

Universidade Federal de Santa Catarina – UFSC

Centro Socioeconômico – CSE

Departamento de Economia e Relações Internacionais – CNM

CAROLINA KOWALSKI PIAZZA

Avaliação do impacto econômico da gratuidade no transporte coletivo

Florianópolis, junho de 2017.

CAROLINA KOWALSKI PIAZZA

Avaliação do impacto econômico da gratuidade no transporte coletivo

Monografia submetida ao curso Graduação em Ciências Econômicas da Universidade Federal de Santa Catarina como requisito obrigatório para a obtenção do grau de Bacharelado.

Orientador: Prof. Dr. Guilherme Valle Moura

Florianópolis, junho de 2017.

Universidade Federal de Santa Catarina – UFSC

Centro Socioeconômico – CSE

Departamento de Economia e Relações Internacionais – CNM

A Banca Examinadora resolveu atribuir a nota 10,0 à aluna Carolina Kowalski Piazza na disciplina CNM 7107 – Monografia, pela apresentação deste trabalho.

Banca examinadora:

Prof. Dr. Guilherme Valle Moura.
Orientador

Prof. Dr. Jaylson Jair da Silveira.

Prof. Dra. Eva Yamila Amanda da Silva
Catela.

AGRADECIMENTOS

Agradeço aos meus pais, por sempre apoiarem minhas decisões, mesmo muitas vezes sem entendê-las, e por se sacrificarem para me dar a melhor formação possível. A eles devo quem sou.

Ao Bruno, pela parceria inabalável e por tornar tudo mais leve.

Agradeço aos amigos, Emilly, Mateus, Dioner, Márcio, Akauã e Aisha, cujo apoio foi fundamental.

Ao meu orientador, professor Guilherme Moura, pelos conselhos, pela orientação, pelo incentivo, pelos livros emprestados, por tudo. Você foi incansável!

Aos diversos professores da UFSC que, cada um à sua maneira, mudaram um pouquinho minha visão de mundo. Em particular, gostaria de agradecer ao professor Jaylson da Silveira e ao professor André Portela, que estiveram sempre com a porta aberta, fosse para tirar dúvidas, fosse para dar conselhos, e contribuíram imensamente para minha formação.

*“It’s the questions we can’t answer that teach us the most. They teach us how to think.
If you give a man an answer, all he gains is a little fact. But give him a question and
he’ll look for his own answers.”*
(Patrick Rothfuss - The Wise Man’s Fear)

Resumo

O presente trabalho tem como objetivo avaliar o impacto da gratuidade universal do transporte coletivo urbano sobre o salário médio e o nível de emprego dos municípios que a implementaram entre os anos de 2004 e 2017. Utilizou-se o método de pareamento com escore de propensão para construção de um grupo de controle comparável ao grupo de municípios que adotaram a política e, em seguida, empregou-se um modelo de painel de efeitos fixos para estimação do impacto sobre as variáveis de resultado. Tanto o impacto estimado sobre o emprego quanto sobre o salário foram positivos, de 1608,74 e de R\$ 74,67 respectivamente, ambos com p-valor inferior a 0,001. O estudo conclui que a implementação da tarifa zero de ônibus pode contribuir positivamente para o crescimento da renda e do emprego e pode ser uma poderosa ferramenta de inclusão social.

Palavras-chave: avaliação de impacto; tarifa zero; pareamento com escore de propensão; efeitos fixos.

Abstract

The aim of this study is to evaluate the impact of the universal gratuity of public urban transport on the average salary and the level of employment of the municipalities that implemented it between the years 2004 and 2017. Propensity score matching was used to build a control group sample comparable to the municipalities that adopted said policy and then a two-way fixed effects model was used to estimate impact on the outcome variables. The impacts on both employment and wage were found to be positive, of 1608,74 and R\$ 74,67 respectively, both with p-value under 0,001. The study concludes that free bus pass contributes positively to income and job growth and can be a powerful tool for social inclusion.

Keywords: impact evaluation; free fare; propensity score matching; fixed effects.

Lista de figuras

Figura 1 – Ilustração da hipótese de suporte comum.	38
Figura 2 – Representação da hipótese de tendência comum e do estimador de diferenças em diferenças.	52
Figura 3 – Mecanismos de impacto do investimento no setor de transporte no crescimento econômico.	67
Figura 4 – Mobilidade por modo de viagem.	70
Figura 5 – Comparação dos histogramas dos escores de propensão antes e depois do pareamento.	82
Figura 6 – Gráfico de dispersão das unidades tratadas e de controle, pareadas e não pareadas. Todas as unidades de tratamento foram pareadas.	82

Lista de tabelas

Tabela 1	– Definições teóricas de efeito causal quando se dispõe de duas unidades i e j .	27
Tabela 2	– Relação entre as definições teóricas de causalidade e as definições observáveis.	31
Tabela 3	– Interpretação dos coeficientes do formato de regressão linear do modelo de diferenças em diferenças.	55
Tabela 4	– Índice de mobilidade de acordo com a renda para os anos de 2007 e 2012. . .	69
Tabela 5	– Mobilidade de acordo com escolaridade para os anos de 2007 e 2012. . .	70
Tabela 6	– Exemplos de municípios com observações discrepantes de salário.	76
Tabela 7	– Médias das variáveis por grupo (tratamento e controle).	78
Tabela 8	– Médias das variáveis de resultado em abril de 2017 por grupo.	78
Tabela 9	– Modelo logit para previsão de participação no programa	81
Tabela 10	– Resultados antes do pareamento. Aqui, os modelos utilizados foram de efeitos fixos de tempo, incluindo a variável de tratamento como regressor. . . .	85
Tabela 11	– Estimação do impacto para a amostra	87
Tabela 12	– Estatísticas descritivas das variáveis do censo de 2000.	94
Tabela 13	– Teste de Mann-Whitney-Wilcoxon	95
Tabela 14	– Amostra resultante do pareamento (grupo de tratados e controle).	96
Tabela 15	– Teste de Mann-Whitney-Wilcoxon após o pareamento.	97

Sumário

1	INTRODUÇÃO	17
1.1	Tema e problema de pesquisa	17
1.2	Objetivos	20
1.3	Justificativa	21
1.4	Estrutura do trabalho	22
2	REFERENCIAL TEÓRICO-METODOLÓGICO	23
2.1	Definição de efeito causal no modelo de Neyman-Rubin	24
2.2	Dados experimentais	32
2.3	Hipótese de valor estável de unidade-tratamento	33
2.4	Viés de seleção	34
3	PAREAMENTO	37
3.1	Mecanismo de seleção	37
3.2	Principais hipóteses	37
3.3	Escore de propensão	40
4	ESTIMADORES DE IMPACTO	49
4.1	Diferenças em diferenças	50
4.2	Modelo de efeitos fixos	56
5	TRANSPORTE E DESENVOLVIMENTO	59
5.1	Pobreza e exclusão social no Brasil	59
5.2	Impactos de intervenções no setor de transporte	64
5.3	Padrão de mobilidade da população de baixa renda	68
5.4	Políticas públicas de transporte	71
6	RESULTADOS E DISCUSSÃO	75
6.1	Origem e tratamento dos dados	75
6.2	Características dos municípios tratados e não tratados	76
6.3	Pareamento	77
6.4	Estimativa de impacto	85
6.5	Discussão	86
7	CONCLUSÃO	89
	REFERÊNCIAS	91

APÊNDICE A – ESTATÍSTICAS DESCRITIVAS	93
---	----

1 Introdução

A Campanha pelo Passe Livre, que em 2005 originou o Movimento Passe Livre, nasceu em Florianópolis no ano de 1999 (MPL, 2017). Desde então, diversas cidades do Brasil foram palco de manifestações que permeiam a questão do transporte público, algumas contra o aumento da tarifa e outras na defesa da universalização do serviço de transporte coletivo. A implementação da tarifa zero envolve a mobilização de recursos públicos e, enquanto política pública, deve ser avaliada no tocante às suas implicações econômicas. Mediante a utilização de dados dos municípios que a adotaram, este trabalho busca avaliar seu impacto sobre o nível de emprego e salário médio dos mesmos.

1.1 Tema e problema de pesquisa

Muito se discute a respeito de como enfrentar os problemas sociais no Brasil e, dentre eles, a pobreza. O conceito de pobreza é contextual — depende de características geográficas e históricas — todavia, pode-se caracterizar a situação de pobreza como uma de carência, em que os indivíduos não têm suas necessidades básicas satisfeitas; ela engloba não apenas a insuficiência de renda em si, mas acesso a condições mínimas que proporcionem uma vida digna ao indivíduo, como acesso à água encanada, energia elétrica e aos demais serviços públicos essenciais.

Barros, Henriques e Mendonça (2001) realizam uma análise da pobreza no Brasil a partir de duas óticas: a escassez de recursos e sua má distribuição. No tocante à escassez, os autores utilizaram três critérios, a saber, uma comparação do Brasil com outros países, um exame do padrão de consumo das famílias brasileiras e também da estrutura de renda média do país. O estudo conclui que a origem da pobreza não se encontra na escassez de recursos, isto é, o Brasil não é um país pobre. Dado que o nível médio de renda no país é substancialmente superior à linha da pobreza, os autores concluem que a desigualdade é determinante no panorama da pobreza no país.

Diante desse quadro, uma saída seria recorrer às políticas públicas para promover o combate à desigualdade econômica e social e à pobreza, desenhadas com o intuito de enfrentar as mazelas sociais do país e garantir os direitos fundamentais dos cidadãos. No marco teórico do desenvolvimento como liberdade, de Sen (2000), a intervenção governamental viria no sentido de expandir as liberdades substantivas dos indivíduos, conferindo-lhes capacidade de agir sobre seu próprio destino.

As grandes questões que permeiam o tema das políticas públicas geram controvérsia: desde se os problemas sociais deve ser resolvidos pelos agentes estatais ou em que medida a intervenção estatal é bem-vinda, até seus pormenores, como qual o estilo de política a ser executado, a que parcelas da população ela se dirige, se haverá uma parceria com empresas privadas, e assim por diante. Uma das ferramentas que permite balizar tais decisões de cunho político é uma avaliação de impacto. Ela permite verificar se o objetivo das políticas públicas é atingido, quantificá-lo e compreender as principais características e mecanismos que atuam para seu sucesso (ou falha) (PEIXOTO *et al.*, 2012).

De acordo com os dados do censo demográfico do IBGE, em 2010, 84,36% da população brasileira residia em áreas urbanas, o que coloca o debate a respeito do acesso universal à cidade como fundamental quando se almeja uma sociedade mais igualitária. Para Gomide (2003), um transporte coletivo acessível, eficiente e de qualidade pode aumentar a renda e o tempo disponíveis aos mais pobres simplesmente dando-lhes acesso a todo o espaço urbano — que inclui acesso aos serviços básicos de saúde, educação e lazer, além de oportunidades de trabalho. Sob essa ótica, uma política de transporte pode ser um poderoso instrumento de inclusão social. Isso ficou evidente com as manifestações de 2013, em que as reivindicações contra aumento da tarifa de ônibus culminaram em um levante de grandes proporções, o qual eventualmente incorporou diversas outras exigências de melhorias nos serviços públicos. É partindo dessa hipótese, de que o transporte urbano tem potencial transformador no processo de desenvolvimento econômico das cidades, que elabora-se este trabalho.

Existem diversos tipos de políticas de transporte, com diferentes objetivos. Dentre essas, uma política bastante difundida é a isenção da tarifa para determinados segmentos da população (a Constituição prevê gratuidade ou desconto para idosos, estudantes e pessoas com deficiência). É necessário ressaltar, no entanto, que o valor da tarifa é calculado pelo rateio dos custos de operação pelo número de pagantes (GOMIDE, 2003). Isso significa que trabalhadores de classes mais baixas financiam a passagem de, por exemplo, estudantes de escolas privadas — indicando que essa política pode ter um efeito regressivo se não avaliada com cuidado (PEREIRA *et al.*, 2015). Outra opção de política, muito reivindicada por movimentos sociais é a gratuidade universal no transporte coletivo. Ao redor do mundo, diversas cidades implementaram a tarifa zero¹, financiada pelos recursos municipais e alguns fundos específicos. Nos últimos anos, alguns municípios brasileiros também aderiram à política, a saber:

a) Agudos – SP (Mar 2003)

c) Eusébio – CE (Out 2010)

b) Anicuns – GO (2005)

d) Itatiaiuçu – MG (Jul 2015)

¹ Alguns exemplos são: Mariehamn (2000) na Finlândia, Ovacik (2014) na Turquia, e Talinn (2013) na Estônia.

- | | |
|-------------------------------|---------------------------------|
| e) Ivaiporã – PR (Jun 2001) | j) Pitanga – PR (Fev 2012) |
| f) Maricá – RJ (Dez 2014) | k) Porto Real – RJ (Set 2011) |
| g) Monte Carmelo – MG (1994) | l) Potirendaba – SP (1998) |
| h) Muzambinho – MG (Out 2011) | m) Silva Jardim – RJ (Fev 2014) |
| i) Paulínia – SP (Ago 2013) | |

Contudo, há pouca informação acerca do sucesso dessa política no sentido de gerar encadeamentos positivos que proporcionam o desenvolvimento econômico e a expansão da liberdade. Tendo em mente que qualquer decisão de intervenção estatal incorre em mobilizar recursos públicos, é necessário que se vá além da especulação quanto aos benefícios gerados pela tarifa zero. Diante disso, propõe-se extrair informações quanto ao impacto da gratuidade de ônibus nos indicadores de salário médio e emprego desses municípios, implementando técnicas de determinação causal utilizadas no âmbito das avaliações de políticas públicas.

1.1.1 Causalidade

Conforme será detalhado ao longo do trabalho, a avaliação de políticas públicas comumente se dá no âmbito dos estudos observacionais ou quase experimentais. Em contraste com os estudos experimentais, em que o pesquisador controla as condições em que o experimento é realizado, a avaliação via estudos observacionais se dá a *posteriori* e o pesquisador, em geral, não tem qualquer controle sobre a seleção dos participantes da política (denominado grupo de tratamento) (HOLLAND, 1986). Isso gera problemas na definição teórica de efeito causal e também na operacionalização de meios eficazes de estimá-lo. Em suma, o objetivo de realizar uma inferência causal não é tarefa simples, a começar pelo que se entende por relação de causalidade.

Segundo Box-Steffensmeier, Brady e Collier (2008), inferências causais sólidas devem satisfazer quatro requisitos, a saber, a) uma ligação entre a causa e o efeito (por exemplo, o estabelecimento de correlação), b) a ausência de efeito quando há ausência da causa em um cenário o mais semelhante possível à realidade, isto é, a investigação de um contrafactual, c) presença de efeito quando a causa é manipulada e, por fim, d) a identificação de atividades e processos que conectam causa e efeito.

Duas noções concorrentes do que se entende por causalidade chamam atenção; a primeira, elaborada por David Hume, compreende que uma relação de causa é um “truque psicológico”, em que o observador percebe a repetição das causas, seguidas de efeitos e conclui haver uma conexão mais profunda entre os dois efeitos. Isso significa que, para

Hume, a essência de uma relação causa-efeito é a regularidade (BOX-STEFFENSMEIER; BRADY; COLLIER, 2008). Em abordagens derivadas do pensamento de Hume, entende-se que relações de causa e efeito são como leis e, portanto, incondicionais. Outra visão, mais relevante para o escopo deste trabalho, foi elaborada pelo linguista George Lakoff e pelo filósofo Mark Johnson e descreve causalidade como uma manipulação forçada da realidade para produzir mudanças no ambiente físico e, por conseguinte, consiste em pensar nos fenômenos a partir de uma perspectiva contrafactual, manipulando eventos antecedentes a fim de observar os efeitos que emergem (BOX-STEFFENSMEIER; BRADY; COLLIER, 2008). Essa é a noção de causalidade que prevalece na literatura acerca das avaliações de impacto, cujos métodos se apoiam no modelo causal de Neyman-Rubin.

Embora um experimento controlado, com seleção aleatória de grupos de controle e tratamento, seja o tipo mais desejável, raramente é viável em estudos econômicos, principalmente em se tratando de fenômenos de proporções regionais ou nacionais. Ainda assim, esse tipo de experimento requer que diversas hipóteses sejam satisfeitas para que se possa afirmar que há causalidade entre eventos. No âmbito dos estudos observacionais, a discussão se complexifica e hipóteses alternativas são necessárias para afirmar a existência de uma relação de causa.

A discussão acima evidencia a importância do método em estudos empíricos nos quais se investigam relações causais. Diante disso, optou-se — talvez diferentemente do que seria esperado de um trabalho desta natureza, cujo principal objetivo é a avaliação de impacto — em explorar mais profundamente as questões teóricas que permeiam os métodos de avaliação de políticas e na compreensão de suas hipóteses.

1.2 Objetivos

1.2.1 Objetivo Geral

Determinar o impacto econômico da política de tarifa zero de ônibus no período compreendido entre janeiro de 2004 e abril de 2017.

1.2.2 Objetivos Específicos

- a) Discutir a função de políticas de transporte urbano no contexto de inclusão social e combate à pobreza;
- b) Caracterizar os encadeamentos que relacionam intervenções no setor de transporte de passageiros com o desenvolvimento econômico;
- c) Caracterizar brevemente as principais políticas de transporte urbano do Brasil;

- d) Apresentar detalhadamente os métodos de avaliação de impacto escolhidos, abordando o corpo teórico dos modelos de determinação causal, bem como suas principais hipóteses e limitações;
- e) Estimar o efeito médio do programa sobre emprego e salário médio a partir dos métodos apresentados.

