification matching*, *kernel matching* e *genetic matching*.

### 3.2.1 O Propensity Score

O método PSM tenta capturar os efeitos que diversas covariáveis observáveis  $X$  tem sobre a participação do indivíduo, num único índice. Logo, indivíduos pertencentes ao grupo de tratados e indivíduos do grupo de controle que possuem o mesmo *propensity score* são pareados, e os resultados da variável de interesse (ex:renda) são comparados para que se obtenha o efeito do programa. Rosenbaum e Rubin (1983, 1984) mostraram que sob certas hipóteses, parear condicional a  $p(\mathbf{x})$  é tão bom quanto sob  $\mathbf{x}$ . Abaixo seguem as hipóteses introduzidas pelos autores.

**Hipótese.1** (Independência Condicional):

$$(y(0), y(1)) \perp t | \mathbf{x} \quad (3.5)$$

Essa hipótese é também chamada na literatura econométrica como, ignorabilidade e seleção nas observáveis. Ela diz que, condicionais a  $\mathbf{x}$ ,  $t$  e  $(y(0), y(1))$  são independentes. Essa hipótese é bastante controversa, pois assume que além das covariáveis observadas  $\mathbf{x}_i$ , não existem características não-observáveis do individuo associadas tanto com o tratamento, quanto com o resultado potencial.

**Hipótese.2 (*Overlap*):** Para todo  $X \in \mathfrak{X}$ , onde  $\mathfrak{X}$  é o suporte das covariáveis

$$0 < \Pr(t = 1|\mathbf{x}) < 1 \quad (3.6)$$

*Overlap* significa que, para qualquer cenário de covariáveis na suposta população, existe uma chance de haver unidades em ambos grupos, de controle e tratamento. Por exemplo, se tivéssemos  $\Pr(t = 1|X = x_0) = 0$ , então as unidades que possuíssem covariáveis com valor  $x_0$  nunca estariam no grupo de tratamento. A combinação dessas duas hipóteses é reconhecida como ignorabilidade forte.

Segundo Imbens (2004), existe uma controvérsia sobre a razoabilidade dessas duas hipóteses em modelagem econômica<sup>8</sup>, e portanto sobre a relevância da literatura econométrica, que sobre essas condições, realiza estimação e inferência para trabalhos empíricos. O autor em seguida oferece três argumentos para utilização delas.

O primeiro é estatístico. O ponto inicial de qualquer avaliação de programa é a comparação entre a média dos resultados das unidades tratadas com as unidades de controle. O próximo passo seria ajustar qualquer diferença nessa média referente às características exógenas (não afetadas pelo programa). Tal análise pode não levar à palavra final sobre a eficácia do programa, mas sua ausência traria uma dificuldade muito maior para tentar compreender as evidências acerca do efeito do tratamento. Segundo, quase toda avaliação de tratamento envolve a comparação de unidades que receberam o tratamento com unidades que não receberam. A questão principal não é sobre a realização da comparação, mas sim, quais unidades devem ser comparadas, ou seja, quais unidades melhor representam os tratados caso eles não tivessem recebido o tratamento. A hipótese de ignorabilidade se limita a afirmar que todas as variáveis que necessitam de ajustes são observadas pelo pesquisador. O terceiro argumento relevante é que mesmo que o agente escolha o seu tratamento otimamente, dois agentes com os mesmos valores para características observáveis podem diferir na escolha do tratamento sem invalidar a hipótese de ignorabilidade, caso as diferenças em suas escolhas sejam devido a características não-observáveis, não relacionadas com os resultados.

Utilizando as hipóteses (1) e (2) (ignorabilidade forte), os efeitos médios do tratamento podem ser identificados. Além dessas duas hipóteses podemos destacar também, que sob versões mais fracas dessas duas hipóteses, o efeito médio do tratamento nos tratados  $\tau_{att}$ , também pode ser identificado (WOOLDRIDGE, 2010). São elas :

<sup>8</sup> Ver Heckman (2005, 2010) e Heckman e Vytacil (2005). Nesse trabalhos, os autores fazem críticas concisas ao que eles chamam de método estatístico em contraposição ao método econométrico, defendido por eles.

**Hipótese.1'** Ignorabilidade na Média

$$E(y(0)|\mathbf{x}, t) = E(y(0)|\mathbf{x}) \quad (3.7)$$

e

$$E(y(1)|\mathbf{x}, t) = E(y(1)|\mathbf{x}) \quad (3.8)$$

A idéia por trás da hipótese (1') é de que se pudermos observar informação (contida em  $\mathbf{x}$ ) suficiente para determinar a participação no programa, então  $(y(0), y(1))$  pode ser independente na média de  $t$ , condicional a  $\mathbf{x}$ .

**Hipótese.2'** Para todo  $X \in \mathfrak{X}$ , onde  $\mathfrak{X}$  é o suporte das covariáveis

$$\Pr(t = 1|\mathbf{x}) < 1 \quad (3.9)$$

Como destacado anteriormente, quando a participação no tratamento não é aleatória, mas depende de um vetor de variáveis observáveis  $X$ , o conceito de *propensity score* (ROSEBAUM; RUBIN, 1983) se torna bastante útil, ele é definido como a probabilidade condicional de receber o tratamento dado suas características observáveis. Formalmente temos

$$p(\mathbf{x}) = P(t = 1|X = \mathbf{x}), \quad \mathbf{x} \in \mathfrak{X} \quad (3.10)$$

Um importante resultado construindo em cima da hipótese de ignorabilidade mostra que não é necessário condicionar os resultados possíveis simultaneamente a todas covariáveis. Todo viés<sup>9</sup> resultante das covariáveis observáveis pode ser removido condicionando somente ao *propensity score*. Rosenbaun e Rubin (1983) propuseram o seguinte.

Se  $p(\mathbf{x}_i)$  é o *propensity score*, então

$$X_i \perp t_i | p(\mathbf{x}_i) \quad (3.11)$$

A equação (3.11) nos diz que, condicional ao *propensity score*, as covariáveis são independente da participação no tratamento. Então, para observações com os mesmos *propensity scores*, a distribuição das covariáveis deve ser a mesma tanto no grupo de tratados

<sup>9</sup> A prova pode ser encontrada em Rubin e Rosenbaum (1983).

quanto no grupo de controle. Condicionando no *propensity score*, cada indivíduo possui a mesma probabilidade de participação no tratamento, como em um experimento aleatório (DEHEJIA; WAHBA, 1999, BECKER; ICHINO, 2002). Uma consequência direta da equação (3.11) é:

$$(y(0), y(1)) \perp t | p(\mathbf{x}) \quad (3.12)$$

Ou seja, os resultados potenciais são independentes do tratamento se condicionarmos ao *propensity score*, então teremos que

$$(y(0), y(1)) \perp t | \mathbf{x} \equiv (y(0), y(1)) \perp t | p(\mathbf{x}) \quad (3.13)$$

Utilizando o *propensity score* para identificarmos  $\tau_{ate}$  e  $\tau_{att}$  obtemos as seguintes expressões:

$$\tau_{ate}(\mathbf{x}) = E \left\{ \frac{[t - p(\mathbf{x})]y}{p(\mathbf{x})[1 - p(\mathbf{x})]} \middle| \mathbf{x} \right\} \quad (3.14)$$

Se a hipótese 2 é válida e as expectativas existem então teremos

$$\tau_{ate} = E \left\{ \frac{[t - p(\mathbf{x})]y}{p(\mathbf{x})[1 - p(\mathbf{x})]} \right\} \quad (3.15)$$

Analogamente,

$$\tau_{att} = E \left\{ \frac{[t - p(\mathbf{x})]y}{\delta[1 - p(\mathbf{x})]} \right\} \quad (3.16)$$

Onde  $\delta = P(t = 1)$ . Como dito anteriormente, se observamos as equações<sup>10</sup> (3.15) e (3.16) poderemos atestar que para identificarmos  $\tau_{ate}$  é necessário que  $0 < \Pr(t = 1 | \mathbf{x}) < 1$  e para identificarmos  $\tau_{att}$  basta somente que  $\Pr(t = 1 | \mathbf{x}) < 1$ . Em estudos observáveis, os grupos de tratamento e controle são geralmente, retirados de amostras diferentes (como veremos no próximo capítulo, é o caso da nossa aplicação econômica, o grupo de tratados é tirado do banco de dados NSW e os grupos de controle são retirados de duas bases de dados

<sup>10</sup> O desenvolvimento dessas expressões podem ser encontrado em Wooldridge (2010, p. 913).

distintas, PSID e CPS), logo o efeito de tratamento que buscamos identificar é o  $\tau_{att}$ , ou seja, é o efeito do tratamento sob a população tratada.

Estimando as equações (3.15) e (3.16), teremos:

$$\hat{\tau}_{ate} = n^{-1} \sum_{i=1}^n \frac{[t_i - \hat{p}(x)]y_i}{\hat{p}(x)[1 - \hat{p}(x)]} \quad (3.17)$$

$$\hat{\tau}_{att} = n^{-1} \sum_{i=1}^n \frac{[t_i - \hat{p}(x)]y_i}{\hat{\delta}[1 - \hat{p}(x)]} \quad (3.18)$$

Onde

$$\hat{\delta} = \frac{1}{n} \sum_{i=1}^n t_i \quad (3.19)$$

É um estimador consistente para  $\delta = P(t = 1)$ , e  $\hat{p}(x)$  pode ser estimado por uma função *logit* ou *probit*.

### 3.2.2 Matching (Pareamento)

Para Dehejia e Wahba (2002) e Cameron e Trivedi (2005), ao implementarmos o PSM três fatos são bastante relevantes: primeiro, se devemos parear com ou sem reposição; segundo, o número de unidades no grupo de controle; e terceiro, a escolha do método de parear. Parear sem reposição significa que qualquer observação no grupo que estamos comparando é pareada somente com um único tratado, enquanto com reposição, podemos obter diversos pareamentos. Caso o número de observações do grupo de controle seja baixo, parear sem reposição significa que os pareamentos podem não ser tão próximos em termos de  $p(x)$ , o que irá aumentar o viés do estimador. A escolha do número de casos no conjunto de comparação envolve um *trade-off* entre viés e variância, ou seja, ao utilizarmos um único pareamento bastante próximo do caso tratado, reduzimos o viés, ao passo que, ao incluirmos mais unidades de controle para parearmos, diminuímos a variância, mas aumentamos o viés.

Na nossa aplicação econômica no capítulo 4, serão usados três tipos de pareamento, pareamento simples (1:1), *caliper matching*, e *genetic matching*, além desses três, vamos

abordar dois métodos utilizados por Dehejia e Wahba (1999, 2002), autores do nosso artigo base, são eles, *nearest-neighbour matching* e *stratification matching*.

No método de **Nearest-Neighbor** (NN) cada unidade de tratamento é comparada com a unidade do grupo de controle com o *propensity score* mais próximo. Podem ser escolhidos  $n$  “vizinhos próximos” para se parear. Diversas formas variáveis do NN podem ser utilizadas, como com reposição e sem reposição. Com dito anteriormente, caso optemos por utilizar reposição, teremos um *trade-off* entre o viés e a variância, ou seja, se permitirmos reposição, a qualidade do pareamento vai aumentar e o viés vai diminuir. Segundo Caliendo e Kopeinig (2008), parear com reposição é muito indicado quando os dados possuem distribuições de *propensity score* muito diferentes entre o grupo de tratados e grupo de controle, pois, caso tenhamos muitos indivíduos tratados com elevados *propensity scores* contra poucos indivíduos no grupo de controle, obteríamos pareamentos ruins pois alguns participantes tratados seriam pareados com indivíduos não participantes que possuem baixos *propensity scores*, no entanto, com reposição esse problema seria suplantado.

Como o método NN corre o risco da distância entre o *propensity score* do vizinho mais próximo e do indivíduo tratado ser muito alta, resultando em pareamentos ineficientes, podemos contornar essa situação impondo uma restrição, ou seja, uma distância máxima entre os *propensity score*. No método **Caliper Matching** (CM), uma vizinhança é definida em termos de um raio em torno do  $p(x)$  das observações tratadas, e os pareamentos que ficam de fora dessa vizinhança são excluídos, ou seja, usam-se somente os melhores pareamentos. Esse processo envolve pareamento com reposição, somente entre os *propensity scores* situados num determinado alcance. Smith e Todd (2005) afirmam que um dos problemas de se utilizar essa metodologia é que dificilmente o pesquisador saberá a priori qual a distância tolerável, já Dehejia e Wahba (2002) sugerem uma variante desse método chamado de **Radius Matching**.

O terceiro método é chamado de **Stratification Matching**, cuja ideia é dividir o suporte comum do *propensity score* num conjunto de intervalos (estratos) e assim calcular o impacto em cada intervalo, ao tomar a diferença nos resultados entre as observações tratadas e as de controle. Para Imbens (2004), como todo viés na seleção das observáveis está associado com o *propensity score*, o uso de 5 estratos já remove quase totalmente o viés relativo às covariáveis.

Por último, temos o **Genetic Matching** (SEKHON, 2011), esse método automatiza o processo de se encontrar um bom pareamento, a ideia é utilizar um algoritmo de procura para encontrar o conjunto de pesos para cada covariável tal que o balanceamento ótimo seja alcançado (HO et al., 2011).

Como veremos no capítulo 4, uma característica importante do *propensity score matching* é de que após o pareamento entre os indivíduos do grupo de tratamento e grupo de controle, os indivíduos não pareados do grupo de controle são descartados e não são utilizados na estimação do impacto do programa (DEHEJIA; WAHBA, 2002).

### 3.3 ABORDAGEM NÃO -PARAMÉTRICA

O termo não-paramétrico é usado para se referir às técnicas estatísticas que não requerem a especificação da forma funcional do objeto sendo estimado. Em vez de presumir que a forma funcional do objeto é conhecida através de alguns parâmetros desconhecidos, propõe-se a utilização de hipóteses menos restritivas.