1.3 Justificativa

É importante que se esclareça os motivos pelos quais se fazem necessárias avaliações de impacto. No caso específico das políticas públicas, parte-se da ideia de que os recursos públicos são limitados e, portanto, deve-se buscar sua alocação da maneira mais eficiente possível — ou seja, deseja-se saber se determinada política proporciona um benefício à sociedade que supera seus custos. Nesse sentido, as avaliações de impacto são parte de algo mais amplo, a “realização de políticas baseadas em evidências”² (KHANDKER; KOLWAL; SAMAD, 2009). Portanto, mesmo que um programa seja amplamente utilizado, é possível que não seja condizente com os custos associados ou que seus objetivos não tenham sido atingidos. Desse modo, conforme ressaltam Peixoto *et al.* (2012), a função de uma avaliação de impacto não precisa se restringir a verificar a existência de impacto, mas também sua magnitude e outras características relevantes. Dependendo da metodologia empregada, pode-se também compreender os canais através dos quais tais benefícios se manifestam.

A determinação do caráter do impacto também pode auxiliar na “validação das teorias que serviram de base ao desenho do programa” (PEIXOTO *et al.*, 2012). Isso é essencial pois permite a implementação do programa em outro lugar e até a alteração de suas características para que se adéque a diferentes contextos socioeconômicos.

Há ainda muitos casos de programas sociais que são ofertados gratuitamente. É de se imaginar que a adesão voluntária seria um indicador de impacto. Todavia, Peixoto *et al.* (2012) argumentam que os usuários podem ter pouca informação a respeito da efetividade do programa e a gratuidade da adesão implica que os usuários não perdem nada participando. Outra situação possível é quando a gratuidade só é válida para os segmentos mais pobres da sociedade e o mesmo serviço é oferecido aos ricos de maneira privada, os quais pagam preços significativos. De fato, a partir da teoria econômica deduz-se que a demanda por um bem a um preço positivo deveria ser uma evidência do impacto. Mas ainda restaria dúvidas quanto à sua magnitude (em vista das diferentes prioridades da alocação da renda entre pobres e ricos) e também não há garantia de que o impacto sobre os pobres é similar àquele sobre os ricos (PEIXOTO *et al.*, 2012).

² *Evidence based policy making.*

Muitas vezes, políticas sociais são destinadas a uma determinada parcela da população, um grupo alvo. Assim, a avaliação de impacto serve para verificar se o objetivo é atingido e, mais ainda, para procurar identificar em que medida o sucesso do programa depende das características de tal população alvo (PEIXOTO *et al.*, 2012). Ela pode servir como ferramenta de conscientização do público e responsabilização do governo pelo uso dos recursos públicos (KHANDKER; KOOLWAL; SAMAD, 2009).

Por fim, realiza-se uma avaliação de impacto sempre com o intuito de estabelecer uma relação causal entre os resultados observados e determinada intervenção (KHANDKER; KOOLWAL; SAMAD, 2009). Estima-se, portanto, o impacto médio do programa sobre o bem-estar dos indivíduos afetados, medido através de variáveis de resultado relevantes.

No que tange às políticas de transporte, as avaliações de impacto realizadas costumam ter foco em sua eficiência (GOMIDE, 2003). No entanto, é possível que tal abordagem não favoreça a adoção de medidas em favor das parcelas mais pobres da população. Propõe-se encarar o transporte urbano como meio indireto de inclusão social — ou seja, que ele seja visto como um meio complementar de combate à pobreza na medida em que favorece o crescimento econômico, ao passo que propicia acesso às oportunidades que efetivamente permitem que os indivíduos superem suas condições de carência.

Os resultados encontrados foram positivos, tanto no nível de emprego quanto no de salário. O impacto estimado para o emprego foi de 1608,74 e no salário foi de R\$ 74,67.

1.4 Estrutura do trabalho

No capítulo 2, como referencial teórico-metodológico, será apresentada uma discussão acerca da definição de causalidade que justifica o modelo causal de Neyman-Rubin, base para compreensão dos métodos empregados na avaliação de impacto. No capítulo 3, descreve-se o método de pareamento com escore de propensão, detalhando suas principais hipóteses e justificando sua utilização. No capítulo 4, apresenta-se o método de diferenças em diferenças e sua extensão para o caso de diversos períodos com um modelo de painel de efeitos fixos. No capítulo 5, resgata-se esta introdução e explora-se a relação da pobreza com as questões relativas ao transporte e como as políticas de transporte podem atacar o referido problema e promover o desenvolvimento. Por fim, no capítulo 6, serão apresentados os resultados da aplicação dos métodos, bem como uma discussão dos resultados.

2 Referencial Teórico-Metodológico

O quadro teórico apresentado neste capítulo é hoje a abordagem dominante na análise de efeitos causais (IMBENS; WOOLDRIDGE, 2009). Buscou-se um ponto de vista teórico, visto que tal teoria conforma a base para compreensão dos métodos de inferência causal utilizados em estudos observacionais — não apenas aqueles implementados neste trabalho. Optou-se por uma apresentação geral do modelo e, nesse sentido, as especificidades do caso da avaliação da tarifa zero serão abordadas quando da apresentação dos resultados.

Quanto ao modelo de Neyman-Rubin, sua gênese data de 1923 com o trabalho de Jerzy Neyman e em 1925 com Ronald Fisher, embora o primeiro tenha permanecido relativamente desconhecido por décadas (BOX-STEFFENSMEIER; BRADY; COLLIER, 2008). Esses trabalhos foram realizados apenas no âmbito dos experimentos aleatórios e somente algumas décadas depois, com os trabalhos de Donald Rubin, sobretudo no início da década de 1970, é que a abordagem se popularizou e foi estendida para contextos observacionais — em que se encaixam a maioria das investigações nas ciências sociais (IMBENS; WOOLDRIDGE, 2009). Em decorrência disso, o modelo de Neyman-Rubin também é chamado de modelo causal de Rubin ou ainda modelo de Neyman-Rubin-Holland, por causa das contribuições deste último para a síntese da abordagem em termos mais gerais (BOX-STEFFENSMEIER; BRADY; COLLIER, 2008).

A terminologia utilizada na abordagem de avaliação de impacto provém de estudos farmacológicos, em que os pesquisadores estão interessados no efeito de determinado tratamento (medicamentoso) (CAMERON; TRIVEDI, 2005). Tem-se uma população de N indivíduos aos quais nos referimos como *unidades*, as quais podem ser pessoas, firmas, famílias, municípios, e assim por diante. Deseja-se avaliar o impacto de dois tratamentos sobre as unidades, a saber, implementar determinada política social ou não implementar. Aqui, só é possível determinar o efeito de uma causa (implementar a política) quando a comparamos com outra “causa”, uma situação alternativa (não implementar a política), então o efeito de implementar a política é relativo à situação contrafactual em que a unidade não implementa a mesma. Denomina-se o grupo de unidades que adotam a política de grupo de tratamento e as demais consistem no grupo de controle. A política social a ser avaliada é denominada intervenção, ou apenas tratamento.

Holland (1986) chama atenção para as diferenças entre a chamada inferência associativa e a causal. O termo inferência associativa refere-se à realização de inferências estatísticas sobre parâmetros que associem diferentes características de indivíduos de uma determinada

população (HOLLAND, 1986). Em outras palavras, consiste em buscar uma conexão, como uma medida de correlação, entre duas variáveis aleatórias a qual indique que elas caminham juntas. Nesse sentido, Holland (1986) caracteriza a inferência associativa como sendo estatística descritiva.

Conforme enunciado na introdução deste trabalho, o estabelecimento de uma relação causal exige condições mais fortes do que a pura associação. Nesse sentido, os modelos estatísticos apropriados precisam de uma estrutura mais elaborada, que levem em consideração a diferença entre o que se entende por relação de causa e uma simples conexão entre duas variáveis e é isso que se pretende explorar neste capítulo, com a apresentação do modelo de Neyman-Rubin.

Box-Steffensmeier, Brady e Collier (2008) identificam quatro principais aspectos que compõem o modelo de Neyman-Rubin: (a) o primeiro é a definição de efeito causal calcada na abordagem contrafactual, em que a direção de causalidade costuma ser determinada via precedência temporal; (b) a hipótese de não interferência entre as unidades de tratamento; (c) independência entre a seleção para o tratamento e a variável de resultado e, por fim (d) métodos para garantir que ocorra (c) quando a hipótese de não interferência é válida. Esses quatro componentes serão discutidos ao longo das próximas seções.

2.1 Definição de efeito causal no modelo de Neyman-Rubin

2.1.1 Resultados potenciais

A abordagem contrafactual de causalidade consiste em comparar duas realidades distintas, uma em que ocorre a intervenção e outra em que ela não ocorre e entende-se que o efeito resulta de uma manipulação sobre as unidades, a qual denominamos *tratamento*. Desse modo, no caso mais simples tem-se dois tratamentos: a presença de intervenção e a ausência dela. Essa é uma importante questão a ser levantada: a natureza dos fatos que podem ser considerados candidatos a causa no modelo de Rubin.

Holland (1986) defende que nem tudo pode ser candidato a causa — para ele, causa é aquilo que poderia ser um tratamento em um experimento, é uma *manipulação* (ou ausência de) sobre as unidades, mesmo que condições materiais e éticas façam com que a execução de tal experimento não seja factível e ele não passe de um exercício hipotético. Portanto, um candidato a causa não pode ser uma atribuição individual (cor, gênero, localização), pois isso não seria passível de manipulação, mesmo em uma situação hipotética. Quando se considera uma atribuição individual, é possível delinear uma associação entre a dada atribuição e uma variável-resposta, mas não estabelecer uma relação causal. Evidentemente, nem sempre a separação entre o que são causas e o que são atributos é

tão simples.

Agora, considere uma unidade i da população e duas manipulações ou tratamentos possíveis dessa unidade, representadas sob forma de uma variável binária T_i : a existência de tratamento ($T_i = 1$) e uma situação de controle ($T_i = 0$). Defina-se, para cada indivíduo i , um par de resultados potenciais, $(Y_i(0), Y_i(1))$. $Y_i(1)$ é o valor da variável de resultado caso o indivíduo i receba tratamento 1 e $Y_i(0)$ é o valor da variável de resultado para o mesmo indivíduo caso ele receba tratamento 0. Em geral, essas variáveis de resultado $Y_i(0)$ e $Y_i(1)$ podem ser contínuas, discretas ou variar em um intervalo limitado de valores (CAMERON; TRIVEDI, 2005).

Deseja-se estabelecer uma relação causal entre ambas, ou seja, comparar o valor da variável de resultado Y em uma unidade que recebe tratamento 1 com o valor da variável de resultado para a mesma unidade caso ela tivesse recebido tratamento 0. Desse modo, a quantidade de interesse é uma comparação do tipo

$$C := Y_i(1) - Y_i(0) \quad (2.1)$$

entre duas possíveis realidades que, evidentemente, não podem ser observadas simultaneamente (ROSENBAUM; RUBIN, 1983). C é o efeito causal do tratamento 1 (relativo ao tratamento 0) em i medido por Y (HOLLAND, 1986). O resultado realizado, que é o que pode ser efetivamente observado pelo pesquisador, será denotado como Y_i . Este pode ser definido por

$$Y_i := Y_i(T_i) = Y_i(0)(1 - T_i) + Y_i(1)T_i = \begin{cases} Y_i(0) & \text{se } T_i = 0 \\ Y_i(1) & \text{se } T_i = 1 \end{cases} \quad (2.2)$$

A distinção entre os resultados potenciais, $Y_i(1)$ e $Y_i(0)$ e o resultado realizado Y_i é a marca registrada das análises econométricas e estatísticas de efeitos de tratamento (IMBENS; WOOLDRIDGE, 2009).

Uma vantagem da formulação do problema a partir de resultados potenciais, em relação àquelas que tratam apenas dos resultados realizados, é que ela permite que o efeito causal C seja definido em nível individual, sem hipóteses a respeito de sua distribuição ou forma funcional. Então, não é necessário determinar de antemão se o efeito é constante ou varia entre os indivíduos da população. Além disso, C é definido antes de determinar o procedimento de seleção — não é necessário acrescentar hipóteses de endogeneidade e exogeneidade do mecanismo de seleção para o tratamento. As propriedades estatísticas dos resultados ou do método de seleção tornam-se uma questão separada da definição de causa (IMBENS; WOOLDRIDGE, 2009).

Quando $C \neq 0$, diz-se que o tratamento tem um efeito líquido. Em um contexto mais geral de determinação de causa pela abordagem contrafactual, duas condições devem ser satisfeitas para que seja possível uma inferência causal. A primeira é que o tratamento deve ser associado a um efeito líquido $C \neq 0$ e a segunda é que a ausência de tratamento deve estar associada a um efeito líquido nulo, isto é, $C = 0$. O modelo de Rubin dá ênfase à primeira condição, então um efeito líquido refere-se à validade da afirmação “se o tratamento ocorre, então o efeito líquido ocorre”.

2.1.2 O problema fundamental da inferência causal

Note que, para cada indivíduo, é impossível observar $Y_i(1)$ e $Y_i(0)$, o que consiste no denominado *problema fundamental da inferência causal*¹ (HOLLAND, 1986). Nesse contexto, um dos principais desafios é inferir o que teria acontecido com o indivíduo tratado caso ele não tivesse recebido o tratamento a fim de realizar a comparação, o que caracteriza a avaliação de impacto como um problema de dados faltantes (ROSENBAUM; RUBIN, 1983). Há diferentes saídas para esse problema, geralmente sob forma de hipóteses adicionais ao modelo.

Em contrapartida, o modelo de resultados potenciais apresenta uma vantagem importante no fato de que a origem da incerteza dos estimadores é clara, ou seja, mesmo que uma grande quantidade de dados esteja disponível, sabe-se que o efeito causal é incerto diante do fato de apenas um resultado potencial ser observado (IMBENS; WOOLDRIDGE, 2009). Deve-se salientar que os resultados potenciais $Y_i(1)$ e $Y_i(0)$ são quantidades não estocásticas.

Uma maneira de resolver o problema de dados faltantes seria encontrar uma outra unidade j o mais similar possível à unidade i para fins de comparação. Para tal, é necessário determinar se a definição de efeito causal a partir de tal comparação é compatível com a definição teórica apresentada até aqui, que levava em consideração a diferença entre resultados potenciais apenas em um dado indivíduo.

Para simplificar a notação, considere apenas duas unidades, i e j . Pode-se definir o efeito causal de maneira intuitiva comparando a diferença na variável de resultado quando uma unidade recebe tratamento e outra não. Note que quatro diferentes realidades são possíveis: aquela em que apenas o indivíduo i é tratado e j não, aquela em que j é tratado e i não, aquela em que ambos os indivíduos não são tratados e aquela em que ambos são tratados. Formalmente, tem-se: $(T_i = 1, T_j = 0)$, $(T_i = 0, T_j = 1)$, $(T_i = 0, T_j = 0)$ e $(T_i = 1, T_j = 1)$, mas apenas as duas primeiras fornecem informações relevantes para o efeito causal (BOX-STEFFENSMEIER; BRADY; COLLIER, 2008).

¹ *Fundamental problem of causal inference* (HOLLAND, 1986).

Tabela 1 – Definições teóricas de efeito causal quando se dispõe de duas unidades i e j .

Indivíduo i
$C_i(T_j = 1) = Y_i(1, 1) - Y_i(0, 1)$
$C_i(T_j = 0) = Y_i(1, 0) - Y_i(0, 0)$
Indivíduo j
$C_j(T_i = 1) = Y_j(1, 1) - Y_j(1, 0)$
$C_j(T_i = 0) = Y_j(0, 1) - Y_j(0, 0)$

Fonte: Adaptado de Box-Steffensmeier, Brady e Collier (2008).

Considere agora as variáveis de resultado *potencial* do indivíduo i adicionando uma coordenada que se refere ao indivíduo j . Então, os resultados das quatro realidades supracitadas são denotados por $Y_i(0, 0)$, $Y_i(0, 1)$, $Y_i(1, 0)$ e $Y_i(1, 1)$ ², em que a primeira coordenada se refere à manipulação sobre o indivíduo i e a segunda coordenada se refere à manipulação sobre o indivíduo j . Em princípio, não há razão para crer que a variável de tratamento de um indivíduo não afete o resultado do outro, isto é, que $Y_i(0, 1) = Y_i(0, 0)$ e $Y_i(1, 1) = Y_i(1, 0)$. Para ver isso, imagine que Igor (i) e Júlia (j) queiram prestar vestibular e que ambos decidam participar de um curso preparatório para a prova. Considere que esse curso é o tratamento, de modo que ocorre $Y_i(1, 1)$ e que, em decorrência disso, os estudantes decidam montar um grupo de estudos para compartilhar conhecimentos e tirar dúvidas. Caso Júlia não tivesse participado do curso preparatório (situação $Y_i(1, 0)$), não haveria grupo de estudos, e o desempenho de Igor seria pior nessa situação do que na primeira, mesmo ele ainda escolhendo participar do curso.

Há duas maneiras plausíveis de definir o impacto sobre i :

- a) $C_i(T_j = 0) = Y_i(1, 0) - Y_i(0, 0)$, que é a diferença entre os resultados potenciais de i quando j não é tratado e
- b) $C_i(T_j = 1) = Y_i(1, 1) - Y_i(0, 1)$, que é a diferença entre os resultados potenciais de i quando j é tratado.

De maneira análoga, os resultados potenciais associados ao indivíduo j são $Y_j(0, 0)$, $Y_j(0, 1)$, $Y_j(1, 0)$ e $Y_j(1, 1)$ e é possível definir $C_j(T_i = 0)$ e $C_j(T_i = 1)$. A tabela 1 apresenta tais possíveis definições teóricas de efeito causal. Note que todas elas comparam diferentes resultados *para um mesmo indivíduo*, isto é, mesmo que, por exemplo, $C_i(T_j = 0)$ leve em consideração o tratamento de j , computa-se a diferença nas variáveis

² De maneira similar aos resultados potenciais apresentados no início da seção: $Y_i(0, 0)$ é interpretado como o valor da variável de resultado do indivíduo i caso i não seja tratado e caso o indivíduo j não seja tratado; $Y_i(0, 1)$ é o valor da variável de resultado do indivíduo i caso i não seja tratado e caso o indivíduo j seja tratado, e assim por diante.

de resultado potencial associadas ao indivíduo i (o que evidentemente ainda não resolve o problema fundamental da inferência causal).

Para evitar o problema de comparar situações que ocorrem em realidades diferentes, impossíveis de observar simultaneamente, deseja-se recorrer a *definições observáveis* de efeito causal, isto é, definições que não envolvem conhecer resultados potenciais de tratamentos diferentes associados a um mesmo indivíduo. São duas as possibilidades, a saber, os cenários $(T_i = 1, T_j = 0)$ e $(T_i = 0, T_j = 1)$, cujas respectivas medidas de impacto são

- a) $C^*(1, 0) := Y_i(1, 0) - Y_j(1, 0)$, diferença entre os resultados de i e j quando o primeiro recebe tratamento e o último não; e
- b) $C^*(0, 1) := Y_i(0, 1) - Y_j(0, 1)$, diferença entre os resultados de i e j quando este recebe tratamento e aquele não.

Note que as medidas C^* comparam indivíduos diferentes, contrastando com as definições teóricas apresentadas. Uma questão natural é: como relacionar as definições acima, as quais o pesquisador pode observar, com as definições teóricas da tabela 1, que satisfazem as condições necessárias para estabelecer a definição de efeito causal? Pretende-se explorar algumas respostas a essa pergunta.

Caso os indivíduos sejam idênticos, tem-se a hipótese denominada *homogeneidade*³ das unidades da amostra (HOLLAND, 1986), portanto

$$\begin{aligned} Y_j(1, 0) &= Y_i(0, 1) \text{ e} \\ Y_j(0, 1) &= Y_i(1, 0). \end{aligned} \tag{2.3}$$

Em que a primeira equação lê-se: resultado potencial de j quando i é tratado e j não é igual ao resultado potencial de i quando j é tratado e i não; e a segunda: o resultado potencial de j quando este é tratado e i não é igual ao resultado potencial de i quando este é tratado e j não é⁴.

Segundo a *hipótese de não interferência entre as unidades* (SUTVA)⁵, não há interferência entre as diferentes unidades da amostra no que se refere às variáveis de tratamento e resultado: o fato de o indivíduo i receber ou não tratamento não influencia a variável de resultado para o indivíduo j . Desse modo, sob a SUTVA, existe um único valor para $Y_i(0)$ e $Y_i(1)$, isto é,

$$\begin{aligned} Y_i(0, 1) &= Y_i(0, 0) \text{ e} \\ Y_i(1, 0) &= Y_i(1, 1)^6. \end{aligned} \tag{2.4}$$

³ *Unit Homogeneity.*

⁴ Também vale que $Y_j(1, 1) = Y_i(1, 1)$ e $Y_j(0, 0) = Y_i(0, 0)$.

⁵ *Stable Unit Treatment Value Assumption (SUTVA).*

Além disso, (2.4) garante que os tratamentos estejam plenamente representados por (2.2) (ROSENBAUM; RUBIN, 1983). Primeiro, considerando apenas a hipótese de não interferência (2.4), tem-se que o efeito causal para o indivíduo i quando o indivíduo j não é tratado é igual ao efeito causal para o indivíduo i quando o indivíduo j é tratado. Formalmente:

$$\begin{aligned} C_i(T_j = 0) &:= Y_i(1, 0) - Y_i(0, 0) \\ &= Y_i(1, 1) - Y_i(0, 1) \\ &= C_i(T_j = 1). \end{aligned} \tag{2.5}$$

Em que a segunda igualdade decorre de (2.4) e a terceira da definição de $C_i(T_j = 1)$. Analogamente, efeito causal para o indivíduo j quando o indivíduo i não é tratado é igual ao efeito causal para o indivíduo j quando o indivíduo i é tratado, isto é,

$$\begin{aligned} C_j(T_i = 0) &:= Y_j(0, 1) - Y_j(0, 0) \\ &= Y_j(1, 1) - Y_j(1, 0) \\ &= C_j(T_i = 1). \end{aligned} \tag{2.6}$$

Desse modo, a SUTVA (2.4) permite que se escreva $C_i = C_i(T_j = 0) = C_i(T_j = 1)$ e $C_j = C_j(T_i = 0) = C_j(T_i = 1)$. Adicionando a hipótese de homogeneidade (2.3), é possível relacionar essas definições teóricas às definições observáveis de efeito causal. Então a medida observável de efeito causal de quando i recebe tratamento e j não recebe ($C^*(1, 0)$) é igual ao efeito causal sobre o indivíduo i e também ao efeito causal sobre o indivíduo j , isto é,

$$\begin{aligned} C^*(1, 0) &:= Y_i(1, 0) - Y_j(1, 0) \\ &= Y_i(1, 0) - Y_i(0, 1) \\ &= Y_i(1, 0) - Y_i(0, 0) \\ &= C_i(T_j = 0) \\ &= C_i \end{aligned} \tag{2.7}$$

em que a segunda igualdade decorre de da homogeneidade (2.3) e a terceira decorre da SUTVA (2.4). Analogamente,

$$\begin{aligned} C^*(1, 0) &:= Y_i(1, 0) - Y_j(1, 0) \\ &= Y_j(0, 1) - Y_j(1, 0) \\ &= Y_j(1, 1) - Y_j(1, 0) \\ &= C_j(T_i = 1) \\ &= C_j, \end{aligned} \tag{2.8}$$

⁶ E ainda $Y_j(0, 0) = Y_j(1, 0)$ e $Y_j(0, 1) = Y_j(1, 1)$.

donde conclui-se que $C_i(T_j = 0) = C_j(T_i = 1)$. De maneira simétrica, obtém-se que $C^*(0, 1) = C_i(T_j = 1) = C_j(T_i = 0)$ e ambas as medidas agora podem ser diretamente observáveis. Mais ainda, quando valem a SUTVA e a homogeneidade das unidades, todas essas medidas de causalidade são iguais, isto é, $C^*(0, 1)$, $C^*(1, 0)$, C_i e C_j , são iguais. Holland (1986) denomina essa saída para o problema fundamental da inferência causal de “solução científica”.

A hipóteses de homogeneidade das unidades e de não interferência, embora plausíveis em determinados contextos, não são diretamente testáveis e não é possível ter absoluta certeza de que são válidas. A primeira questão prática com as definições acima é a plausibilidade de se encontrar uma unidade j tão similar a i , a ponto de tornar a comparação direta entre as duas unidades uma boa medida do efeito causal. Enquanto em um experimento de laboratório o objeto de estudo pode ser totalmente controlado pelo pesquisador, nas ciências sociais a homogeneidade das unidades tende a ser uma hipótese forte. Quando ela não é satisfeita, as diferenças entre indivíduos podem ter uma influência significativa na variável de resultado, independente de receber tratamento ou não, mascarando seu real efeito.

Por exemplo, imagine que se deseja avaliar o impacto de um programa de reforço de matemática oferecido aos alunos de uma escola. Suponha que Igor (i) tenha maior dificuldade em matemática e seja selecionado para participar do programa e que sua colega Júlia (j) tenha facilidade com a disciplina e esteja no grupo de controle — situação ($T_i = 1$, $T_j = 0$). É possível que, ao comparar o desempenho de Igor após a participação do programa ($Y_i(1, 0)$) com o de Júlia ($Y_j(1, 0)$), o efeito estimado ($C^*(1, 0)$) fosse baixo ou até mesmo negativo se comparado ao efeito real, $C_i(T_j = 0)$. Isso ocorreria pois as notas de Igor *caso ele não tivesse participado do programa* ($Y_i(0, 0)$) provavelmente teriam sido mais baixas que as notas de Júlia $Y_j(1, 0)$, utilizada como medida contrafactual. Nesse caso, a medida de efeito causal observada ($C^*(1, 0)$) e o efeito causal sobre Igor (C_i) são diferentes, em decorrência da violação da hipótese de homogeneidade. A tabela 2 sintetiza a conexão entre as definições teóricas apresentadas (por simplicidade, foram incluídas apenas aquelas relacionadas à unidade i) e as definições observáveis a elas associadas.

2.1.3 A solução estatística

Embora as hipóteses de homogeneidade e não interferência sejam suficientes para relacionar definições teóricas e observáveis de efeito causal, cabe indagar se são necessárias (BOX-STEFFENSMEIER; BRADY; COLLIER, 2008). Nesse contexto, mantendo a hipótese de não interferência, tem-se $C_i(T_j = 1) = C_i(T_j = 0) = C_i$ e $C_j(T_i = 1) = C_j(T_i = 0) = C_j$.

Tabela 2 – Relação entre as definições teóricas de causalidade e as definições observáveis.

Definição observacional	Homogeneidade	SUTVA	Definição Teórica
$C^*(1, 0) = Y_i(1, 0) - Y_j(1, 0)$	$Y_j(1, 0) = Y_i(0, 1)$	$Y_i(0, 1) = Y_i(0, 0)$ $Y_i(1, 0) = Y_i(1, 1)$	$Y_i(1, 0) - Y_i(0, 0) = C_i(T_j = 0)$ $Y_i(1, 1) - Y_i(0, 1) = C_i(T_j = 1)$
$C^*(0, 1) = Y_i(0, 1) - Y_j(0, 1)$	$Y_j(0, 1) = Y_i(1, 0)$	$Y_i(0, 1) = Y_i(0, 0)$ $Y_i(1, 0) = Y_i(1, 1)$	$Y_i(1, 0) - Y_i(0, 0) = C_i(T_j = 0)$ $Y_i(1, 1) - Y_i(0, 1) = C_i(T_j = 1)$

Fonte: Adaptado de (BOX-STEFFENSMEIER; BRADY; COLLIER, 2008).

Agora, considere a Média aritmética dos Efeitos Causais observáveis apresentados (MEC):

$$\begin{aligned}
MEC &= \frac{1}{2}(C^*(1, 0) + C^*(0, 1)) \\
&= \frac{1}{2}(Y_i(1, 0) - Y_j(1, 0)) + (Y_j(0, 1) - Y_i(0, 1)) \\
&= \frac{1}{2}(Y_i(1, 0) - Y_i(0, 1)) + (Y_j(0, 1) - Y_j(1, 0)) \\
&= \frac{1}{2}(Y_i(1, 0) - Y_i(0, 0)) + (Y_j(0, 1) - Y_j(0, 0)).
\end{aligned} \tag{2.9}$$

A segunda igualdade decorre das definições de $C^*(1, 0)$ e $C^*(0, 1)$, a terceira é um rearranjo dos termos e a quarta decorre da hipótese de não interferência (2.4). Então, podemos escrever

$$MEC = \frac{1}{2}(C_i + C_j). \tag{2.10}$$

Note que (2.10) é uma média dos impactos em i e j . Não é possível observá-la diretamente, entretanto, é possível selecionar qual dos indivíduos receberá tratamento de maneira aleatória e utilizar o efeito observado, $C^*(1, 0)$ ou $C^*(0, 1)$ como estimativa para MEC. Se esse experimento fosse repetido um grande número de vezes, o valor esperado de MEC seria igual ao efeito causal médio real (BOX-STEFFENSMEIER; BRADY; COLLIER, 2008).

A ausência da hipótese de homogeneidade pode trazer problemas: é possível que os indivíduos tratados e os não tratados sejam distintos a ponto de que comparação direta entre eles sequer seja válida — uma possibilidade é restringir a população inicial a um grupo semelhante em todas as características relevantes ao experimento, como propõem os métodos de pareamento abordados mais adiante. Além disso, o resultado acima depende da lei dos grandes números. Por isso, seria desejável ter uma amostra grande de indivíduos no estudo, de modo que se conheça a possibilidade de que o impacto medido entre os dois grupos seja meramente aleatória ou de que seja decorrente de alguma diferença real entre os dois grupos (que espera-se que seja o tratamento, desde que o experimento seja bem elaborado) (BOX-STEFFENSMEIER; BRADY; COLLIER, 2008).

Embora os resultados potenciais para um dado indivíduo não sejam estocásticos, pode-se considerar $Y(1)$ e $Y(0)$ como variáveis aleatórias, cujos valores realizados para o indivíduo i são $Y_i(0)$ e $Y_i(1)$. Generalizando o raciocínio que acaba de ser exposto, Rosenbaum e Rubin (1983) definem o Efeito Médio do Tratamento (EMT) como a média

$$EMT := \mathbb{E}(Y(1) - Y(0)), \quad (2.11)$$

em que $\mathbb{E}(\cdot)$ é o operador de esperança. Logo, EMT é o valor esperado do efeito causal $Y_i(1) - Y_i(0)$ sobre todos os indivíduos i da população. Essa é a “solução estatística” de Holland (1986), que implica que informações sobre o impacto podem ser obtidas comparando unidades diferentes, isto é, substitui-se o efeito impossível de observar do tratamento 1 sobre o indivíduo i pela estimativa do efeito causal médio sobre a população inteira.

Normalmente, busca-se Efeito Médio do Tratamento sobre os Tratados (EMTT), definido em Cameron e Trivedi (2005) tomando a esperança de EMT restrita ao grupo de tratados:

$$EMTT := \mathbb{E}(Y(1) - Y(0)|T = 1). \quad (2.12)$$

2.2 Dados experimentais

Considere uma amostra de n indivíduos da população. Denote por n_1 o número de indivíduos tratados da amostra e $n_0 = n - n_1$, o número de indivíduos não tratados, candidatos a compor um grupo de controle. Quando a seleção dos grupos é feita de maneira aleatória, a participação do programa (variável de tratamento T) ignora o possível impacto do mesmo sobre os indivíduos, ou seja, um determinado indivíduo não é incluído no programa com a justificativa de sua participação apresentar grande benefício em potencial nem excluído porque o resultado esperado é insatisfatório (CAMERON; TRIVEDI, 2005). Formalmente,

$$Y(1), Y(0) \perp T, \quad (2.13)$$

que implica $\mathbb{E}(Y(0)|T = 1) = \mathbb{E}(Y(0)|T = 0)$. A hipótese (2.13) é denominada *hipótese de independência condicional* (CAMERON; TRIVEDI, 2005). Aplicando a relação (2.13) à equação (2.12), obtemos:

$$EMTT := \mathbb{E}(Y(1)|T = 1) - \mathbb{E}(Y(0)|T = 1) = \mathbb{E}(Y(1)|T = 1) - \mathbb{E}(Y(0)|T = 0). \quad (2.14)$$

Isso significa que as características observadas de ambos os grupos são semelhantes e podemos comparar as variáveis de resultado diretamente (PEIXOTO *et al.*, 2012). À

medida que a amostra aumenta, características específicas dos indivíduos que poderiam confundir o resultado observado se equilibram entre os dois grupos, de modo que, quando a aleatorização é possível, o efeito causal médio pode ser estimado com grande precisão (HOLLAND, 1986). Agora, considere os subconjuntos da amostra $I_1 = \{i \in \{1, \dots, n\} \mid T_i = 1\}$ e $I_0 = \{i \in \{1, \dots, n\} \mid T_i = 0\}$. Segue da discussão acima que o efeito médio do tratamento pode ser estimado pela diferença entre as médias aritméticas da variável de tratados de ambos os grupos:

$$\frac{1}{n_1} \sum_{i \in I_1} Y_i - \frac{1}{n_0} \sum_{i \in I_0} Y_i. \quad (2.15)$$

É também possível estimar o efeito direto do tratamento sobre a variável de resultado através de uma regressão linear, cujo formato básico seria

$$Y_i = X_i \alpha + \beta_i T_i + \varepsilon_i. \quad (2.16)$$

Em (2.16), o parâmetro β_i captaria o efeito do programa para o indivíduo i e ε_i é um componente idiossincrático que influencia os resultados potenciais do indivíduo i (KHANDKER; KOOLWAL; SAMAD, 2009). Quando adotamos a hipótese $\beta_i = \beta \forall i \in \{1, \dots, n\}$, ou seja, de que o efeito do programa sobre os indivíduos é homogêneo, podemos reescrever o modelo (2.16) como

$$Y = X\alpha + \beta T + \varepsilon. \quad (2.17)$$

Conforme apresentado, o único elemento estocástico do modelo é a seleção do tratamento, fazendo com que a abordagem difira do contexto econométrico usual, em que o resultado é uma variável aleatória, mas os regressores não (BOX-STEFFENSMEIER; BRADY; COLLIER, 2008). As propriedades do estimador de impacto em (2.16) são examinadas em (FREEDMAN, 2008), mas considera-se que esta discussão está fora do escopo deste trabalho.

2.3 Hipótese de valor estável de unidade-tratamento

Por enquanto, foram expostas duas maneiras de lidar com o problema fundamental da inferência causal; por um lado, assumir não interferência (SUTVA) e homogeneidade entre as unidades do estudo e, por outro, adotar a SUTVA e controlar as diferenças entre as unidades através de um mecanismo de seleção aleatória do tratamento. Nota-se que a

hipótese de não interferência é essencial — e também motivo de preocupação quando se lida com cenários econômicos (IMBENS; WOOLDRIDGE, 2009).