#### 3.3.1 Estimador Kernel de Densidade

Seja  $f(x)$  uma função contínua, e  $X$  um conjunto de variáveis contínuas, então o estimador kernel de densidade é dado por :

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \quad (3.20)$$

Onde  $K(\cdot)$  é chamada de função kernel, que pode assumir diversos formatos, e  $h$  é um número inteiro positivo chamado de *bandwith* (parâmetro de suavidade). Segundo Li e Racine (2007), qualquer função kernel  $k(\cdot)$  não negativa e limitada que satisfaça as seguintes condições

- (i)  $\int K(v)dv = 1$
- (ii)  $K(v) = K(-v)$
- (iii)  $\int v^2 K(v)dv = k_2 > 0$

é um estimador consistente de  $f(x)$ , ou seja,  $\hat{f}(x) \rightarrow f(x)$ . A prova dessa afirmação pode ser encontrada em Wand e Jones (1995), Pagan e Ullah (1999) e Li e Racine (2007).

Similarmente, para o caso multivariado, supondo que  $X_1, \dots, X_n$  é um vetor i.i.d. onde  $X_i \in \mathbb{R}^q$ , para  $q > 1$ , que possui a mesma fdp  $f(x) = f(x_1, x_2, \dots, x_q)$ . Seja  $X_{is}$  o  $s$ -ésimo componente de  $X_i (s = 1, \dots, q)$ , então usando uma função produto kernel, temos então

$$f(x) = \frac{1}{nh_1 h_2 \dots h_q} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \quad (3.21)$$

Onde

$$K\left(\frac{X_i - x}{h}\right) = k\left(\frac{X_{i1} - x_1}{h_1}\right) \times \dots \times k\left(\frac{X_{iq} - x_q}{h_q}\right)$$

E  $k(\cdot)$  é uma função kernel univariada como na primeira equação.

Nessa primeira definição, as variáveis utilizadas no método são contínuas, ou seja, não há espaço para utilização de variáveis discretas. Porém, sabe-se que para calcularmos o efeito médio de um tratamento utilizamos características observáveis, que podem ser tanto contínuas quanto discretas. A partir do artigo de Aitchison e Aitken (1976), alguns autores começaram a desenvolver uma abordagem não paramétrica que também contemplasse as covariáveis discretas, ver Li e Racine (2003), Hall, Racine e Li (2004), Racine e Li (2004).

### 3.3.2 Least Squares Cross Validation

Um componente essencial na estimação não-paramétrica é o método que utilizamos para encontrar as *bandwidths*. Essa etapa é muito importante pois diferentes *bandwidths* podem gerar resultados radicalmente diferentes (LI; RACINE, 2007). O método utilizado nesse trabalho é conhecido como *Least Squares Cross Validation* (LSCV), ele seleciona os parâmetros de suavidade  $h$ , e foi originalmente proposto por Rudemo (1982), Stone (1984) e Bowman (1984). Ele baseia-se no princípio de que devemos selecionar as *bandwidthsh* que minimizam o erro quadrático integrado da estimativa encontrada.

De acordo com Li e Racine (2007), temos que a diferença quadrática integrada entre  $\hat{f}$  e  $f$  é dada por:

$$\int [\hat{f}(x) - f(x)]^2 dx = \int \hat{f}(x)^2 dx - 2 \int \hat{f}(x)f(x) dx + \int f(x)^2 dx \quad (3.22)$$



Como o terceiro termo não contém o parâmetro  $h$ , então devemos minimizar a equação abaixo

$$\int \hat{f}(x)^2 dx - 2 \int \hat{f}(x)f(x)dx \quad (3.23)$$

O segundo termo  $\int \hat{f}(x)f(x)dx$  pode ser reescrito como  $E_X[\hat{f}(X)]$ , onde o  $E_X(\cdot)$  denota a esperança com respeito a  $X$  e não com respeito as observações aleatórias  $\{X_j\}_{j=1}^n$  utilizadas no cálculo de  $\hat{f}(x)$ . Logo, podemos estimar o segundo termo  $E_X[\hat{f}(X)]$  por  $n^{-1} \sum_{i=1}^n \hat{f}_{-i}(X_i)$ , onde

$$\hat{f}_{-i}(X_i) = \frac{1}{(n-1)h} \sum_{j=1, j \neq i}^n k\left(\frac{X_i - X_j}{h}\right) \quad (3.24)$$

É conhecido como estimador kernel *leave one out* de  $f(X_i)$ <sup>11</sup>. O primeiro termo  $\int \hat{f}(x)^2 dx$  pode ser estimado da maneira que vimos na seção anterior, ou seja:

$$\begin{aligned} \int \hat{f}(x)^2 dx &= \frac{1}{n^2 h^2} \sum_{i=1}^n \sum_{j=1}^n \int k\left(\frac{X_i - x}{h}\right) k\left(\frac{X_j - x}{h}\right) dx \\ \int \hat{f}(x)^2 dx &= \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1}^n \bar{k}\left(\frac{X_i - X_j}{h}\right) \end{aligned}$$

Onde  $\bar{k}(v) = \int k(u)k(v-u)du$ . Portanto, o método LSCV escolhe  $h$  que minimiza

$$CV_f(h) = \frac{1}{n^2 h} \sum_{i=1}^n \sum_{j=1}^n \bar{k}\left(\frac{X_i - X_j}{h}\right) - \frac{2}{n(n-1)h} \sum_{i=1}^n \sum_{j \neq i, j=1}^n k\left(\frac{X_i - X_j}{h}\right) \quad (3.25)$$

Generalizando para o caso multivariável, teremos :

$$CV_f(h_1, \dots, h_q) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \bar{K}_h(X_i, X_j) - \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i, j=1}^n K_h(X_i, X_j) \quad (3.26)$$

<sup>11</sup> Li e Racine (2007) enfatizam a utilização desse tipo de kernel para estimar  $E_X(\cdot)$  pois o operador presume que  $X$  e  $X_j$ 's são independentes entre si, caso contrário, o método de *Cross Validation* não irá funcionar.

Onde

$$K_h(X_i, X_j) = \prod_{s=1}^q h_s^{-1} k\left(\frac{X_{is} - X_{js}}{h_s}\right)$$

$$\bar{K}_h(X_i, X_j) = \prod_{s=1}^q h_s^{-1} \bar{k}\left(\frac{X_{is} - X_{js}}{h_s}\right)$$

Nas equações acima, as variáveis são contínuas, mas como veremos mais a frente, no caso de utilizarmos uma base de dados com variáveis discretas utilizaremos funções kernel próprias para esse tipo de variável. Para Hayfield e Racine (2008), a seleção da *bandwidth* deve ser considerada o aspecto mais importante da estimação não-paramétrica e semi-paramétrica que utilizam funções kernel.

### 3.3.3 Regressão Não-Paramétrica

Segundo Li e Racine (2007), considere o modelo de regressão abaixo:

$$Y_i = g(X_i) + u_i, \quad i = 1, \dots, n$$

Assumindo que  $(Y_i, X_i)$  são i.i.d., e  $g(\cdot)$  é desconhecida. Caso  $g(\cdot)$  seja uma função diferenciável, podemos estimá-la utilizando os métodos kernel expostos acima. Como na regressão paramétrica,  $g(x)$  é interpretada como a média condicional de  $Y$  dado  $X = x$ , ou seja,  $g(x) = E[Y_i | X_i = x]$ .

Sabemos que

$$E(Y|X = x) = \int y f_{y|x}(y|x) dy \quad (3.27)$$

Então

$$g(x) = E(Y|X = x) = \frac{\int y f_{y,x}(x, y) dy}{f(x)} \quad (3.28)$$

Onde  $f_{y,x}(x,y)$  é a função de densidade conjunta de  $(Y,X)$ <sup>12</sup>. A equação acima é facilmente estimada dado as definições no começo dessa seção. Para estimarmos  $g(x)$  podemos substituir  $f_{y,x}(x,y)dy$  pela estimativa  $\hat{f}_{y,x}(x,y)dy$ , onde

$$\hat{f}_{y,x}(x,y)dy = \frac{1}{nh_0h_1 \dots h_q} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) k\left(\frac{y - Y_i}{h_0}\right) \quad (3.29)$$

Onde  $K\left(\frac{X_i - x}{h}\right) = k\left(\frac{X_{i1} - x_1}{h_1}\right) \times \dots \times k\left(\frac{X_{iq} - x_q}{h_q}\right)$  e  $h_0$  é a *bandwidth* associada a  $Y$ .

Então, temos:

$$\begin{aligned} \int y \hat{f}_{y,x}(x,y) dy &= \frac{1}{nh_0h_1 \dots h_q} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \int y k\left(\frac{y - Y_i}{h_0}\right) dy \\ &= \frac{1}{nh_0h_1 \dots h_q} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \int (Y_i + h_0 v) k(v) h_0 dv \\ &= \frac{1}{nh_0h_1 \dots h_q} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) Y_i \end{aligned} \quad (3.30)$$

Onde  $v = \left(\frac{y - Y_i}{h_0}\right)$ ,  $\int k(v) dv = 1$  e  $\int vk(v) dv = 0$ . Esse desenvolvimento nos leva ao

Estimador Kernel Local Constant abaixo:

$$\hat{g}(x) = \frac{\int y \hat{f}_{y,x}(x,y) dy}{\hat{f}(x)} = \frac{\sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)} \quad (3.31)$$

Esse estimador kernel foi originalmente proposto por dois autores Nadaraya (1965) e Watson (1964), por isso, é conhecido também na literatura econométrica como **Estimador Nadaraya-Watson**. No caso de utilizarmos o método LSCV para obtermos as *bandwidths*  $(h_1, h_2, \dots, h_q)$ , teremos que minimizar o seguinte função objetiva:

$$CV_{ND}(h_1, \dots, h_q) = n^{-1} \sum_{i=1}^n (Y_i - \hat{g}_{-i}(X_i))^2 M(X_i)$$

<sup>12</sup> Essa transformação vem de  $f_{y|x}(x,y) = \frac{f_{y,x}(x,y)}{f(x)}$

onde

$$\hat{g}_{-i}(X_i) = \frac{\sum_{l \neq i}^n \frac{Y_l K(X_i - X_l)}{h}}{\sum_{l \neq i}^n \frac{K(X_i - X_l)}{h}}$$

é o estimador kernel *leave-one-out* de  $g(X_i)$  e  $0 \leq M(\cdot) \leq 1$  é uma função ponderada (LI; RACINE, 2007).

O estimador Nadaraya-Watson de  $g(x)$  exposto acima pode ser obtido resolvendo o seguinte problema de minimização (HOROWITZ, 2009) :

$$g_n(x) = \arg \min_b \sum_{i=1}^n (Y_i - b)^2 K\left(\frac{X_i - x}{h}\right) \quad (3.32)$$

Logo, o estimador NW pode ser interpretado como um método que aproxima  $g(X_i)$  por uma constante  $b$  quando  $X_i$  está próximo de  $x$ , por isso o nome Local Constant. Outra alternativa é aproximar  $g$  por uma função linear próxima de  $x$ . Então, temos que

$$g(X_i) \approx b_0 + b_1(X_i - x) \quad (3.33)$$

quando  $X_i$  está próximo de  $x$ . Análogo ao problema exposto na equação (3.32), escolhemos  $(b_0, b_1)$  que soluciona o seguinte problema

$$(\hat{b}_0, \hat{b}_1) = \arg \min_b \sum_{i=1}^n (Y_i - b_0 + b_1(X_i - x))^2 K\left(\frac{X_i - x}{h}\right) \quad (3.34)$$

tal que  $g(X_i)$  é estimada por  $\hat{g}_{LL}(x) = \hat{b}_0$ . Esse procedimento é conhecido como estimação local linear, e  $\hat{b}_1$  estima  $g'(X_i)$  quando ela existe. Formalmente, o estimador de  $g(X_i)$  é :

$$\hat{g}_{LL}(x) = \frac{\sum_{i=1}^n a_i Y_i}{\sum_{i=1}^n a_i} \quad (3.35)$$

onde

$$a_i = \frac{1}{nh} K\left(\frac{X_i - x}{h}\right) [s_{n2} - s_{n1}(X_i - x)] \quad (3.36)$$

e também

$$s_{nj} = \frac{1}{nh^3} \sum_i^n K\left(\frac{X_i - x}{h}\right) (X_i - x)^j; j = 1, 2. \quad (3.37)$$

onde,  $\hat{g}_{LL}$  é conhecido como **Estimador Local Linear**. Como no estimador NW, no estimador LL as *bandwidths* podem ser encontradas utilizando o método LSCV.

### 3.3.4 Propensity Score Não-Paramétrico

Esse método é exposto no artigo de Li, Racine e Wooldridge (2009). Agora, utilizaremos os dois tipos de variáveis, assim, utilizamos  $x_i^c$  e  $x_i^d$  para denotar os componentes contínuos e discretos de  $x_i$ , com  $x_i^c \in R^q$  e  $x_i^d$  sendo de dimensão  $r$ . Seja  $w(\cdot)$  a função kernel univariada para as variáveis contínuas e defina a função produto kernel como

$$W_h(x_i^c, x_j^c) = \prod_s^q h_s^{-1} w\left(\frac{x_{is}^c - x_{js}^c}{h_s}\right) \quad (3.38)$$

Onde  $x_{is}^c$  é o  $s$ -ésimo componente de  $x_i^c$ . Assumimos que algumas das variáveis discretas possuem uma ordem natural (ex: condições de saúde). Seja  $\tilde{x}_i^d$  um vetor  $r_1$  (tal que, os primeiros  $r_1$  componentes de  $x_i^d$  sejam de variáveis discretas que possuem uma ordem natural  $0 \leq r_1 \leq r$ ) e  $\bar{x}_i^d$  denote o restante  $r_2 = r - r_1$  variáveis que não possuem uma ordem. Usamos  $x_{it}^d$  para denotar o  $t$ -ésimo componente de  $x_i^d$  ( $t = 1, \dots, r$ ).

Para variáveis ordenadas, usamos o Kernel Habbema:

$$\tilde{l}(\tilde{x}_{it}^d, \tilde{x}_{jt}^d, \lambda_t) = \begin{cases} 1, & \text{se } \tilde{x}_{it}^d = \tilde{x}_{jt}^d \\ \lambda_t^{|\tilde{x}_{it}^d - \tilde{x}_{jt}^d|}, & \text{se } \tilde{x}_{it}^d \neq \tilde{x}_{jt}^d \end{cases} \quad (3.39)$$

Quando  $\lambda_t = 0$  ( $\lambda_t \in [0, 1]$ ),  $l(\tilde{x}_{it}^d, \tilde{x}_{jt}^d, \lambda_t = 0)$  se torna uma função indicadora, e quando  $\lambda_t = 1$ ,  $l(\tilde{x}_{it}^d, \tilde{x}_{jt}^d, \lambda_t = 1) = 1$  se torna uma função uniforme.