SUTVA implies [...] that each unit is a separate, isolated possible world that is unaffected by what happens to the other units. SUTVA is the master assumption that makes controlled or randomized experiments a suitable solution to the problem of making causal inferences. SUTVA ensures that treatment and control units really do represent the closest possible worlds to one another except for the difference in treatment. (BOX-STEFFENSMEIER; BRADY; COLLIER, 2008, p. 265)

Box-Steffensmeier, Brady e Collier (2008) chamam atenção para a diferença entre a SUTVA e a hipótese padrão de independência em modelos de regressão. Enquanto o MQO fornece estimadores não enviesados mesmo quando não há independência dos termos de erro⁷, o mesmo não ocorre na ausência da hipótese de não interferência (SUTVA). Se esta falha, incorre-se em má especificação do modelo e, conseqüentemente, risco de viés. Desse modo, pode ser difícil generalizar os resultados de um experimento ou mesmo interpretá-los⁸ (BOX-STEFFENSMEIER; BRADY; COLLIER, 2008).

Imbens e Wooldridge (2009) assinalam duas possíveis soluções para os casos em que a SUTVA não é uma hipótese razoável e interações entre indivíduos podem provocar viés nas estimativas de impacto. A primeira é redefinir a unidade de interesse, isto é, quando as interações entre indivíduos ocorrem em âmbito local, agregá-los em uma mesma classe pode tornar a hipótese mais plausível (por exemplo, em vez de avaliar o impacto para um estudante, avaliar o impacto para a escola inteira). Todavia, essa estratégia pode custar em termos de precisão das estimativas e pode não ser satisfatória, dependendo do contexto. Outra possibilidade sugerida pelos autores é a modelagem direta de tais interações, levando em consideração os diferentes grupos em que elas podem ocorrer, suas respectivas sobreposições e a intensidade com que ocorrem.

2.4 Viés de seleção

Quando o grupo de tratados não é selecionado aleatoriamente e a hipótese de homogeneidade não é razoável, não se pode admitir independência condicional (2.13)⁹. Além disso,

⁷ A ausência de viés no MQO exige apenas a hipótese de exogeneidade dos regressores, isto é, de que $\mathbb{E}(\varepsilon|Z) = 0$, em que Z é a matriz de regressores e ε é o termo de erro. Assim, para o modelo básico dado por $y = Z\beta + \varepsilon$ (e supondo que esta é a especificação correta), o estimador de MQO para β é $\hat{\beta} = (Z^T Z)^{-1} Z^T y = (Z^T Z)^{-1} Z^T (Z\beta + \varepsilon) = \beta + (Z^T Z)^{-1} Z^T \varepsilon$. Tomando a esperança, $\mathbb{E}(\hat{\beta}) = \beta + \mathbb{E}((Z^T Z)^{-1} Z^T \varepsilon) = \beta$. Então, a ausência de viés desse estimador não depende de os erros serem independentemente distribuídos (DAVIDSON; MACKINNON, 2004).

⁸ Novamente, formalização da discussão do viés no modelo de regressão no contexto de experimentos aleatórios é tratada em Freedman (2008).

⁹ Portanto, em geral $\mathbb{E}(Y(0)|T = 1) \neq \mathbb{E}(Y(0)|T = 0)$.

caso a seleção não seja aleatória — seja porque a participação é voluntária e determinada por características não observadas, seja porque ela é definida segundo alguns critérios de elegibilidade — não se pode garantir a exogeneidade de T , ou seja, $\mathbb{E}(\varepsilon|T) \neq 0$. A exogeneidade dos regressores é uma hipótese fundamental para garantir que os estimadores de mínimos quadrados não sejam enviesados. Esse possível viés é denominado *viés de seleção* (ou viés de autoseleção para Peixoto *et al.* (2012)). “Sua denominação decorre do fato de que são os próprios tratados que se selecionam para participar do programa” (PEIXOTO *et al.*, 2012, p. 42). Quando isso ocorre, ao calcular o efeito médio do tratamento sobre os tratados, obtém-se

$$\begin{aligned}
 EMTT^* &= \mathbb{E}(Y(1)|T = 1) - \mathbb{E}(Y(0)|T = 0) \\
 &= \mathbb{E}(Y(1)|T = 1) - \mathbb{E}(Y(0)|T = 1) + \mathbb{E}(Y(0)|T = 1) - \mathbb{E}(Y(0)|T = 0) \\
 &= EMTT + \mathbb{E}(Y(0)|T = 1) - \mathbb{E}(Y(0)|T = 0) \\
 &= EMTT + V_s,
 \end{aligned} \tag{2.18}$$

em que a segunda igualdade é obtida ao somar e subtrair o termo $\mathbb{E}(Y(0)|T = 1)$ e V_s é o viés de seleção. Heckman, Ichimura e Todd (1997) referem-se ao viés de seleção como sendo “decorrente de diferenças em [características] não observáveis”, isto é, a participação voluntária pode refletir diferenças sistemáticas entre os grupos que têm influência sobre a variável de resultado e, portanto, sobre a estimativa de impacto. Por exemplo, pessoas que se inscrevem voluntariamente em um programa de capacitação podem ser, em média, mais pró-ativas que aquelas que não se inscrevem. Além do programa de capacitação, elas podem tomar mais iniciativas para buscar oportunidades de emprego e expandir seus conhecimentos, aferindo maior sucesso profissional. Então, o melhor desempenho dessas pessoas seria atribuído integralmente ao programa de capacitação, inflando o impacto estimado.

Por questões éticas e técnicas, a aleatorização nem sempre é possível, principalmente no que tange à realização de políticas públicas e o caso da tarifa zero não é exceção. Faz-se necessário buscar métodos para inferência causal utilizando dados observacionais, isto é, em que o pesquisador não tem controle sobre a associação de cada unidade a um tratamento (1 ou 0). Em termos de mimetizar um experimento controlado, trata-se de construir um grupo de controle *a posteriori* que represente bem a situação contrafactual, $\mathbb{E}(Y(0)|T = 1)$. Assim, $V_s = 0$ e é possível obter uma estimativa melhor do impacto.

Segundo Box-Steffensmeier, Brady e Collier (2008), quando adota-se a SUTVA, há três maneiras de assegurar a independência condicional: (a) experimentos controlados com unidades homogêneas; (b) experimentos estatísticos com seleção aleatória; (c) estudos

observacionais (também chamados de quase-experimentais).

Os métodos estudados ao longo deste trabalho dizem respeito à terceira alternativa. Nos estudos observacionais, as correções são feitas por meio de covariáveis, as quais garantem independência condicional média entre a seleção para o tratamento e a variável de resultado. Todavia, no contexto quase-experimental não há meios de determinar exatamente as covariáveis relevantes e também não há como garantir a hipótese de independência condicional (BOX-STEFFENSMEIER; BRADY; COLLIER, 2008).

3 Pareamento

3.1 Mecanismo de seleção

A aleatorização dos grupos de tratamento e controle não é o único mecanismo de seleção disponível. Na verdade, há relativamente poucos experimentos práticos em economia, sejam eles com aleatorização simples ou até mesmo outros tipos, como estratificada, em pares, e assim por diante. Em todos esses casos, o pesquisador conhece a distribuição de T . Além da aleatorização, uma outra classe de mecanismo de seleção é aquela que mantém uma versão da hipótese de independência, mas em que as probabilidades de seleção não são mais uma função conhecida das covariáveis. Formalmente,

$$Y(0), Y(1) \perp T | X, \quad (3.1)$$

sendo X a matriz cujas colunas são X_i , os quais são vetores de características de cada indivíduo i . Note que (3.1), denominada *hipótese de ignorabilidade*¹, é uma condição de independência mais fraca do que a de independência condicional (2.13). Neste caso, supõe-se que as variáveis observadas contêm todas as informações sobre o resultado potencial do tratamento que o indivíduo utiliza para tomar a decisão de participar ou não do projeto. Então, (3.1) implica que não há viés de variável omitida desde que as características X sejam incluídas na regressão (ROSENBAUM; RUBIN, 1983). O método de pareamento se encaixa neste tipo de mecanismo de seleção.

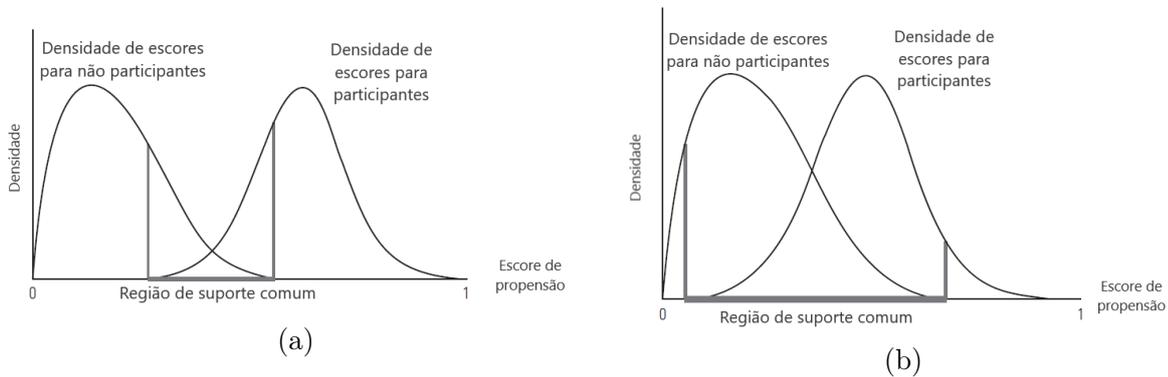
Uma terceira classe de mecanismo de seleção é aquela que relaxa a hipótese de ignorabilidade (3.1). Métodos como regressão descontínua, variáveis instrumentais e diferenças em diferenças são maneiras comuns de lidar com as implicações de que haja algum grau de dependência entre a seleção de tratamento e as variáveis de resultado (IMBENS; WOOLDRIDGE, 2009).

3.2 Principais hipóteses

O método de pareamento consiste em encontrar pares semelhantes formados por indivíduos do grupo tratado e do grupo de controle. Isto é, procura-se buscar dentre os indivíduos não tratados, aqueles que são similares aos indivíduos tratados no que se re-

¹ *Ignorability assumption*, também chamada de *unconfoundedness assumption*, seleção por observáveis, exogeneidade ou ainda independência condicional (IMBENS; WOOLDRIDGE, 2009).

Figura 1 – Ilustração da hipótese de suporte comum.



Fonte: (KHANDKER; KOOLWAL; SAMAD, 2009).

ferre às características pré-tratamento. Essas características são variáveis observadas pelo pesquisador (PEIXOTO *et al.*, 2012). No entanto, o que exatamente se entende por similaridade pode variar caso a caso e é uma questão que está sujeita a amplo debate na literatura (IMBENS; WOOLDRIDGE, 2009). O pareamento sustenta-se sobre duas hipóteses centrais: a primeira é a ignorabilidade (3.1); a segunda é denominada *hipótese de sobreposição*, ou suporte comum, que afirma que

$$0 < \Pr(T = 1 | X = x) < 1. \quad (3.2)$$

A condição (3.2) significa que existe, para todo valor realizado de X , uma probabilidade positiva de não participação (CAMERON; TRIVEDI, 2005). Desse modo, garante que haja indivíduos no grupo de controle com características observáveis similares aos tratados e elimina a separação completa entre tratados e não tratados por meio de alguma característica que prediz perfeitamente a participação (ou não participação); quando essa hipótese é válida, diz-se que há suporte comum.

A figura 1 ilustra situações com ausência (à esquerda) e presença (à direita) de suporte comum. Nas duas figuras, pode-se observar as densidades de probabilidade das características observadas X . Enquanto ao lado esquerdo quase não há sobreposição das curvas, isto é, o suporte comum é fraco, no lado direito há grande sobreposição das distribuições. Isso significa que a hipótese (3.2) é respeitada em quase toda a região de suporte das curvas e torna-se uma hipótese adequada.

De antemão, vale ressaltar que o método de pareamento é comumente utilizado para selecionar as amostras, assim como estimar o efeito causal. Quando utilizado para realizar a avaliação de impacto, geralmente está implícita a hipótese de que variáveis não observáveis não influenciam na decisão de participação do programa. Essa pode ser uma hipótese

muito forte, tornando necessário combinar outros métodos que resolvam o problema decorrente da influência de características não observáveis na decisão de participação, como uma abordagem de diferenças em diferenças (KHANDKER; KOOLWAL; SAMAD, 2009).

Para o fim deste trabalho, em que ele é utilizado apenas para selecionar as amostras e é combinado com diferenças em diferenças, a hipótese de ignorabilidade é prescindível, isto é, o método de diferenças em diferenças permite que se leve em consideração as características não observáveis que poderiam afetar os resultados e não é necessário supor que X contém toda informação relevante para a seleção de tratamento. No entanto, é mantida a hipótese de não interferência entre as unidades, SUTVA, (2.4) — o que implica que o tratamento de um indivíduo não afeta indiretamente as observações não tratadas (CAMERON; TRIVEDI, 2005).

Imbens e Wooldridge (2009) salientam que a ignorabilidade implica que X_i é um conjunto que deve explicar grande parte do indicador de tratamento, de modo que controlar o modelo por diferenças nos valores dessas covariáveis conduza a estimativas válidas de efeitos causais. Ademais, essa hipótese justifica o uso de modelos de regressão linear, se combinada com uma hipótese de linearidade da expectativa condicional dos resultados potenciais, dado X , isto é, a linearidade de $\mathbb{E}(Y(0)|X)$ e de $\mathbb{E}(Y(1)|X)$. Considerando que o efeito do tratamento é constante e igual a τ , que $Y_i(0) = \alpha + \beta' X_i + \varepsilon_i$ — e o resíduo capta a diferença de resultado provocada por idiosincrasias das observações na ausência de tratamento — o resultado observado Y_i , definido como em (2.2), pode ser escrito como

$$Y_i = \alpha + \tau \cdot T_i + \beta' \cdot X_i + \varepsilon_i. \quad (3.3)$$

Então, a ignorabilidade é equivalente à independência de ε_i e T_i condicional a X_i . No entanto, os mesmos autores demonstram preocupação com a possibilidade de viés decorrente da hipótese de linearidade, além de enorme sensibilidade dos resultados de regressões lineares nesse contexto. Uma maneira de lidar com isso seria combinar o *escore de propensão* com regressões lineares.

Existem métodos de pareamento exato, que são aqueles em que se busca escolher um indivíduo do grupo de controle com o mesmo vetor de características dos indivíduos do grupo tratado. Isso só tende a ser factível quando X não tem dimensão muito grande e caso as variáveis observadas sejam discretas e assumam um número reduzido de valores. Já os métodos de pareamento inexato operam de maneira a projetar o vetor de características X a uma medida de menor dimensão, $f(X)$. Rosenbaum e Rubin (1983) definem um *escore de balanceamento*² como uma função $b(X)$ das covariáveis observadas X tal que a

² *Balancing score.*

distribuição de X dado $b(X)$ é a mesma para os tratados e não tratados, isto é,

$$X \perp T | b(X). \quad (3.4)$$

O *escore de propensão*, por exemplo, é um tipo de escore de balanceamento em que $b(X)$ é um escalar.

Duas abordagens de pareamento são mais comuns, um tipo baseado no escore de propensão e as multivariadas, baseadas em alguma medida de distância, como a de Mahalanobis. Desde que as distribuições das covariáveis sejam elipsoidais (como a distribuição normal e a t), as propriedades estatísticas de ambos os tipos são desejáveis, com redução de vies em combinações lineares das covariáveis (BOX-STEFFENSMEIER; BRADY; COLLIER, 2008).

3.3 Escore de propensão

O escore de propensão é uma medida de distância unidimensional, o que confere maior simplicidade ao processo de pareamento, preservando suas propriedades mais desejáveis. O escore de propensão pode ser interpretado como a propensão à participação de um indivíduo no tratamento, dadas as suas características pré-tratamento. Define-se $\epsilon(X_i)$, o escore de propensão do indivíduo i , como

$$\epsilon(X_i) := \Pr(T_i = 1 | X_i) = \mathbb{E}(T_i | X_i). \quad (3.5)$$

Rosenbaum e Rubin (1983) demonstram que, se são válidas as hipóteses de ignorabilidade (3.1) e sobreposição (3.2), e também se os tratamentos dos diferentes indivíduos da amostra T_1, \dots, T_n , são independentes entre si, dadas as covariáveis X_1, \dots, X_n , o que é formalmente expresso por

$$\Pr(T_1, T_2, \dots, T_n | X_1, X_2, \dots, X_n) = \prod_{i=1}^n \epsilon(X_i)^{T_i} (1 - \epsilon(X_i))^{1-T_i}, \quad (3.6)$$

então essas hipóteses também são válidas para o escore de propensão, isto é,

$$0 < \Pr(T_i = 1 | \epsilon(X_i)) < 1 \quad (3.7)$$

e

$$Y_i(0), Y_i(1) \perp T_i | \epsilon(X_i). \quad (3.8)$$

Desse modo, em subpopulações com o mesmo valor para o escore de propensão, as co-variáveis são independentes do indicador de tratamento e, portanto, não provocam viés caso omitidas (IMBENS; WOOLDRIDGE, 2009). Então, Rosenbaum e Rubin (1983) demonstram ainda que o efeito médio do tratamento seria dado por³

$$EMT = \mathbb{E}(Y(1)|\epsilon(X_i), T_i = 1) - \mathbb{E}(Y(0)|\epsilon(X_i), T_i = 0) = \mathbb{E}(Y(1) - Y(0)|\epsilon(X_i)). \quad (3.9)$$

Para obter o EMTT, basta aplicar o operador de esperança em (3.9) condicionado a $T_i = 1$.

Esses resultados são extremamente relevantes pois deles conclui-se que é suficiente controlar o modelo apenas para diferenças no escore de propensão entre unidades tratadas e de controle, quando mantida a hipótese de ignorabilidade. Como a distribuição de $\epsilon(X_i)$ não é conhecida, é necessário estimá-la e é importante escolher um modelo apropriado para fazê-lo. A hipótese (3.6) sugere o uso modelos de resposta binária como probit ou logit e, no caso de dois tratamentos, esses dois modelos fornecem resultados similares (CALIENDO; KOPEINIG, 2008).

3.3.1 Modelo logit

Em geral, para associar ao vetor de características X um número real $p(X)$, utiliza-se modelos paramétricos, como logit e probit. Como este é o caso, isto é, requer-se associar um vetor de características X_i a um escore de propensão $\epsilon(X_i)$, o modelo logit será apresentado brevemente. O caso do modelo probit é inteiramente análogo, com exceção da função da distribuição do termo de erro.

A variável dependente do modelo, que é a variável de tratamento denotada por T_i , assume apenas dois valores, 0 e 1. Denota-se por \mathcal{P}_i a probabilidade da observação i assumir valor 1. Isto é, a probabilidade de o município i participar do programa com base nas características observadas (que é o conjunto de informação, denotado por Ω_i). Nesse sentido, \mathcal{P}_i representa uma expectativa condicional de T_i :

$$\mathcal{P}_i = \Pr(T_i = 1|\Omega_i) = \mathbb{E}(T_i|\Omega_i) \quad (3.10)$$

É necessário, portanto, impor a condição de que $0 \leq \mathbb{E}(T_i|\Omega_i) \leq 1$ — o que não é possível através de um modelo de regressão linear usual (DAVIDSON; MACKINNON, 2004). Para solucionar esse problema, especifica-se que

³ A rigor, Rosenbaum e Rubin (1983) demonstram esse resultado em um contexto mais geral, isto é, para um escore de balanceamento $b(X)$ qualquer, não apenas para o escore de propensão.

$$\mathcal{P}_i \equiv \mathbb{E}(T_i|\Omega_i) = F(X_i\beta) \quad (3.11)$$

em que $X_i\beta$ é uma função índice e $F(\cdot)$ uma função transformação. $F(\cdot)$ possui as mesmas propriedades de uma função de distribuição de probabilidade acumulada, a saber:

$$\begin{aligned} F(-\infty) &= 0 \\ F(\infty) &= 1 \\ f(x) &\equiv \frac{dF(x)}{dx} > 0. \end{aligned}$$

As propriedades acima garantem justamente que, mesmo que $X_i\beta$ assuma qualquer valor real, $F(X_i\beta)$ estará entre 0 e 1. Além disso, $F(x)$ não é linear, ou seja, alterações nos valores de X_i alteram $\mathbb{E}(T_i|\Omega_i)$ de maneira não linear (DAVIDSON; MACKINNON, 2004). No modelo logit, função F é uma função logística:

$$\Lambda(x) := \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x}, \quad (3.12)$$

cuja derivada é

$$\lambda(x) = \frac{e^x}{(1 + e^x)^2} = \Lambda(x)\Lambda(-x). \quad (3.13)$$

Note que (3.13) é simétrica em torno da origem e $\Lambda(-x) = 1 - \Lambda(x)$.⁴ No modelo logit, utiliza-se, no lugar de $F(\cdot)$ em (3.11), a função Λ :

$$\mathcal{P}_i = \Lambda(X_i\beta). \quad (3.14)$$

Já quando $F(\cdot)$ toma a forma da função de densidade de probabilidade acumulada da distribuição normal padrão, $\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}X^2} dX$, tem-se um modelo probit.

Após a implementação de um dos modelos acima, utiliza-se o vetor de valores ajustados $X\hat{\beta}$ como estimativa dos escores de propensão, $\hat{\epsilon}(X)$. Isso evita que haja concentração de muitas unidades nos valores extremos (muitas unidades aglomeradas perto de zero e um), incorrendo em um pareamento mais eficaz (BOX-STEFFENSMEIER; BRADY; COLLIER, 2008). Ressalta-se que estatísticas de teste como a estatística t e o coeficiente R^2

⁴ $1 - \Lambda(x) = 1 - \frac{1}{1+e^{-x}} = \frac{e^{-x}}{1+e^{-x}} = \Lambda(-x)$

ajustado não são de interesse nessa etapa, apenas a correlação entre X e T (KHANDKER; KOOLWAL; SAMAD, 2009).

3.3.2 Critérios de pareamento

Uma vez obtido o escore de propensão, existem diversas maneiras de realizar o pareamento. Khandker, Koolwal e Samad (2009) citam:

Vizinho mais próximo O critério do vizinho mais próximo⁵ consiste em buscar os M indivíduos com escore de propensão mais próximos de cada agente tratado. Ele pode ser realizado com ou sem reposição, isto é, uma mesma unidade de controle pode ser pareada com duas ou mais unidades tratadas distintas. Realizar o pareamento sem reposição pode resultar em selecionar unidades distantes da unidade tratada, incorrendo em piora nas estimativas de impacto, então só é indicado em casos em que há grande disponibilidade de candidatos a controle.

Pareamento pelo raio É possível que o vizinho mais próximo ainda assim tenha escore de propensão distante do indivíduo tratado. O pareamento pelo raio serve para evitar que isso ocorra. Ele funciona de maneira similar ao anterior, mas impondo um limite de tolerância para a distância máxima do indivíduo do grupo de controle escolhido. Comumente o pareamento é feito com reposição nesse caso, aumentando a quantidade de participantes não tratados descartados para compor o grupo de controle.

Estratificação Também chamada de pareamento por intervalos, consiste em particionar a região de suporte comum e calcular o impacto em cada estrato. Posteriormente, utiliza-se uma média ponderada de cada impacto para estimar o efeito causal médio da amostra completa.

Pareamento local linear ou *kernel* Quando os métodos anteriores são utilizados, é comum que apenas um pequeno grupo de indivíduos não tratados satisfaça os critérios necessários para compor o grupo de controle. O pareamento local linear e o pareamento *kernel* utilizam uma média ponderada dos não tratados para fabricar um par para cada indivíduo tratado. No pareamento *kernel*, os pesos dependem das distâncias entre os escores de propensão. O pareamento local linear estima uma regressão não paramétrica localmente ponderada.

Observe que a diferença entre escores de propensão é uma possível medida de distância entre as unidades, mas os critérios acima poderiam ser implementados com outras medidas

⁵ *Nearest neighbor*.

de distância, como uma métrica baseada em uma norma, a distância de Mahalanobis, e assim por diante.

3.3.3 Viés no escore de propensão

No capítulo 2, ao discorrer sobre o modelo de Neyman-Rubin e as hipóteses necessárias para definir causalidade sob a ótica dos resultados potenciais, foi mencionado que as técnicas utilizadas em estudos observacionais visam garantir alguma versão da hipótese de independência condicional média, dada em (3.1). Na prática, outras fontes de viés podem impedir a comparação direta entre tratados e não tratados: dados de má qualidade ou obtidos sob critérios distintos, métodos de avaliação mal empregados ou ainda a falha na hipótese de independência condicional, a qual não pode ser diretamente testada — e que provoca a emergência do viés de seleção. Assim sendo, pode-se esperar algum viés em qualquer estudo observacional e, evidentemente, este não pode ser observado. Cabe questionar qual a magnitude desse viés e também a importância relativa de seus diferentes componentes.

Heckman, Ichimura e Todd (1997) buscaram responder a essa pergunta. Os autores do estudo realizaram uma análise de sensibilidade de diferentes métodos de pareamento, utilizando dados de um programa de capacitação realizado nos EUA, o *Job Training Partnership Act* (JTPA). Os pesquisadores contavam com uma amostra aleatória dentre os participantes do programa, além de diferentes grupos de controle obtidos através de dados observacionais. O estudo compara os estimadores de impacto, os quais obtidos a partir de distintas formas de pareamento, entre si, além de compará-los com os resultados obtidos a partir da amostra aleatória. Desse modo, foi possível estimar o viés em que esses métodos incorriam, comparando-os com o grupo de controle resultante da aleatorização.

Primeiramente, os autores investigaram as diferentes fontes de viés quando se utiliza dados observacionais em contraste com as amostras aleatórias. Considere S_1 a região de suporte de X para os tratados, S_0 a região de suporte de X para o grupo de controle e S_{10} a região de suporte comum, isto é, sobreposição dos suportes dos dois grupos. Além disso, \setminus denota a operação de diferença usual de conjuntos. Considerando a figura 1, tem-se que S_1 e S_0 são as regiões em que as curvas do lado direito e esquerdo assumem valores positivos, enquanto S_{10} é a interseção de ambas. Convencionalmente, define-se o viés de maneira similar ao capítulo 2, como a diferença do resultado potencial de não tratamento entre o grupo de tratados (contrafactual) e o de controle (que é a medida observada), mas agora considerada sobre a distribuição das características observáveis X :

$$V = \int_{S_1} \mathbb{E}(Y(0)|X, T = 1)f(X|T = 1)dX - \int_{S_0} \mathbb{E}(Y(0)|X, T = 0)f(X|T = 0)dX. \quad (3.15)$$

Em que $f(\cdot)$ representa a função de densidade de probabilidade conjunta das covariáveis condicional a T . A partir da definição dada por (3.15), Heckman, Ichimura e Todd (1997) identificam três principais componentes específicas dos métodos de pareamento com base em características observáveis, a saber: (a) V_1 , da ausência de suporte comum entre as distribuições das covariáveis X dos dois grupos, dado por

$$V_1 = \int_{S_1 \setminus S_{10}} \mathbb{E}(Y(0)|X, T = 1)f(X|T = 1)dX - \int_{S_0 \setminus S_{10}} \mathbb{E}(Y(0)|X, T = 0)f(X|T = 0)dX ; \quad (3.16)$$

(b) V_2 , das diferenças nas distribuições das variáveis X entre os dois grupos, dado por

$$V_2 = \int_{S_{10}} \mathbb{E}(Y(0)|X, T = 0)(f(X|T = 1) - f(X|T = 0))dX ; \quad (3.17)$$

(c) e V_3 , das diferenças que restam após condicionar o modelo com as variáveis observáveis e realizando as comparações na região de suporte comum, isto é, o próprio viés de seleção, dado por

$$V_3 = \int_{S_{10}} (\mathbb{E}(Y(0)|X, T = 1) - \mathbb{E}(Y(0)|X, T = 0))f(X|T = 1)dX . \quad (3.18)$$

Então, $V = V_1 + V_2 + V_3$. Heckman, Ichimura e Todd (1997) concluem que o viés de seleção, V_3 , tem uma baixa importância relativa, se comparado aos componentes V_1 e V_2 . Ademais, o estudo evidencia que o método de pareamento acaba com as essas duas fontes de viés: (a) elimina V_1 ao realizar o pareamento apenas na região de suporte comum e (b) elimina V_2 pois a amostra do grupo de controle é construída justamente balanceando as características observáveis entre os dois grupos, conduzindo a uma situação em que as distribuições de $\epsilon(X)$ são iguais para tratados e não tratados. As evidências apresentadas no artigo indicam, portanto, que métodos baseados na seleção por observáveis eliminam a maior parte do viés, desde que o parâmetro de avaliação seja estimado na região de suporte comum (HECKMAN; ICHIMURA; TODD, 1997).

As propriedades assintóticas dos estimadores de efeito causal via escore de propensão são similares e devem fornecer os mesmos resultados (BOX-STEFFENSMEIER; BRADY; COLLIER, 2008). Contudo, no universo das amostras finitas, em que o escore de propensão deve ser estimado, o método escolhido tem influência sobre os resultados obtidos (CALIENDO; KOPEINIG, 2008). Em se tratando da escolha do método, não há uma

solução que se encaixe perfeitamente em todos os contextos e há um número de decisões a serem tomadas: quais variáveis observadas utilizar, qual o modelo utilizar para estimar o escore de propensão, qual critério de pareamento deve ser utilizado para selecionar a amostra, como estimar o efeito causal, e assim por diante.

Nessa linha, Heckman, Ichimura e Todd (1997) comparam a performance de diferentes estimadores de impacto baseados no método de pareamento. Quando se trata da implementação dos métodos, outras fonte de viés tornam-se relevantes: (a) seleção em não observáveis (ou seja, a existência de características que não podem ser observadas ou mensuradas pelos pesquisadores que influenciam a decisão de tratamento); (b) ausência de suporte comum; (c) falha na construção de um grupo de controle com características similares aos tratados. Estimadores de efeito causal baseados em pareamento podem incorrer em (a) ou (b). (b) pode ser eliminado quando o pareamento é realizado apenas na região de suporte comum. No entanto, o parâmetro de interesse deve ser redefinido de modo a restringir-se à região S_{10} . Quando a distribuição da amostra original é muito distinta da distribuição da nova amostra, pareada, o estimador de impacto restrito à região S_{10} é distinto do estimador para a amostra inteira. Então, pode-se esperar que haja viés decorrente da realização do pareamento ser possível apenas na região de suporte comum, o que é confirmado pelo estudo de Heckman, Ichimura e Todd (1997).

Ao testar diferentes grupos de covariáveis e diferentes configurações dos métodos de pareamento, Heckman, Ichimura e Todd (1997) mostraram que a compreensão sobre o contexto do programa avaliado pode influenciar consideravelmente as estimativas da probabilidade de participação e a omissão de variáveis importantes podem ser grandes fontes de viés. Por outro lado, a inclusão de variáveis não garante a eficácia dos estimadores. Incluir variáveis que predizem perfeitamente quem participa e quem não participa do programa impossibilita a realização do pareamento, isto é, a hipótese (3.2) não seria satisfeita para nenhum X (HECKMAN; ICHIMURA; TODD, 1997). Isso não é desejável, uma vez que a amostragem via pareamento tem o intuito de mimetizar uma situação em que a seleção para o programa é aleatória. Então, busca-se variáveis suficientes para obter independência condicional de $Y(0)$ e T , mas não excessivamente boas a ponto de prever T perfeitamente.

Os autores do estudo, no entanto, defendem que as maiores fontes de viés podem decorrer da origem dos dados. Por exemplo, quando questionários distintos eram aplicados a participantes e não participantes (no contexto do JTPA) e, portanto, a definição das características observadas era distinta para os dois grupos; ou quando eram comparados grupos de regiões geográficas (e, conseqüentemente, mercados de trabalho) distintas e contextos econômicos distintos, os métodos de pareamento incorriam em viés considerável. Em suas palavras, “*comparing the incomparable is a major source of evaluation bias*”

(HECKMAN; ICHIMURA; TODD, 1997, p. 607). Caso esses problemas não ocorram, Heckman, Ichimura e Todd (1997) defendem que as estimativas de impacto não experimentais podem ser próximas às suas contrapartidas experimentais, embora o viés não desapareça.

4 Estimadores de impacto

No que se refere à avaliação de impacto, a principal pergunta a ser respondida é: qual é o impacto (ou efeito causal) de um programa sobre um resultado de interesse? (KHANDKER; KOOLWAL; SAMAD, 2009) Até agora, foi discutido como construir um grupo similar ao grupo de tratamento, que pudesse se aproximar da situação contrafactual, isto é, do resultado que teria sido observado com os indivíduos tratados caso não o tivessem sido.

Naturalmente, pode-se indagar se o próprio grupo de tratados, mas em sua condição pré-tratamento, serve de contrafactual, isto é, utilizar as mudanças ocorridas ao longo do tempo com os tratados como medida de impacto. Todavia, outros fatores podem exercer influência sobre a variável de resultado ao longo do tempo — o contexto macroeconômico, características microeconômicas locais, mudanças na legislação, e assim por diante — de maneira que a tendência temporal da variável de resultado pode ter outras causas que não sejam o tratamento, por vezes difíceis de identificar. Portanto, a simples observação das mudanças pré e pós tratamento nas variáveis de interesse do grupo de controle tampouco fornece uma medida confiável de impacto. Em outras palavras, o grupo de tratamento pré programa não é um contrafactual do mesmo grupo pós programa.

Os métodos de pareamento lidam com a ausência de independência entre resultados potenciais e tratamento (2.13) primordialmente utilizando características observáveis dos indivíduos como indicadores da seleção ao tratamento, fazendo valer sua versão mais fraca, a hipótese de ignorabilidade (3.1). No entanto, deseja-se também levar em conta características que não podem ser observadas pelo pesquisador, mas que possam interferir na decisão de participar do tratamento — exemplos comuns são personalidade ou outras características inatas dos indivíduos. Nesse caso, diz-se que há seleção por (características) não observáveis e não há uma classe bem determinada de métodos a serem utilizados, a não ser em situações específicas.

Alguns procedimentos que levam em conta a seleção por não observáveis são variáveis instrumentais, modelos de regressão descontínua, modelos de efeitos fixos ou aleatórios para dados em painel, além de diferenças em diferenças (IMBENS; WOOLDRIDGE, 2009). Neste capítulo serão apresentados o método de diferenças em diferenças (suas principais hipóteses de identificação do efeito do tratamento, o modelo de regressão linear correspondente e o parâmetro de interesse a ser estimado) e o modelo de painel de efeitos fixos. Embora o método utilizado ao final do trabalho para avaliar o impacto da tarifa

zero tenha sido o modelo de painel de efeitos fixos (uma vez que se dispunha de dados em painel com diversos períodos de tempo), o método de diferença em diferenças confere a intuição básica do procedimento realizado. Então, optou-se por apresentá-lo em conjunto, devido à sua simplicidade.