Para uma variável não-ordenada, é utilizado um função kernel baseada no artigo de Aitchison e Aitken (1976), definido por

$$\bar{l}(\bar{x}_{it}^d, \bar{x}_{jt}^d, \lambda_t) = \begin{cases} 1, & \text{se } \bar{x}_{it}^d = \bar{x}_{jt}^d \\ \lambda_t, & \text{se } \bar{x}_{it}^d \neq \bar{x}_{jt}^d \end{cases} \quad (3.40)$$

Novamente,  $\lambda_t = 0$  leva a uma função indicadora e  $\lambda_t = 1$  a função uniforme.

Seja  $\mathbf{1}(A)$  uma função indicadora que assume o valor 1 se  $A$  ocorre e 0 caso contrário.

Combinando as equações (3.39) e (3.40), obtemos a função kernel produto dado por

$$L(x_i^d, x_j^d, \lambda) = \left[ \prod_{t=1}^{r_1} \lambda_t^{|\bar{x}_{it}^d - \bar{x}_{jt}^d|} \right] \left[ \prod_{t=r_1+1}^r \lambda_t^{\mathbf{1}(\bar{x}_{it}^d \neq \bar{x}_{jt}^d)} \right] \quad (3.41)$$

Os autores recomendam a utilização de mínimos quadrados *cross-validation* (LSCV) para selecionar  $\lambda_t$  ( $t = 1, \dots, r$ ), não somente por causa do erro quadrático médio, mas também pela sua habilidade (assintótica) de remover as variáveis discretas irrelevantes. Assim, a *cross-validation* fornece uma maneira eficiente de proteger os modelos não-paramétricos da ocorrência de sobre-especificação, e portanto mitigando a maldição da dimensionalidade.

Como  $\mu(x_i) = Pr(t_i = 1|x_i) \equiv E(t_i|x_i)$ , podemos utilizar um estimador de probabilidade condicional ou um estimador de média condicional, para estima  $\mu(x_i)$ . Nesse método, o segundo foi a opção escolhida. E definimos  $\hat{e}(x_i)$  (*propensity score*) como o estimador não-paramétrico de  $\mu_i \equiv \mu(x_i)$  como

$$\hat{e}(x_i) = \frac{\sum_{j=1}^n t_j K_{n,ij}}{\sum_{j=1}^n K_{n,ij}} \quad (3.42)$$

Onde  $K_{n,ij} = W_h(x_i^c, x_j^c) L(x_i^d, x_j^d, \lambda)$ . Observando que  $var(t_i|x_i) = \mu_i(1 - \mu_i)$ , podemos estimar o efeito do tratamento médio por

$$\hat{t}_{ate.np} = \frac{1}{n} \sum_{i=1}^n \frac{(t_i - \hat{e}(x_i)) y_i M_{ni}}{\hat{e}(x_i)(1 - \hat{e}(x_i))} \equiv \frac{1}{n} \sum_{i=1}^n \left[ \frac{t_i y_i}{\hat{e}(x_i)} - \frac{(1 - t_i) y_i}{1 - \hat{e}(x_i)} \right] M_{ni} \quad (3.43)$$

Onde  $M_{ni} = M_n(x_i)$  é um conjunto de recorte que elimina observações perto da fronteira. Podemos observar que a equação (3.43) é similar à equação (3.17), a diferença é

que  $\hat{p}(x)$  é estimada parametricamente por um modelo *logit* ou *probit*, enquanto  $\hat{e}(x)$  é estimado não-parametricamente. Analogamente, tal qual equação (3.18), temos

$$\hat{\tau}_{att,np} = n^{-1} \sum_{i=1}^n \left[ \frac{[t_i - \hat{e}(x)]y_i}{\hat{\delta}[1 - \hat{e}(x)]} \right] M_{ni} \quad (3.44)$$

Para derivar a distribuição assintótica de  $\hat{\tau}$  são necessárias algumas hipóteses acerca da sua diferenciabilidade e também das suas derivadas parciais, para mais detalhes ver Robinson (1988).

Os parâmetros de suavidade serão escolhidos utilizando o método de LSCV exposto acima. O estimador kernel *leave-one-out* de  $E(y_i|x_i)$  é dado por:

$$\hat{t}_{-i}(x_i) = \frac{\sum_{j \neq i}^n t_j K_{n,ij}}{\sum_{j \neq i}^n K_{n,ij}}$$

Onde  $K_{n,ij} = W_h(x_i^c, x_j^c) L(x_i^d, x_j^d, \lambda)$ .

E escolhemos  $(h, \lambda) = (h_1, \dots, h_q, \lambda_1, \dots, \lambda_r)$  minimizando a seguinte função de mínimos quadrados *cross-validation* (LSCV)

$$CV(h, \lambda) = \frac{1}{n} \sum_{i=1}^n [t_i - \hat{e}_{-i}(x_i)]^2 M_n(x_i) \quad (3.45)$$

Um fato importante de se utilizar o LSCV para se obter as *bandwidths* é que, segundo Hall, Racine e Li (2004) e Hall, Li e Racine (2007) quando  $x_s^d$  e  $x_s^c$  são variáveis irrelevantes, o parâmetro de suavidade  $\lambda_s(h_s)$  da *cross-validation* irá convergir para  $\mathbf{1}(\infty)$  em probabilidade, logo, variáveis irrelevantes (discreta ou contínuas) serão automaticamente suavizadas.

As aplicações empíricas e os resultados das simulações presente em Hall, Racine e Li (2004) e Hall, Li e Racine (2007) revelam que a estimação não-paramétrica baseada no método de *cross-validation* tem uma performance melhor que o estimador frequencial convencional, pois o primeiro não dividi a amostra em células, o que causaria perda de eficiência. Tendo obtido os  $\hat{h}_s$ s e  $\hat{\lambda}_s$ s utilizando o método de *cross-validation*, nós podemos estimar  $\tau_{att}$  usando a equação (3.44) com  $\hat{t}(x_i)$  computado usando  $\hat{h}_s$ s e  $\hat{\lambda}_s$ s.

### 3.4 ABORDAGEM SEMI-PARAMÉTRICA

A estimação não-paramétrica da média condicional tem a capacidade de maximizar a flexibilidade e minimizar o risco de errarmos na especificação do modelo. Entretanto, segundo Horowitz (2009), o preço a se pagar por essa flexibilidade pode ser muito alto devido a algumas razões: primeira, a precisão da estimação decresce a medida que a dimensão de  $X$  aumenta (*curse of dimensionality*), Stone (1980) afirma que a taxa de convergência mais rápida possível do estimador  $E(Y|X = x)$  decresce a medida que adicionamos mais componentes contínuos em  $X$ ; segundo, a dificuldade na interpretação dos resultados obtidos quando  $X$  é multidimensional; terceiro, outro problema com estimação não-paramétrica é que não é possível extrapolar os dados, no caso de uma média condicional, não conseguimos obter previsões de  $E(Y|X = x)$  nos pontos  $x$  que estão situados fora do suporte das variáveis aleatórias  $X$ ; e por último, ao estimarmos um modelo não-parametricamente, pode ser bastante difícil impor restrições sugeridas pela teoria econômica.

Uma maneira de contornar os problemas citados acima é fazermos uso de modelos semi-paramétricos. Como dito na introdução do trabalho, modelos semi-paramétricos conciliam modelagem paramétrica e modelagem não-paramétrica, de tal forma que unem a flexibilidade obtida na estimação não-paramétrica e a precisão dos modelos paramétricos. As hipóteses feitas acerca da forma funcional são mais fortes do que no caso não-paramétrico, mas menos restritivas do que em modelos paramétricos, logo, reduzem a possibilidade de erro na especificação do modelo. Além disso, segundo Hardle e Linton (1994), quando os dados possuem uma dimensão elevada, ou é necessário levar em conta a forma funcional e uma correlação natural, métodos não-paramétricos podem não obter uma boa performance.

Para esse trabalho será usado um tipo especial de modelo semi-paramétrico conhecido como Modelo Aditivo Generalizado, mais conhecido como GAM (*Generalized Additive Models*), inicialmente proposto por Hastie e Tibshirani (1986, 1990).

#### 3.4.1 GAM

Um modelo GAM tem a forma proposta abaixo

$$G[\mu(X)] = \alpha + f_1(X_1) + f_2(X_2) + \dots + f_p(X_p) \quad (3.46)$$



A abordagem desenvolvida por Hastie e Tibshirani (1986, 1990) estima cada função  $f_j$  com algum tipo de estimador não-paramétrico e usa um algoritmo *Backfitting* para estimar simultaneamente todas as  $p$  funções. Onde a média condicional  $\mu(X)$  de  $Y$  se relaciona com uma função aditiva de regressores através de  $G(\cdot)$ , conhecida como função link. Essa função  $G(\cdot)$  pode assumir diversas formas, mas no nosso caso, estamos interessados em apenas duas, quando temos  $G(\mu) = \text{logit}(\mu)$  ou  $G(\mu) = \text{probit}(\mu)$ .

Segundo Horowitz (2009), nesse modelos, a estimação dos componentes aditivos  $f_j$  alcançaria a mesma precisão assintótica caso todos os componentes fossem conhecidos e fosse necessário estimar um único componente com um argumento escalar, essa propriedade é conhecida como eficiência oráculo.

Na literatura econométrica mais recente, são encontrados três tipos de procedimentos para se estimar os modelos GAM, o método de Integração Marginal (ver Linton e Nielsen, 1995), o método Two-Step (HOROWITZ; MAMMEN, 2004), e por último, o método utilizado nesse trabalho, Algoritmo *Backfitting*. Esse método foi primeiramente proposto por Buja, Hastie e Tibshirani (1989) e posteriormente foi desenvolvido por Hastie e Tibshirani (1990), para uma leitura complementar ver também Ruppert, Wand e Carroll (2003), Hastie, Tibshirani e Friedman (2009) e Horowitz (2009)

### 3.4.2 Estimadores *Spline*

No algoritmo *Backfitting* desenvolvido por Hastie e Tibshirani (1986, 1990) para estimação do GAM, a parte não-paramétrica do modelo pode ser estimada utilizando qualquer técnica de suavização. Nesse trabalho, o estimador utilizado é o *spline*.

Segundo Hardle e Linton (1994), o estimador *spline* cúbico  $\hat{m}_\lambda$  é o único minimizador da função abaixo

$$R_\lambda(\hat{m}, m) = \sum_{i=1}^n [Y_i - \hat{m}(X_i)]^2 + \lambda \int [\hat{m}''(v)]^2 dv \quad (3.47)$$

O parâmetro de suavização  $\lambda$  controla o grau de suavidade do estimador  $\hat{m}_\lambda$ . A medida que  $\lambda \rightarrow 0$ ,  $\hat{m}_\lambda$  interpola as observações, enquanto se  $\lambda \rightarrow \infty$ ,  $\hat{m}_\lambda$  tende para uma linha de OLS. Silverman (1984) mostra que estimadores *splines* são equivalentes assintoticamente à estimadores kernel, onde o *splines* equivale à um kernel de quarta ordem abaixo:

$$K(v) = \frac{1}{2} \exp\left(-\frac{|v|}{\sqrt{2}}\right) \sin\left(\frac{|v|}{\sqrt{2}} + \frac{\pi}{4}\right) \quad (3.48)$$

E a *bandwidth* equivalente é igual a

$$h(\lambda, X_i) = \lambda^{1/4} n^{-1/4} f(X_i)^{-1/4} \quad (3.49)$$

### 3.4.3 Propensity Score Semi-Paramétrico

Para estimarmos o *propensity score* semi-paramétrico utilizaremos a equação (3.46)

$$G[\mu(X)] = \alpha + f_1(X_1) + f_2(X_2) + \dots + f_p(X_p) \quad (3.46)$$

De tal maneira que a função link  $G(\cdot)$  assuma duas formas possíveis, ou seja,  $G(\mu) = \text{logit}(\mu)$  ou  $G(\mu) = \text{probit}(\mu)$ , como dito anteriormente. Utilizando o *logit* como exemplo (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

$$\log\left(\frac{\mu(X)}{1 - \mu(X)}\right) = \alpha + f_1(X_1) + \dots + f_p(X_p)$$

Onde,  $\mu(X) = \Pr(Y = 1|X)$ , e cada  $f_j$  pode ser estimado de forma não-paramétrica, o que no nosso caso é utilizado o estimador *spline*. No capítulo 4, utilizaremos os métodos expostos nesse capítulo para estimação do efeito médio do tratamento sobre os tratados, ou  $\tau_{att}$ .

## 4 APLICAÇÃO ECONÔMICA

Como dito no primeiro capítulo, a aplicação econômica será feita a partir de um programa de treinamento para trabalhadores (*job training*) realizado na década de 1970 pelo governo norte americano.

### 4.1 INTRODUÇÃO

Os dados foram obtidos do NSW (*National Supported Work Demonstration*). O NSW foi um programa federal implementado na década de setenta que possuía o intuito de fornecer experiência de trabalho, por um período de 6 a 18 meses, para indivíduos que enfrentavam problemas financeiros e sociais. Os participantes do programa, escolhidos de forma aleatória, eram recrutados para trabalhar em diversos tipos de trabalho. Tanto o grupo de tratados quanto o grupo de controle participaram de avaliações em intervalos específicos. No artigo original de Lalonde (1986), a análise do tratamento do participantes é dividida entre homens e mulheres, já em Dehejia e Wahba (1999, 2002), a análise é restrita somente aos homens. Nesse trabalho, focamos somente no grupo de participantes homens também.

### 4.2 BASE DE DADOS<sup>13</sup>

Durante a década de 70 a *Manpower Demonstration Research Corporation* (MDRC) era responsável pelo NSW. Os indivíduos escolhidos para o grupo de tratados tinham garantido de 9 a 18 meses de trabalho. O grupo de tratamento foi dividido em grupos de 3 a 5 pessoas que trabalhavam juntos e se encontravam regularmente com um conselheiro da NSW para discutir suas performances. A NSW pagava aos tratados por seu trabalho.