4.1 Diferenças em diferenças

Há registros da utilização de diferenças em diferenças tão antigos quanto 1854, quando o epidemiologista John Snow reuniu dados acerca de uma epidemia de cólera em Londres na época. O pesquisador almejava descobrir a via de transmissão da doença — se pela água ou pelo ar. Snow manipulou o fornecimento de água de algumas regiões da cidade — uma parte recebia água do Tâmis no centro da cidade e outra recebia água limpa de outra parte do rio. Ao expor vizinhanças próximas (e, portanto, com qualidade similar do ar) a diferentes fontes de água pôde estabelecer uma conexão entre a água poluída e a taxa de mortes por cólera (LECHNER *et al.*, 2011). Snow dispunha dos dados acerca da taxa de mortes por cólera antes da mudança no fornecimento da água, o que permitiu distinguir a diferença na incidência de cólera entre bairros decorrente de outros fatores, já presentes antes da intervenção.

Em economia, uma das aplicações mais antigas do método foi feita por Obenauer e Nienburg (1915) em um relatório do *Bureau of Labour Statistics* que visava avaliar o impacto do estabelecimento de um salário mínimo para as mulheres em Oregon, Estados Unidos (KENNAN, 1995). Mais tarde, foram os trabalhos seminais de Ashenfelter (1978) e Orley e David (1985), os quais avaliaram o impacto de programas de treinamento nos mercados de trabalho locais, que tornaram o método difundido nos trabalhos empíricos em economia (IMBENS; WOOLDRIDGE, 2009).

Na formulação mais simples do método de diferença em diferenças, considera-se quatro tipos de objetos, saber: o grupo de indivíduos tratados pós tratamento, tratados pré tratamento, o grupo de não tratados pós tratamento e não tratados pré tratamento. Isto é, são dois grupos (tratados e controles) e dois períodos de tempo (antes e depois do tratamento). Ele leva esse nome pois considera duas diferenças: a primeira diz respeito à comparação das variáveis antes e depois do programa em ambos os grupos (controle e tratados). Em seguida, é computada a diferença entre o resultado obtido nos dois grupos. Essa dupla subtração remove o viés de comparar ambos os grupos apenas no segundo período, que poderia ser resultado de diferenças permanentes entre os dois grupos, assim como o viés da comparação ao longo do tempo do grupo tratado resultante de tendências temporais não relacionadas ao tratamento sob avaliação (IMBENS; WOOLDRIDGE, 2009).

Esse método impõe uma restrição a mais nos dados, considerando que são necessárias observações pré-tratamento das variáveis de impacto que estão sendo avaliadas. Já uma vantagem é que ele pode ser implementado a partir de dados muitas vezes disponíveis ao público e sua atratividade encontra-se no fato de não serem necessárias ferramentas computacionais sofisticadas para obter a estimativa de impacto, ao menos quando não são utilizados modelos de inferência muito complexos (LECHNER *et al.*, 2011).

4.1.1 Identificação do efeito causal

Em linha com a notação utilizada até agora, o grupo ao qual o indivíduo i pertence é representado pela variável binária T , com $T_i = 1$ se i for tratado e $T_i = 0$ caso contrário. Primeiramente, suponha que as variáveis estão disponíveis em dois períodos de tempo indicados por $t \in \{0, 1\}$. O período zero indica uma observação pré-tratamento, enquanto o período um indica uma observação pós tratamento. Agora, as variáveis de resultado potencial serão representadas com um índice adicional t para indicar o período. Assim, $Y_{it}(0)$ é o resultado potencial de não tratamento do indivíduo i no tempo t e $Y_{it}(1)$ é o resultado potencial de receber tratamento do mesmo indivíduo no período de tempo t . O resultado observado é denotado por Y_{it} . Deseja-se estimar o efeito médio sobre os tratados, agora também associado a cada período t . Então, ainda tomando a esperança com relação aos indivíduos,

$$EMTT_t = \mathbb{E}(Y_t(1) - Y_t(0)|T = 1). \quad (2.12 \text{ revisitada})$$

Algumas hipóteses são necessárias para identificação do parâmetro de interesse. Primeiramente, mantém-se a hipótese de não interferência entre os membros da população, a SUTVA. Segue dessa hipótese que um, e apenas um, dos resultados potenciais é observável para cada membro da população (LECHNER *et al.*, 2011). Assim, de forma coerente com o que foi apresentado nos capítulos anteriores, o resultado observado está plenamente representado pela equação

$$Y_{it} = T_{it} \cdot Y_{it}(1) + (1 - T_{it}) \cdot Y_{it}(0). \quad (2.2 \text{ revisitada})$$

Note que novamente foi acrescentado um subscrito t para indicar o período de tempo. Adicionalmente, como em $t = 0$ a intervenção ainda não ocorreu, determina-se que nesse período o tratamento não teve efeito na população, isto é,

$$EMTT_0 = 0 \quad (4.1)$$

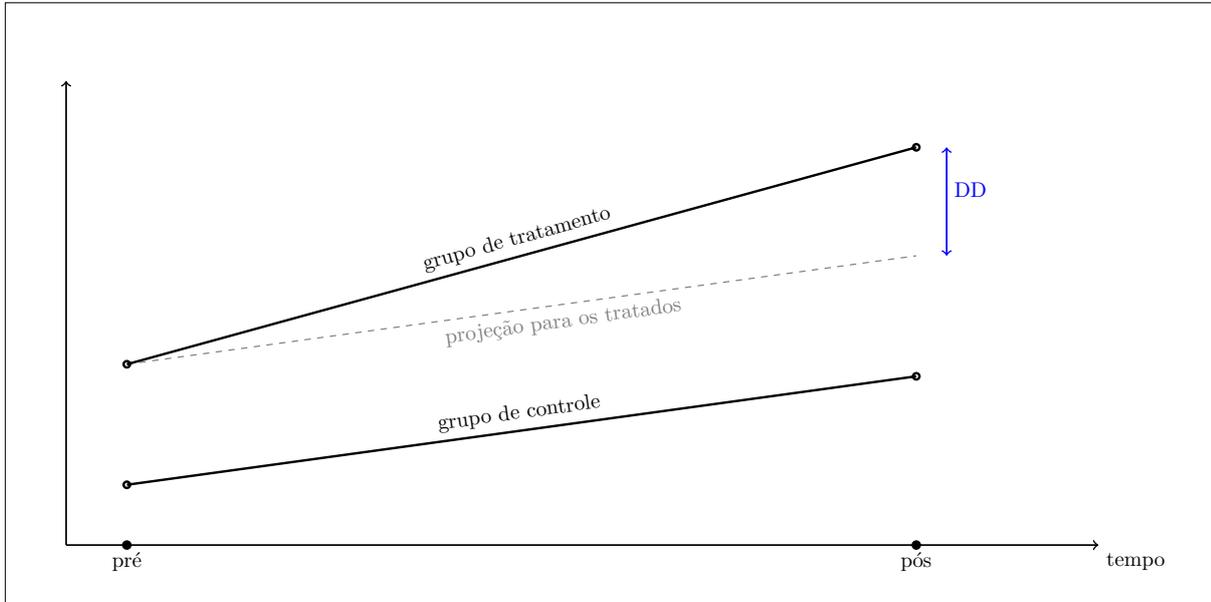


Figura 2 – Representação da hipótese de tendência comum e do estimador de diferenças em diferenças.

A importância de (4.1) é que ela exclui mudanças de comportamento entre os indivíduos tratados que possam influenciar o resultado pré-tratamento pela antecipação de tratamento futuro.

Outra hipótese central do método é a de que há uma *tendência comum*, de que a diferença do resultado potencial de não tratamento ao longo do tempo ($Y_1(0) - Y_0(0)$) é a mesma para o grupo de tratados e o grupo de controle. Em outras palavras, supõe-se que a tendência temporal do grupo de tratamento caso não tivesse recebido tratamento teria sido a mesma tendência observada no grupo de controle, que coincide com a tendência temporal do resultado potencial de não tratamento para esse grupo. Formalmente,

$$\begin{aligned}
 & \mathbb{E}(Y_1(0)|T = 1) - \mathbb{E}(Y_0(0)|T = 1) \\
 &= \mathbb{E}(Y_1(0)|T = 0) - \mathbb{E}(Y_0(0)|T = 0) \\
 &= \mathbb{E}(Y_1(0)) - \mathbb{E}(Y_0(0)).
 \end{aligned} \tag{4.2}$$

Quando são incluídas covariáveis, Lechner (2008) aponta devem incluir aquelas que poderiam levar a diferenças na tendência temporal dos dois grupos. A tendência comum é a hipótese chave do método de diferenças em diferenças: caso se espere que os resultados potenciais de não tratamento dos dois grupos ($T = 1$ e $T = 0$) sigam a mesma trajetória ao longo do tempo, um eventual desvio dessa trajetória observado no grupo de tratamento pode ser atribuído justamente à incidência do tratamento. A figura 2 ilustra a hipótese de tendência comum, em que a linha tracejada representa a tendência temporal do grupo de controle, projetada a partir da condição inicial do grupo de tratamento para inferir

qual teria sido o resultado na situação contrafactual.

A última hipótese é denominada *viés estável*¹. Imagine que se suponha (falsamente) que não há características não observáveis que afetem a medida de efeito causal, isto é, suponha falsamente que condicional a X , os resultados potenciais são independentes do tratamento. Adotando a hipótese (4.1), se o efeito causal estimado no período $t = 0$ for diferente de zero, tem-se viés do estimador de impacto e a hipótese de ignorabilidade (3.1) não seria verdadeira. Mas se esse viés é constante ao longo do tempo, a estimativa de impacto pode ser corrigida para isolar o efeito do tratamento — essa é a hipótese de viés estável. Defina as medidas de viés para cada período t ($t \in \{0, 1\}$) como $V_t = V_0(x) = \mathbb{E}(Y_t(0)|T = 1) - \mathbb{E}(Y_t(0)|T = 0)$, ou seja:

$$V_0 = \mathbb{E}(Y_0(0)|T = 1) - \mathbb{E}(Y_0(0)|T = 0) \text{ e} \quad (4.3)$$

$$V_1 = \mathbb{E}(Y_1(0)|T = 1) - \mathbb{E}(Y_1(0)|T = 0). \quad (4.4)$$

A hipótese de *viés estável* pode ser formalizado por:

$$V_0 = V_1. \quad (4.5)$$

Então, $V_0 = V_1 = V_s$, isto é, o viés de seleção é constante ao longo do tempo. Agora, deve-se determinar o efeito médio sobre os tratados no período pós tratamento, $EMTT_1$, com base em medidas que podem ser estimadas a partir dos dados observados, isto é, a partir de Y_{i1} e Y_{i0} . De sua definição,

$$\begin{aligned} EMTT_1 &:= \mathbb{E}(Y_1(1) - Y_1(0)|T = 1) \\ &= \mathbb{E}(Y_1|T = 1) - \mathbb{E}(Y_1(0)|T = 1), \end{aligned} \quad (4.6)$$

em que o primeiro termo, $\mathbb{E}(Y_1|T = 1)$ pode ser estimado pela média da variável de resultado dos tratados no período pós tratamento. Quanto ao segundo termo, referente ao resultado potencial de não tratamento dos tratados (contrafactual) no período 1, é possível utilizar a hipótese de tendência comum (4.2), rearranjar seus termos, e obter:

$$\begin{aligned} \mathbb{E}(Y_1(0)|T = 1) &= \mathbb{E}(Y_1(0)|T = 0) - \mathbb{E}(Y_0(0)|T = 0) \\ &\quad + \mathbb{E}(Y_0(0)|T = 1) \\ &= \mathbb{E}(Y_1|T = 0) - \mathbb{E}(Y_0|T = 0) \\ &\quad + \mathbb{E}(Y_0(0)|T = 1), \end{aligned} \quad (4.7)$$

¹ *Bias stability* ou *constant bias* (LECHNER *et al.*, 2011).

em que os dois primeiros termos podem ser estimados, uma vez que consistem simplesmente nas médias das observações pós e pré tratamento do grupo de controle. Por fim, basta determinar $\mathbb{E}(Y_0(0)|T = 1)$. Decorre de (4.1), isto é, do fato de do efeito médio do tratamento antes do tratamento ser nulo que

$$\begin{aligned}\mathbb{E}(Y_0(0)|T = 1) &= \mathbb{E}(Y_0(1)|T = 1) \\ &= \mathbb{E}(Y_0|T = 1).\end{aligned}\tag{4.8}$$

E $\mathbb{E}(Y_0|T = 1)$ pode ser estimado pela média das observações dos tratados pré tratamento. Substituindo (4.8) em (4.7), obtém-se que a esperança do resultado potencial de não tratamento no período pós tratamento ($t = 1$) no grupo de tratados é a diferença nos resultados do grupo de controle antes e depois da intervenção mais o resultado do grupo de tratados antes do tratamento (sua condição inicial). Formalmente,

$$\begin{aligned}\mathbb{E}(Y_1(0)|T = 1) &= \mathbb{E}(Y_1|T = 0) - \mathbb{E}(Y_0|T = 0) \\ &\quad + \mathbb{E}(Y_0|T = 1)\end{aligned}\tag{4.9}$$

Agora, substituindo (4.9) na relação (4.6), obtém-se

$$\begin{aligned}EMTT_1 &= \mathbb{E}(Y_1|T = 1) - \mathbb{E}(Y_1(0)|T = 1) \\ &= \mathbb{E}(Y_1|T = 1) - \mathbb{E}(Y_1|T = 0) \\ &\quad + \mathbb{E}(Y_0|T = 0) - \mathbb{E}(Y_0|T = 1) \\ &= \mathbb{E}(Y_1|T = 1) - \mathbb{E}(Y_0|T = 1) \\ &\quad - (\mathbb{E}(Y_1|T = 0) - \mathbb{E}(Y_0|T = 0)).\end{aligned}\tag{4.10}$$

A última igualdade em (4.10) é apenas um rearranjo dos termos.

4.1.2 Regressão linear

4.1.2.1 Caso com dois períodos de tempo

O modelo de diferenças em diferenças pode também ser formulado em um formato de regressão linear. Supõe-se que o resultado potencial de não tratamento do indivíduo i tem uma estrutura aditiva que pode ser especificada como

$$Y_i(0) = \alpha + \beta \cdot t + \gamma \cdot T_i + u_i\tag{4.11}$$

Isso significa que, na ausência de tratamento, o nível da variável de resultado é determinada por um efeito invariante no tempo atribuído ao grupo (tratamento ou controle,

Tabela 3 – Interpretação dos coeficientes do formato de regressão linear do modelo de diferenças em diferenças.

	Pré	Pós	Diferença (pós - pré)
Grupo de controle	α	$\alpha + \beta$	β
Grupo tratado	$\alpha + \gamma$	$\alpha + \beta + \gamma + DD$	$\beta + DD$
Diferença (tratado - controle)	γ	$\gamma + DD$	DD

Fonte: Adaptado de (PEIXOTO *et al.*, 2012).

denotado pela variável T) e um efeito do período de tempo que é comum a ambos os grupos (ANGRIST; PISCHKE, 2008). u_i representa as características não observáveis dos indivíduos, que supõe-se constante ao longo do tempo e independente da variável de tratamento, isto é, $u_i \perp (T_i, t)$. Supondo que o efeito do tratamento DD é constante, a equação da variável de resultado potencial do tratamento é: $Y_i(1) = Y_i(0) + DD$. O efeito causal pode ser estimado com uma regressão linear:

$$Y_i = \alpha + \beta \cdot t + \gamma \cdot T_i + DD \cdot P_i + \varepsilon_i, \quad (4.12)$$

em que a variável $P_i = t \cdot T_i$ é uma *dummy* de interação entre as variáveis t e T_i . Desse modo, $P_i = 1$ para observações do grupo de tratados no período pós tratamento e $P_i = 0$ caso contrário e, evidentemente, seu coeficiente DD é o parâmetro de interesse no modelo. Dessa maneira, o estimador de diferença em diferenças é tal que:

$$\widehat{DD} = (\bar{Y}_{11} - \bar{Y}_{10}) - (\bar{Y}_{01} - \bar{Y}_{00}). \quad (4.13)$$

Aqui, com uma amostra de N indivíduos, se N_{gw} é o número de indivíduos da amostra que receberam tratamento g no período de tempo w , tem-se que $\bar{Y}_{gw} = \sum_{i \in I_g} \frac{Y_i}{N_{gw}}$ é o resultado médio desse subgrupo de indivíduos (IMBENS; WOOLDRIDGE, 2009). Essas quantidades são representadas na tabela 3.

4.1.2.2 Diversos períodos de tempo

O modelo de regressão apresentado na subseção anterior pode ser estendido para o caso de haver disponibilidade de dados para mais períodos de tempo — sejam eles em um formato de painel ou múltiplos cortes transversais. Considere que há K períodos de tempo e que D_k é uma variável binária referente ao período k . Então a equação (4.11) pode ser reescrita como

$$Y_i(0) = \alpha + \sum_{k=1}^K \lambda_k \cdot D_k + \gamma \cdot T_i + \varepsilon_i, \quad (4.14)$$

com parâmetros separados para cada período de tempo. Combinando o modelo com a equação $Y_i(1) = Y_i(0) + DD$, obtém-se

$$Y_i(1) = \alpha + \sum_{k=1}^K \lambda_k \cdot D_k + \gamma \cdot T_i + DD \cdot P_i + \varepsilon_i. \quad (4.15)$$

De maneira análoga ao caso com dois períodos de tempo, supõe-se um efeito aditivo linear e, portanto, (4.15) pode ser estimada utilizando um modelo de Mínimos Quadrados Ordinários (MQO) (IMBENS; WOOLDRIDGE, 2009). No entanto, a especificação em (4.15) não leva em consideração o caso em que diferentes indivíduos recebem tratamento em períodos de tempo distintos. Os dados em painel, por outro lado, acompanham os mesmos indivíduos ao longo do tempo. É possível levar isso em consideração utilizando um modelo de painel de efeitos fixos, o qual será apresentado na próxima seção.

4.2 Modelo de efeitos fixos

Quando há disponibilidade de dados em painel, tanto de períodos anteriores quanto posteriores ao tratamento, estimador de diferenças em diferenças pode ser obtido através de um modelo de painel com efeitos fixos de tempo e individuais. Na próxima seção, o modelo de efeitos fixos será apresentado em um contexto geral e posteriormente adaptado às necessidades deste trabalho.

4.2.1 Componentes de erro

Considere o modelo de regressão linear simples dado por

$$y_{it} = X_{it}\beta + u_{it}, \quad (4.16)$$

em que y_{it} é a variável dependente, X_{it} é a matriz de variáveis independentes, β é o vetor de parâmetros a serem estimados e u_{it} é o termo de erro. Os índices it dizem respeito à observação para o indivíduo i no período t . Considere que haja n indivíduos e k períodos de tempo e que os dados se encontrem em uma estrutura de pilha com $m = n \cdot k$ linhas

Em se tratando de dados em painel, é possível que as hipóteses padrão do MQO não sejam válidas: pode haver choques que afetam todos os indivíduos em um determinado período de tempo, de modo que há correlação entre u_{it} e u_{is} , para $s \neq t$ ou choques que

afetam um determinado indivíduos em todos os períodos de tempo, o que se reflete em correlação entre u_{it} e u_{jt} para $i \neq j$. Isso viola a hipótese do MQO de que os termos de erro não são correlacionados², tornando o estimador ineficiente. Além disso, a presença de autocorrelação leva a estimativas de erro padrão subestimadas e estimativas de estatísticas t infladas (CAMERON; TRIVEDI, 2005).

Uma maneira de lidar com esse problema envolve dividir o termo de erro u_{it} em diferentes componentes,

$$u_{it} = \lambda_t + v_i + \varepsilon_{it}, \quad (4.17)$$

em que λ_t afeta todos os indivíduos no período t e v_i afeta o indivíduo i em todos os períodos de tempo. Supõe-se que λ_t são independentes em períodos de tempo diferentes, v_i são independentes para indivíduos diferentes e ε_{it} são independentes entre si e são específicos à observação it ³.

É possível lidar com o modelo de componentes de erro (4.17) de duas maneiras, com modelos de efeitos aleatórios ou fixos. Caso vistos como efeitos aleatórios, é necessário identificar a matriz de covariância dos termos u_{it} como função dos termos λ_t , v_i e ε_{it} e utilizar o modelo de mínimos quadrados generalizados. Já com efeitos fixos, λ_t e v_i são vistos como parâmetros a serem estimados utilizando o MQO acrescido de variáveis *dummy* (DAVIDSON; MACKINNON, 2004).

A equação do modelo pode, portando, ser reescrita como

$$y_{it} = X_{it}\beta + \lambda_t + v_i + \varepsilon_{it}. \quad (4.18)$$

Denotando por ε o vetor cujos elementos são ε_{it} , adota-se a hipótese de que $\mathbb{E}(\varepsilon\varepsilon^T) = \sigma_\varepsilon^2 I$, em que I é a matriz identidade.

Davidson e MacKinnon (2004) reiteram que, apesar de denominados efeitos fixos, λ_t e v_i podem ser aleatórios, desde que independentes de ε_{it} . Pode, no entanto, haver correlação com as variáveis X_{it} . Além disso, para evitar colinearidade entre os regressores, X_{it} não deve incluir uma constante ou mesmo uma variável que tem o mesmo valor para todos os indivíduos. Os parâmetros em (4.18) podem ser estimados via MQO.

² $\mathbb{E}(uu^T) = \sigma^2 I$, em que I é a matriz identidade (DAVIDSON; MACKINNON, 2004).

³ As aplicações mais comuns em livros-texto ((DAVIDSON; MACKINNON, 2004), (CAMERON; TRIVEDI, 2005)) são aquelas em que há apenas o componente relativo aos indivíduos, o modelo de efeitos fixos *one-way*. A versão que inclui efeitos fixos de tempo é comumente denominada de modelo *two-way* (CAMERON; TRIVEDI, 2005).

4.2.2 Estimação de impacto

No início deste capítulo, levantou-se a questão da seleção para o tratamento por características não observáveis. Para lidar com esse problema, propõe-se utilizar o modelo de painel de efeitos fixos. Deve-se ressaltar da discussão apresentada que a variável de tratamento, que identifica os grupos de controle e tratamento, não é mantida, uma vez que é invariante no tempo. Retomando a notação de resultados potenciais das seções anteriores, mantém-se a variável P_{it} , variável binária que identifica períodos pós tratamento de indivíduos tratados.

No contexto de efeitos fixos, supõe-se que a seleção para o tratamento é aleatória se condicional à variável de tratamento, eventuais características observáveis X_{it} , e uma variável não observada A_i . Formalmente,

$$\mathbb{E}(Y_{it}(0)|A_i, P_{it}, X_{it}, t) = \mathbb{E}(Y_{it}(0)|A_i, X_{it}, t). \quad (4.19)$$

A principal hipótese do modelo de efeitos fixos, assim como com diferenças em diferenças, é que a característica não observada A_i é invariante ao longo do tempo. Assim, de maneira análoga à equação (4.14), tem-se o modelo para o resultado potencial de não tratamento dado por $\mathbb{E}(Y_{it}(0)|A_i, X_{it}, t) = \alpha + \lambda_t + A_i \cdot \gamma + X_{it}\beta$.

Novamente, supõe-se que o efeito do tratamento é constante e aditivo, de modo que para o resultado potencial de tratamento é

$$\mathbb{E}(Y_{it}(1)|A_i, X_{it}, t) = \mathbb{E}(Y_{it}(0)|A_i, X_{it}, t) + DD. \quad (4.20)$$

Por fim, o modelo de regressão considerando os componentes de erro é

$$Y_{it} = v_i + \lambda_t + DD \cdot P_{it} + X_{it}\beta + \varepsilon_{it}, \quad (4.21)$$

em que $v_i = \alpha + A_i \cdot \gamma$. Quando há um grande número de observações, há muitos parâmetros a serem estimados. Na prática, Angrist e Pischke (2008) apontam que tratar os efeitos fixos como parâmetros é o mesmo que estimar os desvios da média de cada observação.

5 Transporte e desenvolvimento

5.1 Pobreza e exclusão social no Brasil

O termo pobreza geralmente refere-se a uma medida de caráter pecuniário, que evoca a insuficiência de renda. Uma medida de pobreza pode ser absoluta ou relativa. A pobreza absoluta é avaliada quando é estabelecido um nível mínimo de subsistência caracterizado em termos de um nível de renda, uma “linha de pobreza” e classifica-se os indivíduos como pobres quando sua renda está abaixo de tal patamar. Desse modo, uma medida de pobreza absoluta é o número ou a proporção de habitantes de um país ou região que encontram-se abaixo da linha de pobreza. Já a pobreza relativa é medida através da comparação entre o nível de renda dos indivíduos que se encontram nos níveis inferiores da distribuição de renda com aqueles que se encontram em camadas superiores, usualmente as médias (GANNON; LIU, 1997).

Apesar do cunho unidimensional dessa definição, não há dúvidas de que o fenômeno da pobreza se manifesta em todas as dimensões da vida de um indivíduo, como subnutrição, analfabetismo, falta de acesso a oportunidades de trabalho e condições precárias de moradia e de vida. Esse caráter multifacetado da condição de pobreza pode ser melhor compreendido a partir de Sen (2000). O autor vê o processo de desenvolvimento como o de expansão das diferentes liberdades de que usufruem os indivíduos, isto é, a liberdade não é mera consequência do desenvolvimento, mas seu fim primordial. Esse é o denominado papel constitutivo da liberdade no desenvolvimento e relaciona-se à importância da liberdade substantiva para a vida humana (SEN, 2000).

Por outro lado, a expansão da liberdade é o principal meio de desenvolvimento, isto é, tem papel instrumental no processo. As liberdades substantivas são capacidades elementares, como evitar situações de privação material, assim como a liberdade de ser capaz de ler, fazer cálculos aritméticos, ter liberdade de expressão e de participação da vida política. Sen aponta que medidas que promovem a expansão dessas liberdades também têm impacto sobre a renda e o produto, isto é, promovem o desenvolvimento econômico no sentido tradicional do termo (SEN, 2000).

A condição de pobreza é caracterizada não apenas por baixa renda e incapacidade de adquirir bens, mas pela falta de acesso a serviços essenciais (água, luz, esgoto, educação, saúde), discriminação social, a negação dos direitos sociais e ao acesso aos mercados de bens e de trabalho (SEN, 2000). Além disso, é marcada pela exclusão das decisões políticas

e ausência de representatividade no processo de tomada de decisão e impossibilidade de reivindicação de direitos sociais (GOMIDE, 2003). Em suma, a condição de pobreza se caracteriza pela ausência de liberdades substantivas essenciais à vida e, no sentido proposto por Sen (2000), pode-se dizer que a superação da pobreza está imbricada no processo de desenvolvimento como expansão das liberdades.

A exclusão social decorrente da pobreza também se manifesta como exclusão espacial, sobretudo (mas não somente) no meio urbano (GOMIDE, 2003). Há uma verdadeira cisão da cidade: de um lado, a parcela rica, repleta de oportunidades e dotada de infraestrutura e atenção, ainda que muitas vezes insuficiente, do poder público. De outro, a cidade informal, marcada pelas favelas, cortiços e ocupações ilegais da terra, pobre, distante das oportunidades rentáveis e carente de serviços essenciais. Conforme aponta Grieco (2013), quanto à condição de (falta de) acesso em toda a América Latina:

Poverty has a distinctive geography; and that geography has not been well served by either the absence of appropriate localized provision of pro-poor services or by the absence of pro-poor compensating mobility systems. Poverty has largely been left unaddressed in respect of accessibility and mobility (GRIECO, 2013, p. 2).

Assim, a condição de pobreza é agravada pela distância geográfica daqueles locais onde as pessoas poderiam ter suas necessidades supridas e também pela impossibilidade ou dificuldade de acesso a esses estabelecimentos. Em outras palavras, a distribuição das instalações onde são oferecidos serviços urbanos — públicos ou não — tem implicações em termos de equidade. Nesse contexto, baixos níveis de acesso aos serviços de educação, saúde, emprego, fornecimento de água, esgoto e luz, atividades culturais e de lazer e os próprios serviços de transporte podem ser vistos como maneiras específicas através das quais essas iniquidades podem se manifestar (GRIECO, 2015).

Tendo isso em vista, a garantia de acesso — e por consequência da mobilidade — se coloca como liberdade substantiva essencial no tocante às discussões de intervenções de combate à pobreza por parte do poder público. Entende-se que o acesso pode ocorrer de duas maneiras. A primeira é através do deslocamento até o local onde os serviços e oportunidades são oferecidas. Por vezes, o acesso torna-se possível quando esses serviços “se deslocam” até os usuários, isto é, em vez de promover a jornada, pode-se formular maneiras de oferecer esses serviços nos locais onde moram os necessitados.

Em uma pesquisa realizada no ano de 2001, Barros, Henriques e Mendonça investigaram o caráter da pobreza no Brasil. Seu diagnóstico foi que a forte exclusão social que marca toda a história do país não é decorrente de uma escassez de recursos, mas de um padrão de crescimento e distribuição de renda extremamente desigual. Pelo contrário, o país

é abundante em recursos. Na opinião dos autores, a pobreza no Brasil — no aspecto da insuficiência de renda — tem a desigualdade como principal causa e, portanto, é o principal nó que as autoridades políticas devem buscar desatar caso se queira superar a condição de pobreza do país.

Desde então, algum tempo (e diversas políticas de assistência social) se passou, mas o diagnóstico continua semelhante. Apesar de o Brasil ter conquistado enorme redução da pobreza entre 2004 e 2014, ainda somos o país mais desigual do mundo (SKOUFIAS; NAKAMURA; GUKOVAS, 2017). Barros, Henriques e Mendonça parecem ter identificado um aspecto extremamente relevante, a saber, que a desigualdade na sociedade brasileira é, além de profunda, estável:

Desigualdade que surpreende tanto por sua intensidade como, sobretudo, por sua estabilidade. Desigualdade extrema que se mantém inerte, resistindo às mudanças estruturais e conjunturais das últimas décadas. Desigualdade que atravessou impassível o regime militar, governos democraticamente eleitos e incontáveis laboratórios de política econômica, além de diversas crises políticas, econômicas e internacionais (BARROS; HENRIQUES; MENDONÇA, 2001, p. 23).

Apesar do diagnóstico pessimista, verifica-se um acentuado declínio da pobreza no decênio de 2004-2014. Além disso, o coeficiente de Gini, principal indicador de desigualdade, também caiu no mesmo período (SKOUFIAS; NAKAMURA; GUKOVAS, 2017). O avanço no nível de emprego formal consiste em uma melhora qualitativa no que se refere à exclusão social. Por exemplo, o trabalhador informal pode ter renda acima da linha da pobreza, mas não usufrui de direitos como descanso remunerado ou seguridade social e, nesse sentido, sua cidadania lhe é negada (GOMIDE, 2003).

Além das transferências de renda, boa parte do sucesso na redução da pobreza deveu-se a uma melhoria nas condições de trabalho, impulsionada por condições externas favoráveis. À semelhança de outros países de renda média, nota-se que o desempenho do mercado de trabalho foi essencial na promoção do crescimento e melhoria nas condições dos pobres (SKOUFIAS; NAKAMURA; GUKOVAS, 2017).

No caso brasileiro, a manutenção de taxas de câmbio valorizadas permitiu o crescimento de setores de serviços não transacionáveis, proporcionando oportunidades de emprego em ramos que não exigem alta qualificação. O correspondente aumento de renda aumentou a demanda por bens e serviços em toda a economia, alimentando — ao menos enquanto a condição externa foi favorável — uma fase de crescimento e redução do nível de pobreza. De acordo com os autores, 58% da redução da extrema pobreza no país ainda deveu-se às transferências de renda e outras fontes que não o trabalho (contra 37% para renda

relativa ao trabalho). Entre os pobres, os números foram de 30% e 58% respectivamente¹ (SKOUFIAS; NAKAMURA; GUKOVAS, 2017). As dificuldades econômicas e políticas pelas quais passa o país atualmente podem ameaçar o sucesso das políticas de combate à pobreza, como já apontam os atuais índices de desemprego e a queda do salário real.

Diante da recessão da economia brasileira, com contração de 3,8% do PIB em 2015 e 3,6% em 2016, o mercado de trabalho sofre com as consequências adversas. A taxa de desemprego subiu de 4,6% em dezembro de 2014 para 11,8% em outubro de 2016 e a previsão de Skoufias, Nakamura e Gukovas (2017) é de que haja uma reversão de parte da queda da pobreza. Ressalta-se que políticas de proteção aos pobres, como o Bolsa Família, cumprem um papel adicional na atual conjuntura. Além de promover uma melhoria no perfil de distribuição de renda do país, elas consistem em uma rede de proteção aos mais pobres e àqueles que Skoufias, Nakamura e Gukovas chamam de novos pobres — os filhos da crise econômica pela qual o país passa. A previsão dos autores é de que a quantidade de brasileiros pobres aumente, em proporção maior nas áreas urbanas do que nas rurais e acompanhada de um aumento na desigualdade de renda.

Em 2015, a proporção de pessoas extremamente pobres na população brasileira era de 3,4% e de “moderadamente pobres”² era de 8,7% (SKOUFIAS; NAKAMURA; GUKOVAS, 2017). Todavia, cerca de 23% da população economicamente ativa recebe menos de um salário mínimo.

O quadro descrito acima coloca a imperativa necessidade de discutir maneiras de combater a pobreza e melhorar as condições de vida dos mais pobres, eliminando situações de privação não só de renda, mas da participação da vida em sociedade. Na linguagem de Sen (2000), deve-se buscar o desenvolvimento através da expansão das liberdades dos indivíduos.

Ao descrever o êxito da redução da mortalidade em alguns países pobres, Sen (2000) classifica duas principais vias, os processos mediados pelo crescimento e aqueles conduzidos pelo custeio público. A primeira ocorre por meio do crescimento econômico rápido e seu êxito depende do processo de crescimento ter uma base ampla e ser economicamente abrangente. É crucial a utilização da prosperidade econômica para expansão dos serviços sociais relevantes. Por outro lado, o processo conduzido pelo custeio público opera por meio de um programa manutenção social dos serviços de saúde, educação, e outras disposições sociais relevantes. Quando esse é o caminho adotado, há melhoria nas condições de vida sem que necessariamente haja grande crescimento econômico (SEN, 2000).

¹ A linha de extrema pobreza foi estabelecida em R\$70 e a linha de pobreza em R\$140 em preços de junho de 2011.

² Abaixo de R\$140!