Os salários recebidos por eles eram menores do que os recebidos nos trabalhos regulares, mas as remunerações cresciam a medida que suas performances melhorassem. Era permitido que os *trainees* ficassem nesse emprego no máximo 18 meses, quando expirava o período de treinamento, e então, eram forçados a procurar empregos normais. Os tipos de trabalhos exercidos pelo participantes variavam de acordo com a localização geográfica e disposição. Os candidatos elegíveis para NSW foram escolhidos aleatoriamente para participação no programa entre março de 1975 e julho de 1977, escolhidos para participar do treinamento recebiam todos os benefícios do NSW, ao passo que os não participantes não.

<sup>13</sup> Os dados utilizados nesse trabalho foram obtidos através do site <<http://users.nber.org/~rdehejia/data/>>.

Tanto o grupo de tratados, quanto o grupo de controle participaram das entrevistas, seguintes à participação, em intervalos específicos de tempo.

A aleatorização ocorreu entre março de 1975 e julho de 1977. No seu artigo, Lalonde (1986) limitou-se a utilizar indivíduos que participaram do programa após dezembro de 1975 com intuito de assegurar que os rendimentos referentes ao ano de 1975 fossem utilizados como os rendimentos pré-intervenção, da mesma forma, limitou-se a utilizar indivíduos que não estavam mais participando do programa a partir de janeiro de 1978, e assim utilizou os rendimentos de 1978 como variável de interesse Y. Sendo assim, sua amostra contava com 297 indivíduos no grupo de tratados e 425 indivíduos no grupo de controle. Já Dehejia e Wahba (1999, 2002) utilizaram um subconjunto<sup>14</sup> dos dados de Lalonde, limitando a utilizar indivíduos cujos rendimentos de 1974 podem ser obtidos, ou seja, indivíduos que começaram a participar cedo o bastante do programa para terem informações sobre o ano de 1974. Esse subconjunto contém 185 indivíduos no grupo de tratados e 260 no grupo de controle. Para estimação semi-paramétrica e não-paramétrica iremos utilizar o subconjunto utilizado por Dehejia e Wahba.

Tabela 1- Médias Amostrais das Características

	Nº de Observações	Idade	Educação	Negro	Hispânico	Sem Diploma	Casado	RE74	RE75
<b>NSW/Lalonde</b>									
Tratados	297	24,63	10,38	0,8	0,09	0,73	0,17		\$ 3.066,00
Controle	425	24,45	10,19	0,8	0,11	0,81	0,16		\$ 3.026,00
<b>NSW/ Dehejia e Wahba</b>									
Tratados	185	25,81	10,35	0,84	0,059	0,71	0,19	\$ 2.096,00	\$ 1.532,00
Controle	260	25,05	10,09	0,83	0,1	0,83	0,15	\$ 2.107,00	\$ 1.267,00
<b>Grupos de Controle</b>									
PSID - 1	2490	34,85	12,11	0,25	0,03	0,31	0,87	\$19.429,00	\$19.063,00
CPS - 1	15992	33,22	12,02	0,07	0,07	0,29	0,71	\$14.016,00	\$13.650,00

Fonte: Dehejia e Wahba (1999)

Na tabela 1 acima, os grupos de controle utilizados foram retirados de dois grupos de comparação: *Panel Study of Income Dynamics* (PSID-1) e *Westat's Matched Current Populations Survey-Social Security* (CPS-1). Além disso, Lalonde (1986) criou subconjuntos desses grupos de controle (CPS-2, CPS-3, PSID-2 e PSID-3), mas nesse trabalho iremos nos ater somente aos grupos PSID-1 e CPS-1. Como iremos utilizar na estimação a base de dados

<sup>14</sup> Segundo Ashenfelter (1978), Ashenfelter e Card (1985), Card e Sullivan (1988) é muito importante observar alguns anos pré-intervenção para podermos identificar melhor o efeito de um treinamento.

utilizada por Dehejia e Wahba (subconjunto criado a partir dos dados do Lalonde), então, na tabela 2 abaixo apresentamos as estatísticas das características pré-intervenção dos grupos de controle PSID-1 e CPS-1 comparadas a nossa amostra de tratados, e também suas definições.

Tabela 2 - Comparação DW, PSID e CPS

Variável	Definição	Tratados	Controle (PSID <sup>15</sup> )	Controle (CPS <sup>16</sup> )
Idade	Idade em anos	25,82	34,85	33,23
Educação	Educação em anos	10,35	12,12	12,03
Negro	1 se for negro e 0 c.c.	0,8432	0,2506	0,07354
Hispânico	1 se for hispânico e 0 c.c.	0,05946	0,03253	0,07204
Sem Diploma	1 se educação >12 e 0 c.c.	0,7081	0,3052	0,2958
Casado	1 se for casado e 0 c.c.	0,1892	0,8663	0,7117
U74	1 se estiver desempregado em 1974 e 0 c.c.	0,7081	0,1196	0,3052
U75	1 se estiver desempregado em 1975 e 0 c.c.	0,6	0,1093	0,1
RE74	Rendimentos Reais em 1974	\$ 2.096,00	\$ 19.429,00	\$ 14.017,00
RE75	Rendimentos Reais em 1975	\$ 1.532,00	\$ 19.063,00	\$ 13.651,00
RE78	Rendimentos Reais em 1978	\$ 6.349,10	\$ 21.554,00	\$ 14.847,00
Nº de Observações		185	2490	15992

Fonte: Dehejia e Wahba (1999).

Na tabela acima, fica evidente a diferença entre o grupo de tratamento e os grupos de comparação PSID e CPS, em relação às características pré-intervenção. Em relação a variável *idade*, o grupo de tratados é mais jovem do que os dois grupos de controle, sendo os dois grupos de controle similares nesse quesito.

Ao analisarmos a variável *educação*, que nos indica a quantidade de anos de estudo obtido pelo indivíduo, observamos que a média dos três grupos está bem próxima, em torno de 12 anos, entretanto, em relação à variável *sem diploma* o panorama muda bastante, enquanto no grupo de tratados 70% possui diploma, nos dois outros grupos de controle, apenas aproximadamente, 30% possuem.

Quanto a etnia, se olharmos a presença de negros, a composição dos grupos é bem distinta, enquanto no grupo de tratados 84% dos indivíduos são negros, no grupo PSID, somente 25% dos indivíduos são e no grupo CPS esse número cai mais ainda, com a presença de 7% de pessoas que se declararam negras. Em relação aos hispânicos, todas as amostras apresentam baixa porcentagem deles.

<sup>15</sup> O grupo PSID é definido como grupo de homens com idade menor que 55 anos que afirmaram não ser aposentados, presentes no *Panel Study of Income Dynamics*.

<sup>16</sup> O grupo CPS representa todos os homens presentes no *Westat's Matched Current Populations Survey-Social Security* (CPS-1) com idade menor do que 55 anos.

A maior discrepância no entanto, é encontrada ao observarmos as variáveis *RE74*, *RE75* e *RE78*, pois na média, os rendimentos reais nesses anos são muito maiores no grupo de controle do que no grupo de tratados, chegando a ser 300% maiores. Esses fatos evidenciam de forma clara a distinção entre os grupos de controle o grupo de tratado, e por isso a necessidade de utilizarmos métodos estatísticos para encontramos no interior desses dois grupos, PSID e CPS, indivíduos compatíveis com os tratados.

As variáveis *U74* e *U75* foram criadas a partir dos dados e são utilizados por Dehejia e Wahba (1999) na estimação do efeito médio de tratamento.

#### 4.3 RESULTADOS EXPERIMENTAIS

Como mostrado na tabela 1, a amostra original contruída por Lalonde (1986) possui 722 observações, sendo 297 indivíduos no grupo de tratados e 425 no grupo de controle. O efeito estimado do treinamento é medido em relação a variável *RE78*, ou seja, em cima dos rendimentos do ano de 1978. Na tabela abaixo apresentamos os resultados originais obtidos por Lalonde.

Tabela 3 - Efeitos Estimados do Treinamento - Amostra Original de Lalonde (1986)

Grupo de Comparação	Efeito Estimado do Treinamento	
	Não -Ajustado	Ajustado
NSW	886 [472]	798 [472]
PSID	-15578 [913]	-8067 [990]
CPS	-8870 [562]	-4416 [577]

Fonte: Lalonde (1986)

Como podemos observar o efeito do treinamento sobre a renda em 1978 é de 886 dólares americanos<sup>17</sup>, já em relação aos outros grupos de controle, as estimativas para o efeito do treinamento já são muito ruins. Como a participação no tratamento se deu de forma aleatória, esses resultados apresentados na coluna 2 são fáceis de se obter aplicando uma diferença entre as médias dos grupos.

Como dito na seção anterior, na amostra original de Lalonde, não existe a variável rendimentos anuais em 1974 (*RE74*), ela foi criada por Dehejia e Wahba no artigo de 1999.

<sup>17</sup> As estimativas estão em dólares de 1982.

Na tabela 4 abaixo, apresentamos os efeitos estimados do treinamento baseados no subconjunto construído a partir da inclusão da variável RE74 (foram excluídos da amostra original de Lalonde indivíduos que não possuíam renda em 1974). Essa amostra possui 445 observações, sendo 185 indivíduos no grupo de tratamento e 260 indivíduos nos grupos de controle.

Tabela 4. Efeitos Estimados do Treinamento - Amostra de Dehejia e Wahba (1999)

Grupo de Comparação	Efeito Estimado do Treinamento	
	Não -Ajustado	Ajustado
NSW	1794 [633]	1688 [636]
PSID	-15205 [1155]	-879 931]
CPS	-8498 [712]	-8 [572]

Fonte: Dehejia e Wahba (1999)

A partir da tabela acima podemos observar que o efeito estimado do treinamento, dado o subconjunto utilizado, é de US\$ 1.974. Comparando as tabelas 3 e 4 observamos que o efeito do treinamento é maior quando utilizamos o subconjunto da amostra, ele muda de US\$ 886 para US\$ 1794, isso ocorre devido às diferenças na composição das duas amostras. De fato, ocorre que se obtém um efeito do treinamento maior para os indivíduos que ingressaram mais cedo no treinamento ou que se encontravam desempregados antes da intervenção (DEHEJIA; WAHBA, 1999). Para efeitos de comparação, durante a nossa estimação semi-paramétrica e não-paramétrica utilizaremos o efeito do treinamento estimado por Dehejia e Wahba, no valor de US\$ 1794, como *benchmark*.

#### 4.4 RESULTADOS NÃO-EXPERIMENTAIS

No trabalho de Lalonde (1986), os valores de US\$ 1.974 e US\$ 1.688 são obtidos quando se compara o grupo de tratados e grupo de controle dentro da amostra da **NSW/ Dehejia e Wahba** (ver tabela 1), já nos resultados não-experimentais duas novas amostras são criadas, uma que contém o grupo de tratados presentes na amostra NSW (185 observações) e grupo de controle PSID (2490 observações) e a outra com o mesmo grupo de tratados mas com o grupo de controle CPS (15992 observações). A estimação realizada por Dehejia e Wahba (1999) é feita em dois passos, no primeiro passo, o *propensity score* é estimado para cada amostra não-

experimental utilizando um modelo *logit*, já no segundo passo, estimado  $\hat{p}(x)$ , usa-se os métodos de estratificação e nearest-neighbor para encontrar o efeito do treinamento<sup>18</sup>. Abaixo, na tabela 5, encontram-se os resultados obtidos pelos autores.

Tabela 5 - Efeitos Estimados do Treinamento *Propensity Score* (Dehejia e Wahba)

Amostra	Efeito Estimado do Treinamento (US\$)				
	Estratificação no $p(x)$			NN no $p(x)$	
	Não -Ajustado	Ajustado	Observações Utilizadas	Não -Ajustado	Ajustado
NSW-PSID	1608	1494	1255	1691	1473
	[1571]	[1581]		[2209]	[809]
NSW-CPS	1713	1774	4117	1582	1616
	[1115]	[1152]		[1069]	[751]

Fonte: Dehejia e Wahba (1999).

Podemos observar através da tabela acima que os resultados obtidos ficam bem próximos aos obtidos por Lalonde, variam de 1494 até 1774 dólares. Dessa forma, Dehejia e Wahba (1999, 2002), introduziram na literatura econométrica uma método de estimação não-experimental que consegue efetivamente “imitar” um experimento aleatório.

#### 4.5 ESTIMAÇÃO NÃO-PARAMÉTRICA<sup>19</sup>

Na abordagem não-paramétrica, a estimação do efeito do treinamento é feita em 3 passos, no primeiro são encontradas as *bandwidths* através do método *Least Squares Cross Validation*, no segundo passo, a partir de Li, Racine e Wooldridge (2009) calculamos o *propensity score* utilizando a equação (3.42) abaixo

$$\hat{e}(x_i) = \frac{\sum_{j=1}^n t_j K_{n,ij}}{\sum_{j=1}^n K_{n,ij}} \quad (3.42)$$

Após encontrarmos o vetor  $\vec{e}$  de *propensity score* não-paramétricos, aplicamos o terceiro passo, que consiste em escolher um modelo para estimarmos o efeito do treinamento sobre os tratados. No nosso trabalho, para encontrarmos o ATT, o *propensity score* vai ser utilizado de duas formas distintas. No primeiro método utilizaremos a equação (3.44), uma

<sup>18</sup> Para mais detalhes ver Dehejia e Wahba (2002).

<sup>19</sup> Para estimação foi utilizado o pacote “np” do software R, desenvolvido por Hayfield e Racine (2008).



regressão com o *propensity score*, para estimarmos o impacto do treinamento sobre os tratados de modo que :

$$\hat{\tau}_{att,np} = n^{-1} \sum_{i=1}^n \left[ \frac{[t_i - \hat{e}(x)]y_i}{\hat{\delta}[1 - \hat{e}(x)]} \right] M_{ni} \quad (3.44)$$

No segundo método vamos unir a estimação não-paramétrica com a utilização de métodos de pareamento. Os métodos escolhidos foram três: Pareamento Simples (1:1), *Caliper* e Genético. Na estimação não-paramétrica, para as variáveis contínuas foi utilizado o Kernel Gaussiano e para as variáveis discretas (não ordenadas), utilizamos o Kernel Li e Racine exposto na equação (3.41).

#### 4.5.1 Regressão com *Propensity Score* Não-Paramétrico

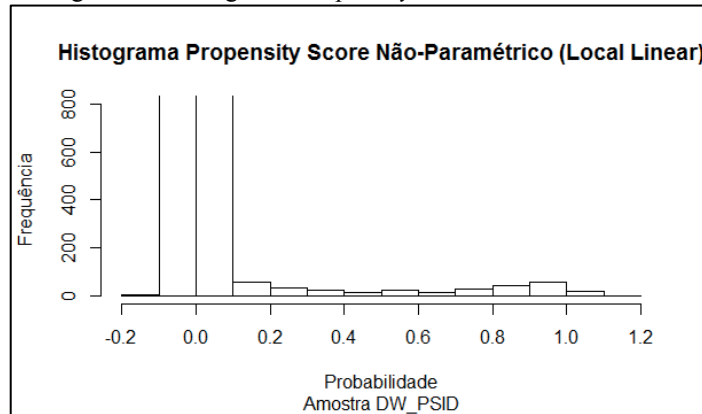
Nesse método, além da equação (3.21), utilizamos a restrição [ $\hat{e}(x) < 0,5$  e  $\hat{e}(x) > 0,95$ ] para eliminarmos estimativas próximas de 0 e 1 (WOOLDRIDGE, 2010). Os resultados obtidos aplicados as duas amostras DW\_PSID e DW\_CPS estão na tabela abaixo.

Tabela 6 - Efeitos Estimados do Treinamento (*Propensity Score* Não-Paramétrico)

Regressão com Propensity Score Não-Paramétrico			
Grupo	Nº de Observações	$\hat{\tau}_{att}(\text{US\$})$	
		NW	LL
DW_PSID	2675	4138 (286)	3547 (328)
DW_CPS	16177	3757 (482)	3792 (508)

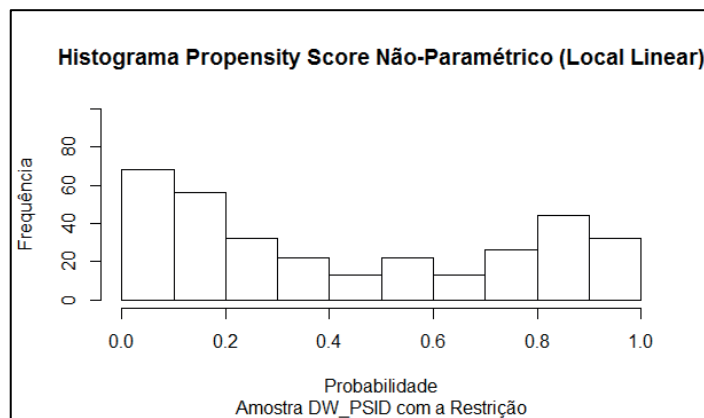
Fonte: elaborado pelo autor (2015).

Entre parênteses se encontram o número de observações após aplicarmos [ $\hat{e}(x) < 0,5$  e  $\hat{e}(x) > 0,95$ ]. Analisando os resultados, a implementação do método desenvolvido por Li, Racine e Wooldridge (2009) gerou estimativas muito elevadas, além disso, o número de observações caiu drasticamente ao aplicarmos essa restrição.

Figura 1 - Histograma *Propensity Score* Não-Paramétrico

Fonte: elaborado pelo autor (2015).

Olhando as figuras 1 e 2 podemos observar a frequência do *propensity score* estimado para a amostra DW\_PSID utilizando local linear. Na figura 1 podemos observar a distribuição dos *propensity scores* não-paramétricos antes da restrição [ $\hat{e}(x) < 0,5$  e  $\hat{e}(x) > 0,95$ ], e é nítida a grande concentração em torno da  $p(x) = 0$ , algumas até negativas. Ao aplicarmos a restrição, geramos a figura abaixo:

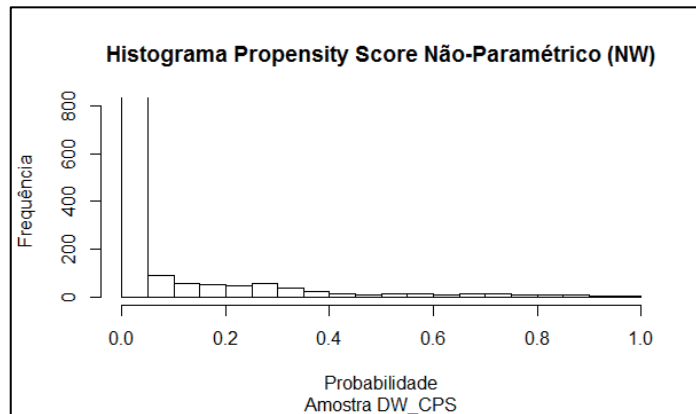
Figura 2 - Histograma *Propensity Score* Não-Paramétrico (pós restrição)

Fonte: elaborado pelo autor (2015).

A restrição imposta melhorou bastante a distribuição do propensity score estimado, mas não foi suficiente para gerar uma boa estimativa para o impacto do treinamento, pois continuamos obtendo a maioria das estimativas perto da fronteira inferior. Nas figuras 1 e 2 conseguimos evidenciar como o grupo de controle PSID (2490 observações) é muito diferente do grupos de tratados retirado da amostra de Lalonde (185 observações), pois o número grande de ocorrências de probabilidade de participação próximos ao limite inferior significa que dentro desse grupo encontramos muitos indivíduos que não são bons pares para os tratados, ou seja, não conseguimos construir um contrafactual crível e consequentemente, a

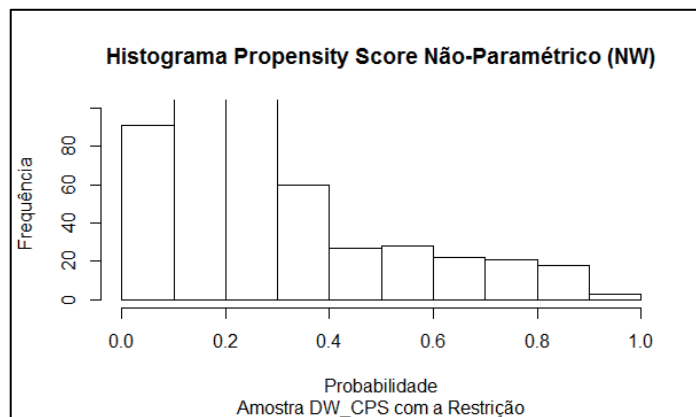
estimação do efeito que o treinamento provocou foi prejudicada. Esse resultado se repete para o três modelos estimados restantes, todos apresentam gráficos semelhantes às figuras 1 e 2 . Abaixo nas figuras 3 e 4 podemos observar o histograma do *propensity score* estimado por Nadaraya-Watson para a amostra DW\_CPS.

Figura 3 - Histograma *Propensity Score* Não-Paramétrico (Nadaraya - Watson)



Fonte: elaborado pelo autor (2015).

Figura 4 - Histograma *Propensity Score* Não-Paramétrico (pós-restrição/Nadaraya - Watson)



Fonte: elaborado pelo autor (2015).

Ao olharmos os resultados nas colunas 3 e 4, na tabela 6, o método proposto, não se mostrou um bom estimador do impacto do treinamento, as estimativas ficaram muito acima do que se esperava, algo em torno do *benchmark* US\$ 1794, encontrado por Dehejia e Wahba (1999, 2002), entretanto, devemos ter em mente que ao estimarmos o *propensity score* dessa maneira, revelamos uma face dos dois grupos de controle, que grande parte dos indivíduos presente neles não são bons pares para os indivíduos que receberam o tratamento. Vamos ver na próxima seção que uma estratégia melhor é estimarmos  $\hat{p}(x)$  e depois aplicarmos um método de pareamento.

#### 4.5.2 PSM Não-Paramétrico

Ao todo foram estimados 12 modelos distintos para obtenção do  $\hat{\tau}_{att}$ . Seis para o grupo de controle PSID e mais seis para o grupo de controle CPS. Abaixo, dividimos os resultados em 2 tabelas distintas, uma para cada amostra, para melhor analisarmos.

Tabela 7 - Efeitos do Treinamento utilizando PSM (Não-Paramétrico):DW\_PSID

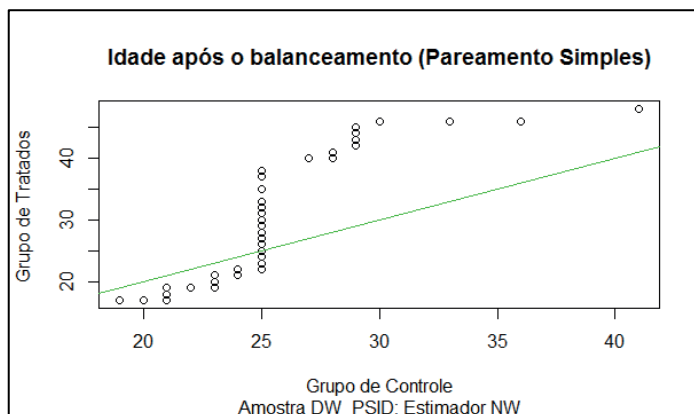
Método de pareamento	PSM Não-Paramétrico: Amostra DW_PSID		$\hat{\tau}_{att}$ (US\$)
	Observações Tratadas Pareadas	Observações Controle Pareadas	
	Estimador Local Constante (Nadaraya-Watson)		
Simples (1:1)	185	189	2356 [3720,3]
Caliper (0,25)	102	106	1285 [1008,6]
Genetic	185	565	2418 [3677]
	Estimador Local Linear		
Simples (1:1)	185	196	1914 [3502]
Caliper (0,25)	106	117	673 [1035]
Genetic	185	204	1969 [3457]

Fonte: elaborado pelo autor (2015).

Acima na tabela 7 podemos observar que grande parte dos resultados foram superestimados, o melhor resultado foi obtido quando utilizamos o método de pareamento *caliper* com o estimador Nadaraya-Watson, mas mesmo assim, abaixo do esperado.

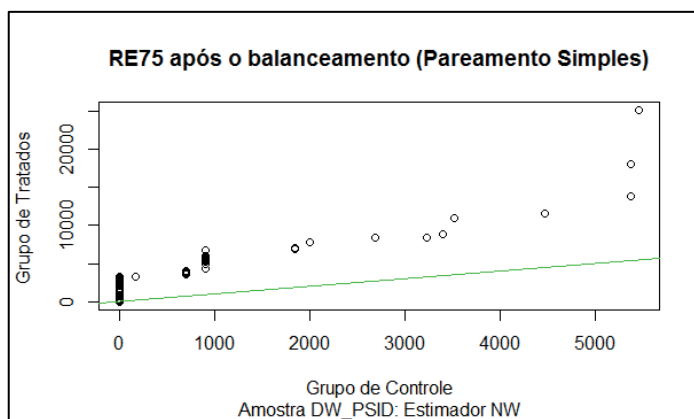
Abaixo na figura 5, está representado o balanceamento da variável idade após o pareamento simples utilizando a amostra DW\_PSID. Nela é possível enxergar claramente o quão ruim ficou o pareamento para essa variável, no gráfico, os círculos deveriam ficar sob a linha de 45° graus, no entanto, não é isso que observamos. Se olharmos para figura 6, a variável *RE75*, teve um balanceamento ainda pior, nela é possível enxergar através da linha vertical, que muitos indivíduos do grupo de controle com  $RE75 = 0$  estão sendo pareados com indivíduos com  $0 \geq RE75 \geq 4000$ .

Figura 5 - Idade após o Balanceamento (Pareamento Simples)



Fonte: elaborado pelo autor (2015).

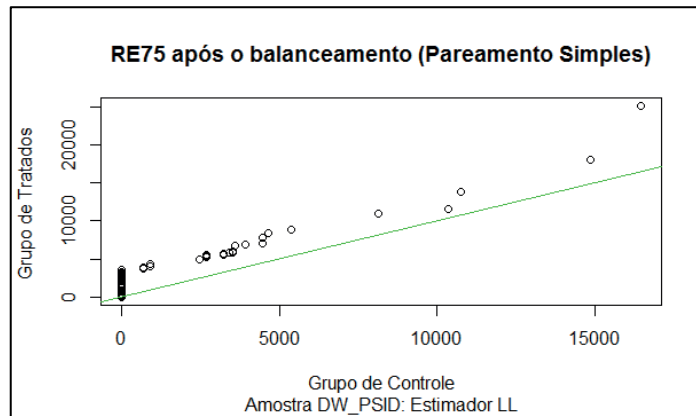
Figura 6 - Rendimentos em 1975 após o Balanceamento (Pareamento Simples/Nadaraya-Watson)



Fonte: elaborado pelo autor (2015).

Apesar do estimador local linear apresentar resultados mais próximos (US\$ 1914 e US\$ 1969) os erros padrões das suas estimativas são muito altos. Apesar disso, o pareamento ficou melhor com relação ao local constant. Abaixo na figura 7 apresentamos o balanceamento da variável RE75 após o pareamento simples.

Figura 7 - Rendimentos em 1975 após o Balanceamento (Pareamento Simples/Local Linear)



Fonte: elaborado pelo autor (2015).

Se compararmos com a figura 6, as observações estão bem mais próximas da linha de 45° graus, agora utilizando o local linear, entretanto, para a obtenção de um pareamento ótimo devíamos ter amostras centradas na linha verde do gráfico.

Logo, observando a tabela 7, os resultados obtidos nesse método não foram satisfatórios, pois não conseguimos criar um grupo contrafactual que refletisse o grupo de tratados, podendo assim, eliminar o viés de seleção. Entretanto, vale ressaltar que se compararmos com os resultados obtidos na Regressão com *Propensity Score*, esse método chegou muito mais próximo de eliminar o viés indesejado. Para efeitos de comparação, os resultados obtidos com a regressão, para a mesma amostra variaram entre US\$ 4138 e US\$ 3547, muito acima de US\$ 2418,5, o maior resultados obtido nesse método.

Na próxima tabela, apresentamos o resultado das estimações obtidos para o grupo de controle CPS. Novamente utilizamos as mesmas variáveis e os mesmos métodos de pareamento, segue abaixo a tabela 8 com os resultados.

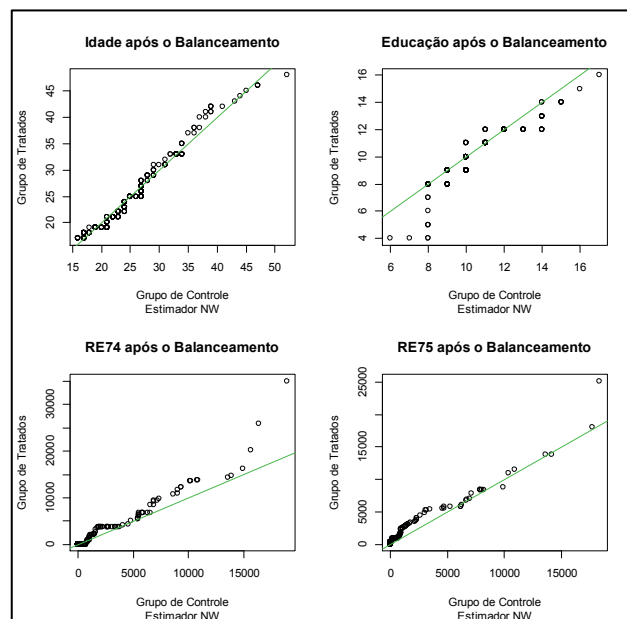
Tabela 8 - Efeitos do Treinamento utilizando PSM (Não-Paramétrico):DW\_CPS

Método de pareamento	PSM Não-Paramétrico: Amostra DW_CPS		$\hat{\tau}_{att}$ (US\$)
	Observações Tratadas Pareadas	Observações Controle Pareadas	
	Estimador Local Constante (Nadaraya-Watson)		
Exato (1:1)	185	306	2109 [1659]
Caliper (0,25)	144	265	2467 [822]
Genetic	185	442	2144 [1659]
	Estimador Local Linear		
Exato (1:1)	185	303	1340 [1520]
Caliper (0,25)	151	269	1883 [975]
Genetic	185	262	1424 [1513]

Fonte: elaborado pelo autor (2015).

Em relação ao estimador Nadaraya-Watson, analisando a parte de cima da tabela 8 também encontramos efeitos superestimados, o menor obtido é de US\$ 2109.

Figura 8 - Balanceamento das variáveis Idade, Educação, RE74 e RE75 (Amostra DW\_CPS/Nadaraya-Watson)



Fonte: elaborado pelo autor (2015).

Na figura 8 acima, podemos analisar o balanceamento das variáveis contínuas para o pareamento simples com reposição utilizando o estimadorNW, nela é possível observar como

o balanceamento foi otimizado para elas. Observem que para essas variáveis, as observações estão mais próximas da linha de 45° graus. Entretanto, na tabela abaixo, apresentamos as estatísticas das variáveis discretas *Sem Diploma e Casamento* presentes no modelo e podemos observar que o pareamento para elas não foi muito bom.

Tabela 9 - Estatísticas das Variáveis Sem Diploma e Casamento

Variável Sem Diploma		
	Antes do Pareamento	Após o Pareamento
Média Tratamento	0,70811	0,70811
Média Controle	0,29584	0,79591
DP da Diferença	90,43	-19,26
Variável Casamento		
	Antes do Pareamento	Após o Pareamento
Média Tratamento	0,18919	0,18919
Média Controle	0,71173	0,12231
DP da Diferença	-133,06	17,031

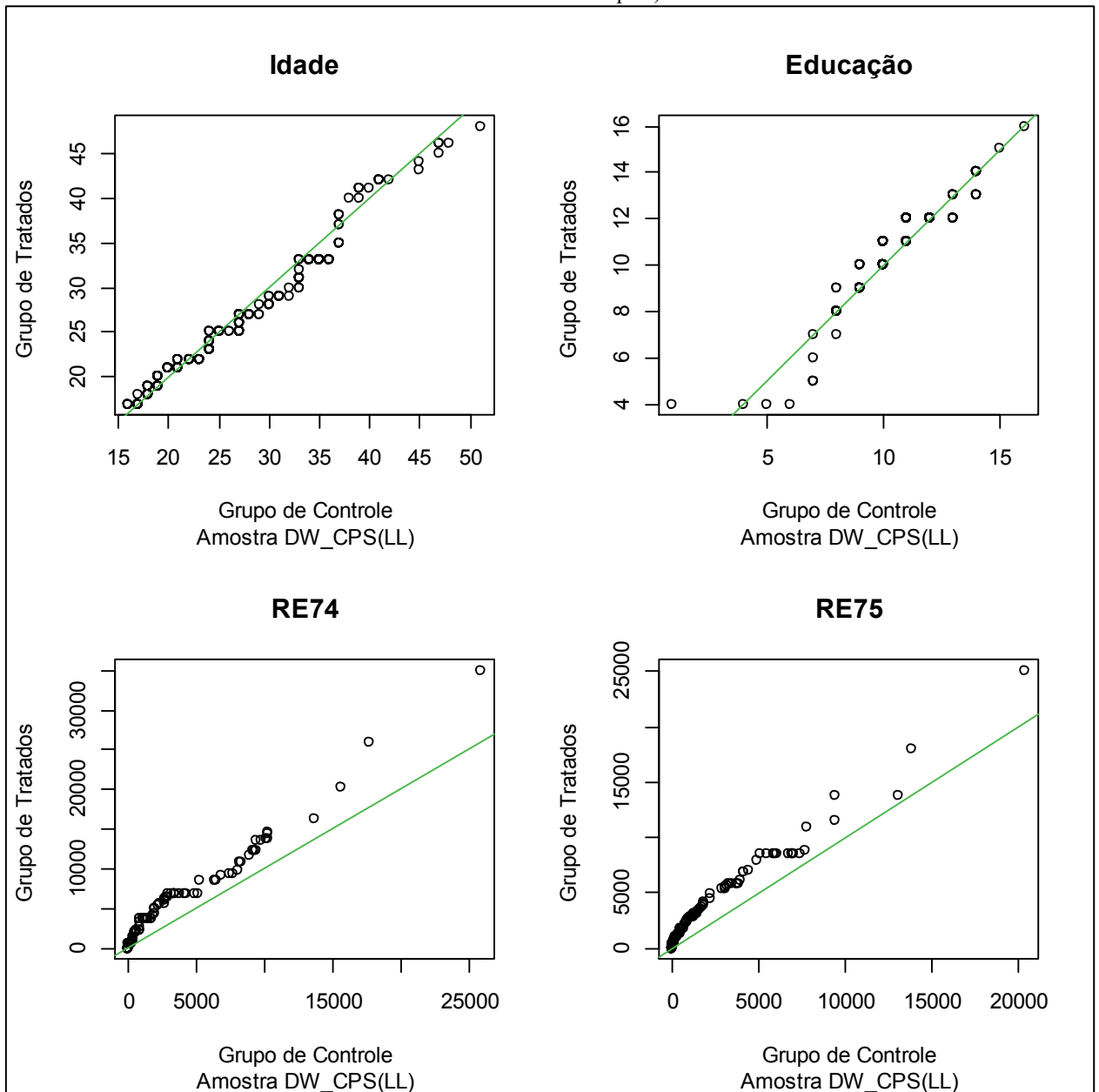
Fonte: elaborado pelo autor (2015).

Nota-se que o balanceamento após o pareamento melhorou bastante, mas se notarmos nos desvios padrões das diferenças, podemos observar que eles estão muito elevados. Nesse caso, os métodos de pareamento não foram suficientes para encontrar pares bons para os indivíduos que receberam o tratamento.

Já quando utilizamos o estimador local linear as estimativas do efeito do treinamento ficaram excelentes, todas ficaram muito próximas do resultado esperado. As figuras 9 e 10 abaixo apresentam o balanceamento das variáveis contínuas *idade, educação, RE74 e RE75* para os dois primeiros pareamentos (simples e *caliper*) respectivamente.



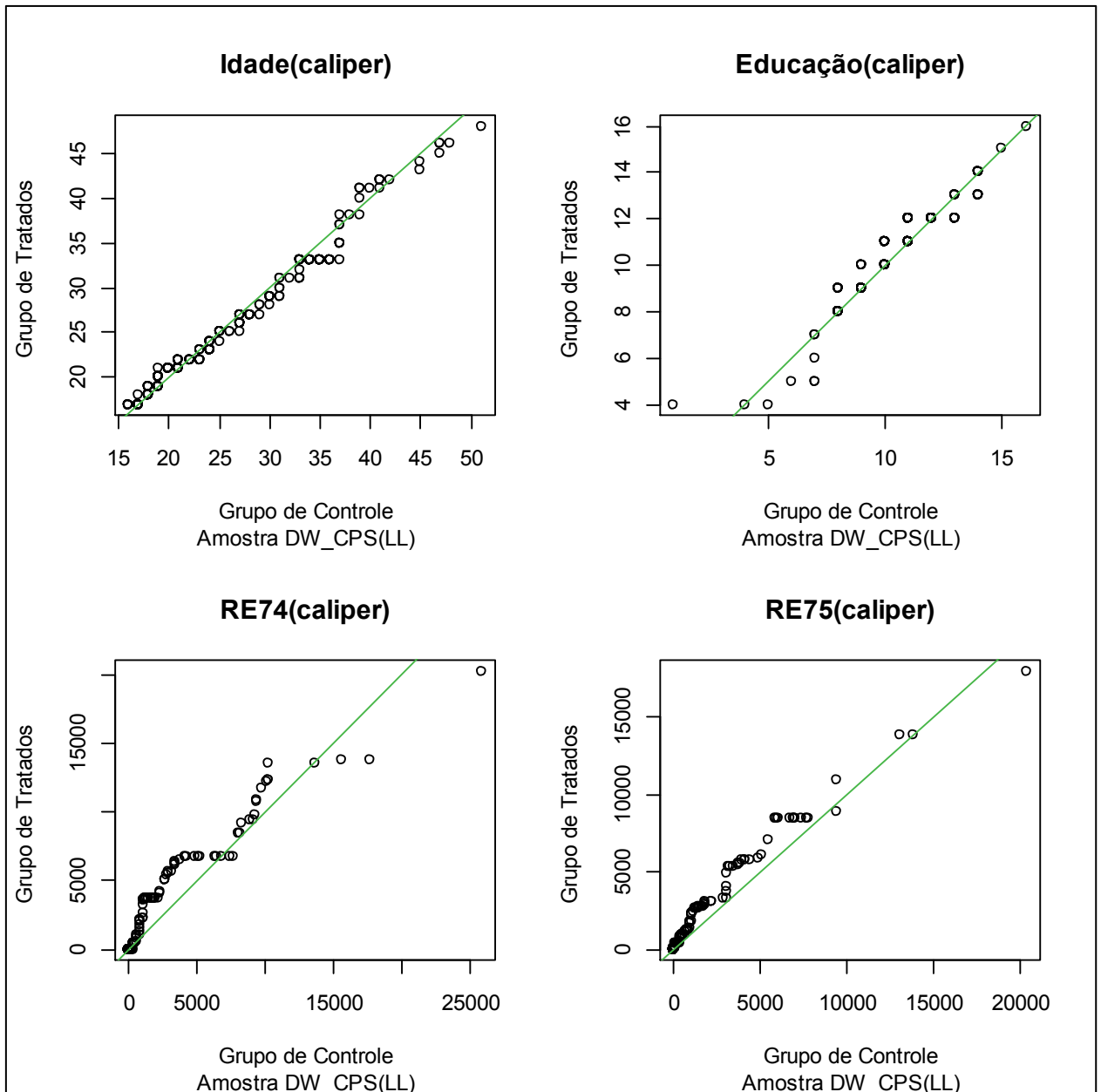
Figura 9 - Balanceamento das variáveis Idades, Educação, RE74 e RE75 (Amostra DW\_CPS/Local Linear/Pareamento Simples)



Fonte: elaborado pelo autor (2015).

Na figura 9 podemos observar que o balanceamento para as variáveis contínuas desse modelo (pareamento simples) ficaram boas. As variáveis RE74 e RE75 ficaram um pouco acima da linha de 45°, já as variáveis idade e educação apresentaram um balanceamento ótimo. Na figura 10, cujo modelo utilizou o método *caliper*, os resultados novamente são muito bons também.

Figura 10 - Balanceamento das variáveis Idades, Educação, RE74 e RE75 (Amostra DW\_CPS/Local Linear/Caliper)



Fonte: elaborado pelo autor (2015).

Portanto ao estimarmos o *propensity score* para a amostra DW\_CPS utilizando o estimador local linear, conseguimos obter resultados bem significantes, diferente de quando utilizamos o estimador Nadaraya-Watson. Agora ,com a união da estimação por local linear com os métodos de pareamento foi possível obter boas estimativas.

#### 4.6 ESTIMAÇÃO SEMI-PARAMÉTRICA<sup>20</sup>

A estimação semi-paramétrica é diferente da não-paramétrica explicada acima, pois agora somente algumas variáveis serão estimadas de forma não-paramétrica. O *propensity score* semi-paramétrico é estimado utilizando um modelo GAM como esse proposto abaixo:

$$\hat{g}(\mathbf{x}) = G[\mu(X)] = \alpha + f_1(X_1) + f_2(X_2) + \dots + f_p(X_p)$$

Onde  $\hat{g}(\mathbf{x})$  é o *propensity score* semiparamétrico, de tal maneira que a função link  $G = \text{Probit}$  ou  $G = \text{Logit}$ . No nosso trabalho ao invés de utilizarmos funções kernel para estimarmos as  $f_i$ , utilizaremos os *splines* cúbicos apresentados na equação (3.47)

A estimação agora é feita em duas etapas, primeiro estimamos o *propensity score*  $\hat{g}(\mathbf{x})$ , e depois, com o vetor  $\vec{g}$  obtido, estimamos o impacto do treinamento novamente de duas formas, a primeira utilizando uma regressão com o *propensity score*, como na equação abaixo

$$\hat{t}_{att,sp} = n^{-1} \sum_{i=1}^n \frac{[w_i - \hat{g}(\mathbf{x})]y_i}{\hat{\delta}[1 - \hat{g}(\mathbf{x})]} \quad (3.44')$$

E a segunda, utilizando o vetor  $\vec{g}$  junto com os três métodos de pareamento já utilizados na estimação não-paramétrica. A estimação do *propensity score* semiparamétrico foi realizada utilizando o pacote “mgcv” encontrado no R.

##### 4.6.1 Regressão com *Propensity Score* Semi-Paramétrico

Nesse método, além da equação (3.44'), novamente, utilizaremos a restrição [ $\hat{e}(x) < 0,5$  e  $\hat{e}(x) > 0,95$ ] para eliminarmos estimativas próximas de 0 e 1. Dessa vez foram estimados quatro modelos para cada grupo de controle, totalizando 8 no total. Os resultados obtidos foram separados para cada amostra, na tabela abaixo seguem os resultados obtidos para a amostra DW\_PSID.

<sup>20</sup> Para estimação semi-paramétrica dos modelos propostos, foi utilizado o pacote “mgcv” presente no software R.

Tabela 10 - Efeitos Estimados do Treinamento (*Propensity Score* Semi-Paramétrico)

Regressão com Propensity Score Semi-paramétrico: Amostra DW_PSID			
Modelo	$\hat{\tau}_{att}(\text{US\$})$	$R^2$	Amostras após a Restrição
Probit (A)	1776	0,675	374
Probit s/ Sem Diploma (B)	1771	0,675	374
Logit (C)	1912	0,681	350
Logit s/ Sem Diploma (D)	1888	0,68	356

Fonte: elaborado pelo autor (2015).

Nos modelos (A) e (C) ao estimarmos a regressão semi-paramétrica, todos os coeficientes não-paramétricos (variáveis: *educação, idade, re74 e re75*) se mostraram significativos, já para a parte paramétrica, os coeficientes da variável *sem diploma* não, logo, optamos por estimar o modelo sem a presença dessa variável.

Na tabela acima, todos os modelos apresentaram estimativas muito boas, principalmente os modelos (B) e (D), pois diferente dos modelos (A) e (C) apresentam todos os coeficientes significativos, tanto na parte paramétrica, quanto na parte não-paramétrica, o modelo (B) inclusive apresenta um resultado mais próximo ao *benchmark* de Lalonde (US\$ 1794) do que os modelos estimados por Dehejia e Wahba (1999, 2002) para amostra DW\_PSID.

Para a amostra DW\_CPS, os mesmos modelos foram estimados, abaixo na tabela são apresentados os resultados.

Tabela 11 - Efeitos Estimados do Treinamento (*Propensity Score* Semi-Paramétrico)

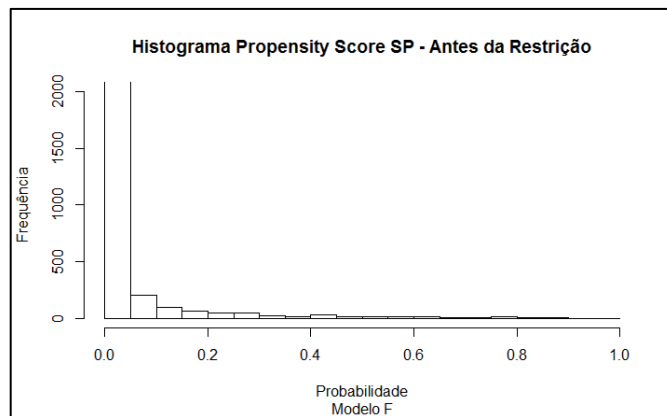
Regressão com Propensity Score Semi-Paramétrico: Amostra DW_CPS			
Modelo	$\hat{\tau}_{att}(\text{US\$})$	$R^2$	Amostras após a Restrição
Probit (E)	2143	0,423	632
Probit s/ Sem Diploma (F)	2164	0,417	644
Logit (G)	2049	0,429	556
Logit s/ Sem Diploma (H)	2096	0,425	557

Fonte: elaborado pelo autor (2015).

Diferente do que ocorreu com a amostra DW\_PSID, agora utilizando o grupo de controle, todas as estimativas do impacto do treinamento ficaram superestimadas. Todos os modelos apresentaram variáveis significativas tanto na parte paramétrica quanto na parte não-paramétrica. Se compararmos a terceira coluna das duas tabelas com os resultados observamos que o  $R^2$  para os modelos do grupo de controle CPS são menores do que os obtidos quando estimados para o grupo PSID.

Analisando o histograma do *propensity score* estimado pelo modelo (F), pré utilização da restrição, podemos observar mais uma vez a presença maciça dos *propensity* estimados no limite inferior  $p(x) = 0$ .

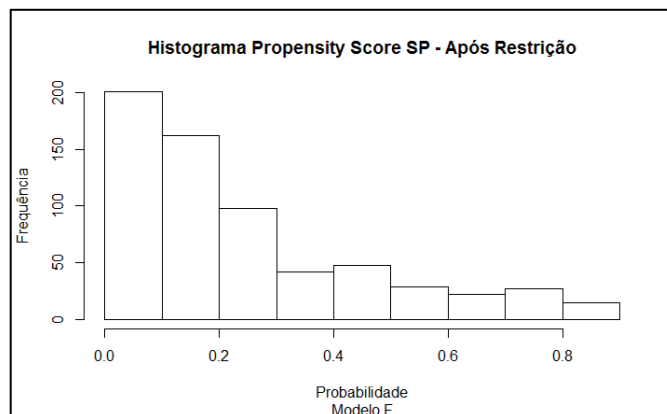
Figura 11 - Histograma *Propensity Score* Semi-Paramétrico (Antes da restrição/modelo F)



Fonte: elaborado pelo autor (2015).

Após utilizarmos a restrição, obtemos o seguinte histograma

Figura 12 - Histograma *Propensity Score* Semi-Paramétrico (Após restrição/modelo F)



Fonte: elaborado pelo autor (2015).

Similarmente como ocorreu com a regressão não-paramétrica, mesmo após a restrição, os *propensity score* estimados ficaram agrupados no limite inferior, com pouca presença no limite superior, prejudicando a estimação. Apesar de apresentar estimativas viesadas, esse método se mostrou mais eficaz do que quando estimado de forma não-paramétrica.

#### 4.6.2 PSM Semi-Paramétrico

Para essa seção utilizamos os mesmos *propensity score* estimados na seção anterior, mas dessa vez excluimos da estimação final, os *propensity score* semi-paramétricos estimados pela função link *logit*. Ao todo foram estimados 12 modelos para esse método, e para melhor compreensão dividimos os resultados abaixo em duas tabelas, uma para o grupo de controle PSID e outra para o grupo de controle CPS. Seguem os resultados obtidos quando aplicamos o PSM ao grupo de controle PSID.

Tabela 12 - Efeitos do Treinamento utilizando PSM (Semi-paramétrico): DW\_PSID

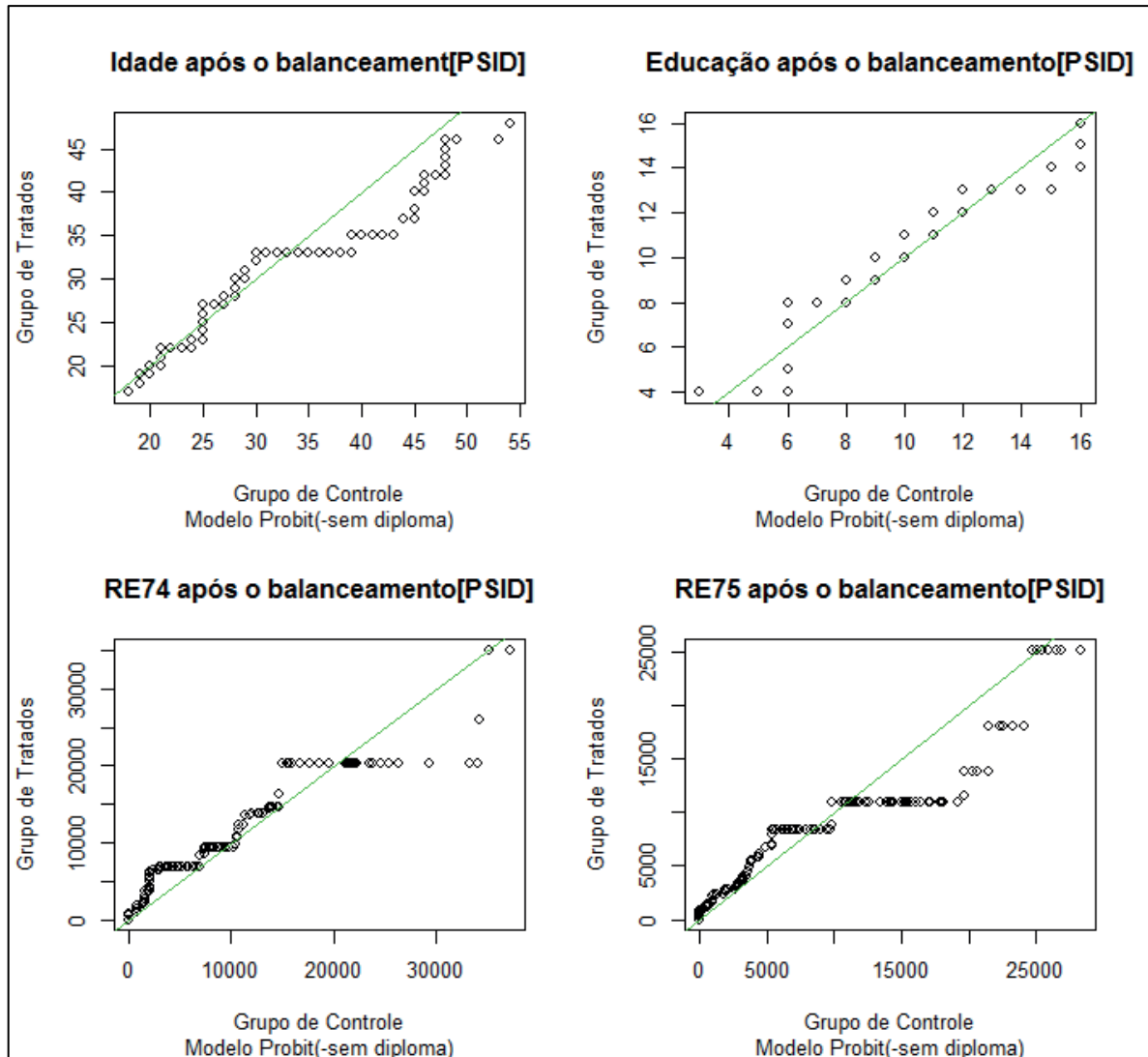
PSM Semi-paramétrico: Amostra DW_PSID				
Método de pareamento	Observações Tratadas Pareadas	Observações Controle Pareadas	$\hat{\tau}_{att}$ (US\$)	
Modelo Probit				
Simples (1:1)	185	333	1407	[2344]
Caliper (0,25)	174	322	1525	[2020]
Genetic	185	218	1268	[2391]
Modelo Probit (- sem diploma)				
Simples (1:1)	185	332	383	[3248]
Caliper (0,25)	169	316	409	[2636]
Genetic	185	526	390	[3232]

Fonte: elaborado pelo autor (2015).

Para o modelo *probit* completo (com todas as variáveis), podemos observar que apesar dos erros padrões elevados, os resultados estão dentro do intervalo esperado, com os valores variando entre US\$ 1268 e US\$ 1525,6. Já quando retiramos a variável *sem diploma* os resultados mudam drasticamente, com os valores entre US\$ 383 e US\$ 409. Esse fato nos

mostra o quão sensível foi a retirada da variável para a estimação. Abaixo na figura 13 apresentamos os gráficos do balanceamento das variáveis contínuas utilizando o método *caliper*.

Figura 13 - Balanceamento das variáveis Idades, Educação, RE74 e RE75 (PSID/Modelo *Probit* s/ diploma)



Fonte: elaborado pelo autor (2015).

Analisando os dois gráficos na parte de baixo da figura podemos observar que o pareamento funciona bem até certo valor para RE74 e RE75, após isso, indivíduos muito diferentes são pareados para estimação do efeito do treinamento, o que acaba prejudicando nossa estimação. Na tabela seguinte é apresentado o resultado quando aplicamos os modelos à amostra CPS.

Tabela 13 - Efeitos do Treinamento utilizando PSM (Semi-paramétrico):DW\_CPS

PSM Semi-paramétrico: Amostra DW_CPS			
Método de pareamento	Observações Tratadas Pareadas	Observações Controle Pareadas	$\hat{\tau}_{att}$ (US\$)
Modelo Probit			
Simple (1:1)	185	4442	2374,6 [984,68]
Caliper (0,25)	170	4427	1665 [760,87]
Genetic	185	321	2293 [1000]
Modelo Probit (- sem diploma)			
Simple (1:1)	185	14881	2039 [984]
Caliper (0,25)	182	14878	1815 [922]
Genetic	185	310	2291 [1000]

Fonte: elaborado pelo autor (2015).

Esses modelos aplicado ao grupo de controle CPS apresentaram pelo menos duas estimativas muito boas, ambas quando aplicamos o método *caliper*, US\$ 1665 e US\$ 1815, entretanto, todas estimativas restantes, para ambos os modelos ficaram superestimadas. O fato do modelo *probit* incompleto ter se saído melhor para o grupo de controle CPS nos mostra que o grupo PSID é mais sensível à utilização da variável *sem diploma*.

Os resultados obtidos do efeito do treinamento utilizando o método PSM, tanto na abordagem não-paramétrica, quanto na abordagem semi-paramétrica, se mostraram bastante sensíveis à escolha do tipo de pareamento, na tabela 13 acima, por exemplo, a utilização do método *caliper* ( $\delta = 0,25$ ) gerou resultados bastante satisfatórios ao contrário dos dois outros. Esses fatos corroboram o quão importante é a escolha do método de pareamento, além disso, o pesquisador após parear os indivíduos do grupo de tratamento com o grupo de controle, deve checar se o método foi benéfico ou não.

Portanto, a estratégia do pesquisador deve ter três passos importante: o primeiro, estimar os *propensity score* de forma não-paramétrica ou semiparamétrica, segundo, aplicar algum tipo de algoritmo de pareamento, além dos três métodos de pareamento utilizados nesse trabalho existem outros tipos que podem ser utilizados para auxiliar o pesquisador na busca do impacto do programa, e por último, devemos checar o quão útil foi o método de pareamento, através do gráfico do balanceamento.



## 5 CONCLUSÕES

Nesse trabalho propomos uma maneira diferente de estimarmos o impacto de um programa ao misturarmos abordagens não-paramétrica e semi-paramétrica com alguns métodos já utilizados na literatura econômica. Conseguimos extrair de alguns modelos estimados resultados bastante significativos, alguns até melhores do que os apresentados por Dehejia e Wahba (1999, 2002).

Ao estimarmos o *propensity score* de forma não-paramétrica, utilizando o método exposto por Li, Racine e Wooldridge (2009), mostramos que a quantidade de indivíduos com probabilidade de participação baixa, nos dois grupos de controle PSID e CPS, é muito elevada, ou seja, os dois grupos de controle escolhidos por Lalonde (1986), são muito diferentes do grupo de tratados que participaram do programa de treinamento, o mesmo ocorreu quando utilizamos o método GAM (HASTIE; TISHBIRANI, 1990) para estimação semi-paramétrica. Esse fato é muito importante, pois quanto maior a precisão na estimação do *propensity score*, melhor serão os resultados obtidos ao utilizarmos algum método de pareamento.

A maioria dos resultados ficaram superestimados, mas próximos dos resultados obtidos por Lalonde (1986), isso se deve ao fato que a qualidade das estimativas dependem também do grupo de comparação que estamos usando. Ao todo, na nossa aplicação econômica, foram estimados 36 modelos distintos para estimação do efeito médio do treinamento sobre os tratados  $\hat{\tau}_{att}$ , sendo 16 não-paramétricos e 20 semi-paramétricos. Na estimação não-paramétrica podemos destacar o método PSM Não-Paramétrico com estimador Local Linear tanto para a amostra DW\_PSID, mas principalmente para a amostra DW\_CPS (tabelas 7 e 8). Já na estimação Semi-paramétrica, destacamos os modelos de regressão semi-paramétrica com *propensity score* para a amostra PSID (tabela 10).

Diante dos resultados obtidos, devemos fazer uma ressalva, nessa aplicação em particular, nós temos disponível um resultado comparável, o *benchmark* US\$1794, o que não ocorre na maioria dos trabalhos, entretanto, a estratégia adotado no trabalho, de primeiro estimarmos os *propensity score* de forma não-paramétrica ou semi-paramétrica, segundo, escolher algum método de pareamento, e por último, checar o quão útil foi o método de pareamento, através do gráfico do balanceamento; parecem garantir a aplicabilidade dos métodos expostos em problemas reais.

Portanto, dito isso, diante das estimativas encontradas, é possível dizer que os métodos expostos nesse trabalho podem auxiliar um pesquisador na busca do impacto de um programa e devem ser adicionados ao conjunto de ferramentas de um pesquisador.

## REFERÊNCIAS

- ABADIE, A.; IMBENS, G.W. Large Sample Properties of Matching Estimators for Average Treatment Effects. **Econometrica**, Chicago, v. 74, p.235-267, 2006.
- AITCHISON, J.; AITKEN, C.G.G. Multivariate Binary Discrimination by the Kernel Method. **Biometrika**, London, v. 63, p. 413-420, 1976.
- ANGRIST, J.D. et al. Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment. **The American Economic Review**, Nashville, p.1535-1558, 2002.
- ASHENFELTER, O. **The Effect of Manpower Training Earnings**: Preliminary Results. Princeton: Princeton University, Department of Economics, Industrial Relations Section, 1974. (Working Papers, 440).
- ASHENFELTER, O. Estimating the Effect of training Programs on Earnings. **The Review of Economics and Statistics**, Cambridge, v. 60, p. 47-57, 1978.
- ASHENFELTER, O.; CARD, D. Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs. **The Review of Economics and Statistics**, Cambridge, v. 67, p. 648-660, 1985.
- BECKER, S.O.; ICHINO, A. Estimation of average Treatment Effects based on Propensity Scores. **The Stata Journal**, College Station, v. 2, p. 358-377, 2002.
- BOWMAN, A.W. An Alternative Method of Cross-Validation for the Smoothing of Density Estimates. **Biometrika**, London, v. 71, p. 353-360, 1984.
- BUJA, A.; HASTIE, T.J.; TIBSHIRANI, R.J. Linear Smoothers and Additive Models. **Annals of Statistics**, Hayward, v. 17, p. 453-555, 1989.
- CALIENDO, M.; KOPEINIG, S. Some Practical Guidance for the Implementation of Propensity Score Matching. **Journal of Economic Surveys**, Clevedon, v. 22, p. 31-72, 2008.
- CAMERON, A.C.; TRIVEDI, P.K. **Microeconometrics Methods and Applications**. Cambridge: Cambridge University Press, 2005.
- CARD, D.; SULLIVAN, D. Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment. **Econometrica**, Chicago, v. 56, p. 597-530, 1988.
- DEHEJIA, R.H.; WAHBA, S. Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. **Journal of the American Statistical Association**, New York, v. 94, p. 1053-1062, 1999.
- DEHEJIA, R.H.; WAHBA, S. Propensity Score-Matching Methods For Nonexperimental Causal Studies. **The Review of Economics and Statistics**, Cambridge, v. 84, p. 151-161, 2002.

DUFLO, E. **Field Experiments in Development Economics**. Cambridge: Massachusetts Institute of Technology, Department of Economics and Abdul Latif Jameel Poverty Action Lab, 2006.

DUFLO, E.; KREMER, M. **Use of Randomization in the Evaluation of Development Effectiveness**. [S.l.]: World Bank Operations Evaluation Department (OED), 2003.

DUFLO, E.; GLENNESTER, R.; KREMER, M. Using Randomization in Development Economics Research: A Toolkit. In: SCHULTZ, T.; STRAUSS, J. (Ed.). **Handbook of Development Economics**. [S.l.]: Elsevier, 2008.

FISHER, R.A. **The Design of Experiments**. Edinburgh: Oliver and Boyd, 1925.

GERTLER, P.J. et al. **Impact Evaluation in Practice**. [S.l.]: World Bank, 2011. Disponível em: <<https://openknowledge.worldbank.org/handle/10986/2550>>. Acesso em: 16 ago. 2015.

HALL, P.; RACINE J.S.; LI, Q. Cross-validation and the Estimation of Conditional Probability Densities. **Journal of the American Statistical Association**, New York, v. 99, p. 1015-1026, 2004.

HALL, P.; LI, Q.; RACINE J.S. Nonparametric Estimation of Regression Functions when there Exist Irrelevant Regressors. **Review of Economic and Statistics**, Cambridge, v. 89, p. 784-789, 2007.

HARDLE, W.; LINTON, O. Applied Nonparametric Methods. In: ENGLE, R. F.; MCFADDEN, D. L. **Handbook of Econometrics**. [S.l.]: Elsevier, 1994. p. 2295-2339.

HASTIE, T.; TIBSHIRANI, R. Generalized Additive Models. **Statistical Science**, Hayward, v. 1, p. 297-318, 1986.

HASTIE, T.; TIBSHIRANI, R. **Generalized Additive Models**. London: Chapman and Hall, 1990.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**. 2. ed. [S.l.]: Springer, 2009.

HECKMAN, J.J.; ICHIMURA, H.; TODD, P. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program. **Review of Economic Studies**, Oxford, v. 64, p. 605-654, 1997.

HECKMAN, J.J.; ICHIMURA H.; TODD, P. Matching as an Econometric Evaluation Estimator. **Review of Economic Studies**, Oxford, v. 65, p. 261-294, 1998.

HECKMAN, J.J.; ROBB, R. Alternatives Methods for Evaluating the Impact of Interventions. **Journal of Econometrics**, Amsterdam, v. 30, p. 239-267, 1985.

HO, D. et al. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. **Journal of Statistical Software**, [S.l.], v. 42, n. 8, p. 1-28, June 2011. Disponível em: <<http://www.jstatsoft.org/v42/i08>>. Acesso em: 23 jan. 2015.

HOLLAND, P.W. Statistics and Causal Inference. **Journal of the American Statistical Association**, New York, v. 81, p. 945-960, 1986.

HOROWITZ, J.L. **Semi parametric and Nonparametric Methods in Econometrics**. [S.l.]: Springer, 2009.

HOROWITZ, J.L.; MAMMEN, E. Nonparametric Estimation of an Additive Model with a Link Function. **Annals of Statistics**, Hayward, v. 32, p. 2412-2443, 2004.

IMBENS, G.W. Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review. **Review of Economics and Statistics**, Cambridge, v. 86, p. 4-29, 2004.

IMBENS, G.W.; WOOLDRIDGE, J.M. Recent Developments in the Econometrics of Program Evaluation. **Journal of Economic Literature**, Nashville, v. 47, p. 5-86, 2009.

KHANDKER, S.H.; KOOLWAL, G.B.; SAMAD, H.A. **Handbook on Impact Evaluation: Quantitative Methods and Practices**. [S.l.]: World Bank, 2010. Disponível em: <<https://openknowledge.worldbank.org/handle/10986/2693>>. Acesso em: 2 mar. 2015.

KRUEGER, A.B. Experimental Estimates of education Production Functions. **Quarterly Journal of Economics**, Cambridge, v. 114, p. 497-532, 1999.

KUSEK, J.Z.; RIST, R.C. **Ten Steps to a Results-Based Monitoring and Evaluation System**. [S.l.]: World Bank, 2004. Disponível em: <<https://openknowledge.worldbank.org/bitstream/handle/10986/14926/296720PAPER0100steps.pdf?sequence=1>>. Acesso em: 13 mar. 2015.

LALONDE, R.J. Evaluating the Econometric Evaluations of Training Programs with Experimental Data. **The American Economic Review**, Nashville, v. 76, p. 604-620, 1986.

LI, Q.; RACINE, J.S. Nonparametric Estimation of Distributions with Categorical and Continuous Data. **Journal of Multivariate Analysis**, New York, v. 86, p. 266-292, 2003.

LI, Q.; RACINE, J.S. **Nonparametric Econometrics**. Princeton: Princeton University Press, 2007.

LI, Q.; RACINE, J.S.; WOOLDRIDGE, J.M. Efficient Estimation of Average Treatment Effects with Mixed Categorical and Continuous Data. **Journal of Business and Economic Statistics**, Washington, v. 27, p. 206-223, 2009.

LINTON, O.B.; NIELSEN, J.B. A Kernel Method of Estimating Structured Nonparametric Regression Based on Marginal Integration. **Biometrika**, London, v. 82, p. 93-100, 1995.

NADARAYA, E.A. On Nonparametric Estimates of Density Functions and Regressions Curves. **Theory of Applied Probability**, [S.l.], v. 10, p. 186-190, 1965.

NEYMAN, J. Statistical Problems in Agricultural Experiments. **Journal of the Royal Statistical Society**, [S.l.], v. 2, p. 107-180, 1923.

- PAGAN, A.; ULLAH, A. **Nonparametric Econometrics**. Cambridge: Cambridge University Press, 1999.
- RACINE, J.S.; LI, Q. Nonparametric Estimation of Regression Functions with both Categorical and Continuous Data. **Journal of Econometrics**, Amsterdam, v. 119, p. 99-130, 2004.
- RAVALLION, M. The Mystery of Vanishing Benefits: An Introduction to Impact Evaluation. **The World Bank Review**, [S.l.], v. 15, p. 115-140, 2001.
- RAVALLION, M. **Evaluating Anti-Poverty Programs**. [S.l.]: The World Bank, 2005. (Policy Research Working Paper Series, 3625).
- ROBINSON, P. Root-N Consistent Semiparametric Regression. **Econometrica**, Chicago, v. 56, p. 931-954, 1988.
- ROSEBAUM, P.R.; RUBIN, D.B. The Central Role of Propensity Score in Observational studies for Causal Effects. **Biometrika**, London, v. 70, p. 41-55, 1983.
- ROY, A. Some Thoughts on the Distribution of Earnings. **Oxford Economic Papers**, Oxford, v. 3, p. 135-146, 1951.
- RUBIN, D.B. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. **Journal of Educational Psychology**, Arlington, v. 66, p. 688-701, 1974.
- RUBIN, D.B. Inference and Missing Data. **Biometrika**, London, v. 63, p. 581-592, 1976.
- RUBIN, D.B. Bayesian Inference for Causal Effects: The Role of Randomization. **The Annals of Statistics**, Hayward, v. 6, p. 34-58, 1978.
- RUDEMO, M. Empirical Choice of Histograms and Kernel Density Estimators. **Scandinavian Journal of Statistics**, Stockholm, v. 9, p. 65-78, 1982.
- RUPPERT, D.; WAND, M.P.; CARROLL, R.J. **Semiparametric Regression**. Cambridge, 2003.
- SEKHON, J.S. Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R. **Journal of Statistical Software**, [S.l.], v. 42, n. 7, 2011.
- SILVERMAN, B.W. Spline Smoothing: The Equivalent of Variable Kernel Method. **Annals of Statistics**, Hayward, v. 12, p. 898-916, 1984.
- SMITH, J.A.; TODD, P.E. Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators? **Journal of Econometrics**, Amsterdam, v. 125, p. 305-353, 2005.
- STONE, C.J. Optimal Rates of Convergence for Nonparametric Estimators. **Annals of Statistics**, Hayward, v. 9, p. 1348-1360, 1980.

STONE, C.J. An Asymptotically Optimal Window Selection Rule for Kernel Densities Estimates. **Annals of Statistics**, Hayward, v. 12, p. 1285-1297, 1984.

WAND, M.P.; JONES, M.C. **Kernel Smoothing**. [S.l.]: Springer, 1995.

WATSON, G.S. Smooth Regression Analysis. **Sankhya**, Série A, Calcutta, v. 26, p. 359-372, 1964.

WOOLDRIDGE, J.M. **Econometric Analysis of Cross Section and Panel Data**. 2. ed. Cambridge: MIT Press, 2010.

YATCHEW, A. Nonparametric Regression Techniques in Economics. **Journal of Economic Literature**, Nashville, v. 36, p. 669-721, 1998.