A distinção acima entre os processos mediados pelo crescimento e os conduzidos pelo custeio público pode ser generalizada para um contexto geral de combate à pobreza e eliminação das situações de privação. Com respeito à elaboração de políticas, o problema pode ser formulado sob duas perspectivas complementares, isto é, intervenções por parte do poder público podem promover o desenvolvimento e a expansão das liberdades em duas frentes: direta e indireta. A abordagem indireta parte do princípio de que o crescimento econômico gera oportunidades rentáveis aos indivíduos pobres. A abordagem direta preocupa-se com o fornecimento de serviços e assistência aos pobres com relação às suas necessidades básicas, como alimentação e educação (GANNON; LIU, 1997).

Barros, Henriques e Mendonça (2001) apontam para a importância de levar em consideração que:

As quedas observadas na magnitude da pobreza em todos os anos posteriores a 1977 resultam, primordialmente, do crescimento econômico. O papel da redistribuição de renda é bastante limitado durante todo o período, com exceção do final da década de 80, em particular no ajuste posterior a 1989, quando vigorava o maior grau de desigualdade das duas décadas (BARROS; HENRIQUES; MENDONÇA, 2001).

Ou seja, na opinião dos autores, verificou-se uma ênfase nas abordagens indiretas de combate à pobreza, sobretudo quando os esforços das autoridades macroeconômicas estavam voltados à estabilização da inflação e à retomada do crescimento econômico pós plano Real. Isso poderia explicar a persistência na desigualdade da distribuição de renda (pobreza relativa) apesar da redução no nível absoluto de pobreza. O combate à pobreza pela via do crescimento econômico dá importância ao aumento da eficiência da alocação de recursos e ao melhor funcionamento dos mercados. Entende-se que o crescimento do produto tende a beneficiar a população como um todo, incluindo os pobres, sob forma de emprego e outras oportunidades econômicas (GANNON; LIU, 1997).

A abordagem direta parte do princípio de que as pessoas em uma situação de pobreza extrema podem não ser capazes de usufruir de novas oportunidades geradas pelo crescimento econômico e o aumento da renda pode não ser suficiente para promoção da liberdade, no sentido proposto por Sen (2000). Então, busca-se elaborar intervenções diretas nas áreas de saúde, educação, nutrição e acesso ao emprego. Essas intervenções podem ter um caráter redistributivo ou apenas buscar promover maior eficiência na prestação desses serviços (GANNON; LIU, 1997). Conforme descrito até aqui, o papel da renda proveniente do trabalho teve grande peso na redução da pobreza verificada no país até o início da recessão, o que torna importante avaliar em que medida os mais pobres têm acesso a essas oportunidades. O fato de a população de renda mais baixa geralmente ocupar a periferia dos centros urbanos (ou áreas geograficamente isoladas no meio rural) torna

imprescindível a existência de um sistema de transporte acessível que permita o acesso aos locais que demandam mão-de-obra.

De modo mais geral, tanto o combate à pobreza pela via indireta do crescimento quanto pela via direta pressupõem a possibilidade de movimentar bens e pessoas. A mobilidade é condição muitas vezes necessária ao acesso, não só às oportunidades de trabalho, mas a todo o espaço urbano, propiciando acesso aos serviços básicos de educação e saúde e atividades de lazer (GOMIDE, 2003). Portanto, uma política de transporte coesa, adequada à realidade local, pode ser um poderoso instrumento de combate à pobreza e à exclusão social.

5.2 Impactos de intervenções no setor de transporte

O investimento em infraestrutura em geral, e o investimento em sistemas de transporte em particular, geralmente é visto como uma estratégia indireta de combate à pobreza, na medida em que pode ser motor de crescimento econômico. Nesse sentido, medidas tomadas pelo poder público na direção dos sistemas de transporte tendem a visar eficiência e não levam em consideração que, ao resolver carências de mobilidade proporcionando acesso aos mercados de bens e de trabalho, o crescimento econômico poderia ganhar mais autonomia, uma vez que os indivíduos a teriam para constantemente buscar novas oportunidades econômicas.

O debate acerca da eficiência dos sistemas de transporte orientados para a mobilidade e o acesso aos serviços essenciais deve levar em consideração o fato de que sistemas técnicos de alta tecnologia podem ter um alcance muito limitado devido à restrição orçamentária dos mais pobres e podem não ter muito sucesso na redução da desigualdade (GRIECO, 2015). Pelo contrário, é possível que os objetivos de eficiência e redução de custos do sistema entrem em conflito com objetivos de redução da desigualdade, isto é, projetos que priorizam a eficiência podem ter um impacto distributivo adverso.

Idealmente, conforme apontam Gannon e Liu (1997), isso não precisa ser um problema — uma vez identificado o custo de um projeto para a parcela menos favorecida da população, é possível desenhar mecanismos de compensação através de impostos e transferências, capazes de promover a redistribuição de renda.

No entanto, os países em desenvolvimento comumente dispõem de um sistema tributário lento, caro e ineficaz, o que dificulta sua utilização para compensação de impactos distributivos. Nesse sentido, torna-se importante pensar em maneiras de melhorar os serviços transporte sem onerar os mais pobres. Ou mais ainda, elaborar políticas de transporte que sejam instrumento direto de combate à pobreza, isto é, que tenham como principal

finalidade a melhoria do padrão distributivo e das condições de vida dos mais pobres (GANNON; LIU, 1997).

Embora o potencial de redução da pobreza seja uma justificativa recorrente de novos empreendimentos no setor de transportes, há evidências de que os pobres geralmente ficam excluídos do grupo de beneficiados. Isso pode ocorrer na forma de desapropriações, forçando-os a buscar residência em zonas mais periféricas das cidades sem receber compensação ou simplesmente de uma desvantagem relativa na parcela de mobilidade: há aumento na mobilidade de grupos de renda superior e diminuição na parcela referente ao grupo inferior (GRIECO, 2015).

Diante disso, surgiu na literatura associada ao planejamento urbano e de infraestrutura a preocupação de elaborar esquemas de transporte pró-pobres, como uma extensão da abordagem de planejamento voltada para a acessibilidade. Esta visa desenvolver o serviços e estabelecimentos urbanos de maneira localizada, reduzindo a necessidade de mobilidade, o congestionamento das vias, a poluição e a qualidade de vida no meio urbano.

Grieco (2015) reitera que é crucial desenvolver infraestrutura que dê suporte a esse padrão localizado de planejamento e que estão alinhados também à preocupação de inclusão social, como a que melhora as condições de transporte não motorizado e de pedestres. A autora defende que planejadores urbanos devem conciliar pressões por melhorias sustentáveis tanto do ponto de vista ambiental quanto do ponto de vista social e é possível que soluções implementadas em cidades ricas do “norte global” não possam ser importadas diretamente aos países do “sul global”, onde se encontram a maioria dos pobres do planeta (GRIECO, 2015).

Desse modo, os projetos realizados no setor de transportes podem ser orientados (a) à redução da pobreza; (b) à promoção da eficiência e do crescimento; ou (c) à maior sustentabilidade ambiental. Eles podem ainda ter uma orientação mista, com componentes que promovam ao mesmo tempo a eficiência e a redução da pobreza, por exemplo.

Sen (2000) classifica cinco tipos de liberdades instrumentais: liberdades políticas, facilidades econômicas, oportunidades sociais, garantias de transparência e segurança protetora. Todas contribuem para a capacidade geral das pessoas de viver livremente, mas têm efeito de complementar umas as outras. A análise do desenvolvimento deve considerar os encadeamentos empíricos que vinculam os tipos distintos de liberdade uns aos outros. A existência de tais encadeamentos é argumento em favor de que a liberdade seja o principal meio do desenvolvimento.

Em geral, espera-se que o impacto de projetos que visam a redução da pobreza sejam

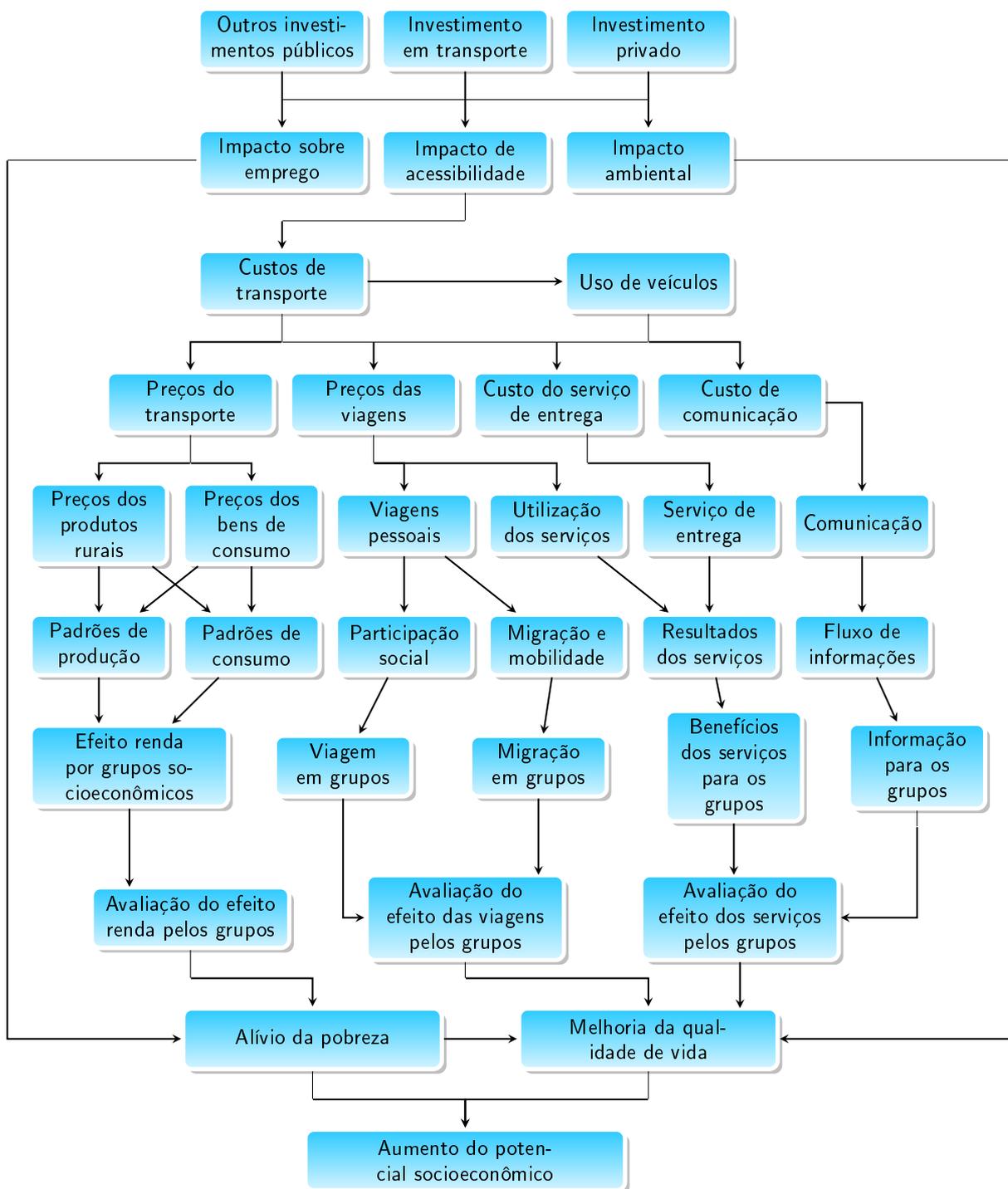
indiretos, através do crescimento econômico, ou diretos, com aumento do bem-estar dos mais pobres. Os mecanismos através dos quais investimentos em transporte se traduzem em melhoria no bem-estar dos mais pobres envolvem diversos encadeamentos e o resultado final pode ser muito difícil de prever (GANNON; LIU, 1997).

Por exemplo, investimentos no setor reduzem custos de transporte e aumentam o acesso a oportunidades, o que já tem efeito direto no bem-estar dos indivíduos. A contração dos custos e maior acesso podem resultar em redução de tarifas tanto de frete quanto de transporte de passageiros e, desde que haja algum nível de competitividade no mercado, pode resultar em redução do preço de bens e serviços, uma expansão geográfica dos mercados, maior mobilidade, entre outras consequências — as quais afetam indivíduos em todas as faixas de renda. Além disso, investimentos em infraestrutura de transporte afetam o emprego, uma vez que demandam mão-de-obra, muitas vezes de baixa qualificação, e então consistem em novas oportunidades de emprego aos mais pobres (GANNON; LIU, 1997).

São diversos os mecanismos através dos quais um sistema de transportes adequado pode colaborar com o crescimento econômico. Uma das funções do sistema de transportes é facilitar a interação entre diferentes atividades econômicas — ele reduz o custo de reunir os diferentes insumos de produção (energia, matérias primas, trabalho, informação e outros bens intermediários), proporcionando redução no custo de produção, o que pode reduzir os preços finais; além disso, a maior mobilidade de bens favorece o comércio local e internacional e permite economias de escala (GANNON; LIU, 1997). Investimento em transporte também contribui para diversificação econômica e permite explorar economias de escopo. Em suma, um bom sistema de transporte urbano traz externalidades positivas no sentido de melhorar a competitividade das cidades, aquecendo a atividade econômica, provocando impacto positivo no emprego e na renda, além de um ambiente mais atrativo para as firmas (GOMIDE, 2003).

Em sua relação com o crescimento econômico, as políticas de transporte podem desempenhar um papel apenas complementar, quando servem para adequar a estrutura existente à demanda já em crescimento, ou um papel de liderança, quando estimulam o crescimento econômico através da expansão dos mercados. É amplamente reconhecido que um bom sistema de transportes é uma condição necessária, conquanto não suficiente, para o crescimento econômico, que por sua vez amplia a demanda por transporte. Há poucos estudos empíricos acerca da contribuição de investimentos em transporte como motores (em vez de como acessórios) do crescimento econômico e eles são feitos em um nível agregado, estimando o efeito de investimentos em infraestrutura como um todo. Apesar de identificar um efeito positivo, tais estudos não são capazes de discernir as condições necessárias para que ele exista. Mesmo assim, é aceito que o papel de infraestrutura de transporte

Figura 3 – Mecanismos de impacto do investimento no setor de transporte no crescimento econômico.



Fonte: Adaptado de Gannon e Liu (1997).

no processo de crescimento econômico é significativo e positivo (GANNON; LIU, 1997).

O que se pode afirmar é que a mobilidade é fundamental para a criação de capital humano, necessário ao crescimento econômico sustentado e também à superação da condição de exclusão. Ademais, a ineficiência do sistema de transporte faz com que as pessoas demorem mais do que o necessário para percorrer um determinado trajeto, exigindo dispêndio de tempo e energia e prejudicando indicadores de produtividade dos trabalhadores (GOMIDE, 2003).

No meio urbano, as famílias mais pobres tendem a estar situadas nas periferias ou em lugares de difícil acesso, frequentemente distantes de seus locais de trabalho e sem condições de realocação residencial. Muitos trabalhadores de baixa renda têm mais de um emprego em locais distintos e um sistema de transportes deficiente ou muito caro pode prejudicar sua habilidade de obter ou manter o emprego (GANNON; LIU, 1997). Segundo Gomide (2003), também pode ocorrer da escolha de moradia estar condicionada à proximidade do local de trabalho. Ao não poder optar pelo local de trabalho, buscam morar perto dele e evitar gastos de transporte. O resultado desse fenômeno são bolsões de pobreza em áreas onde há ocupação irregular dos terrenos das cidades, como favelas e cortiços.

Um fato pouco lembrado é que a própria busca por emprego exige dispêndio com tarifas de transporte, o que pode impedir a parcela de extremamente pobres da população de acessar tais oportunidades.

Famílias que residem no meio rural tendem a encontrar-se em condições de isolamento geográfico e carência um esquema de transporte adequado às suas necessidades, com insuficiência dos serviços públicos existentes e agravada pelo alto custo de aquisição de veículos próprios (GANNON; LIU, 1997).

Em ambos os casos, rural e urbano, é patente que um bom sistema de transporte pode ter enorme impacto no bem-estar das pessoas de baixa renda. Nota-se que o problema adquire complexidade, uma vez que não se trata apenas de apontar que os custos do transporte coletivo ou individual são proibitivos para uma fração significativa da população brasileira.

5.3 Padrão de mobilidade da população de baixa renda

A renda é o principal determinante do comportamento relacionado ao transporte. Rendas mais altas estão associadas a mais viagens, distâncias percorridas maiores e meios de transporte mais rápidos e confortáveis (GANNON; LIU, 1997). Na região metropolitana de São Paulo, o índice geral de mobilidade (número de viagens diárias por habitante, conforme calculado em Companhia do Metropolitano de São Paulo (2013)) cresce com

Tabela 4 – Índice de mobilidade de acordo com a renda para os anos de 2007 e 2012.

2007					
Renda familiar (*)	Índice de mobilidade (viagens diárias/habitante)				
	Coletivo	Individual	Motorizado	Não Motorizado	Total
até 1244	0,59	0,18	0,77	0,82	1,58
1244 a 2488	0,75	0,30	1,05	0,75	1,80
2488 a 4976	0,81	0,67	1,48	0,59	2,08
4976 a 9330	0,66	1,35	2,01	0,42	2,43
9330 e mais	0,42	1,97	2,39	0,34	2,73
TOTAL	0,71	0,58	1,29	0,66	1,95
2012					
Renda familiar (*)	Índice de mobilidade (viagens diárias/habitante)				
	Coletivo	Individual	Motorizado	Não Motorizado	Total
até 1244	0,65	0,22	0,87	0,86	1,73
1244 a 2488	0,88	0,38	1,26	0,78	2,04
2488 a 4976	0,88	0,84	1,72	0,62	2,34
4976 a 9330	0,72	1,39	2,12	0,52	2,63
9330 e mais	0,57	1,78	2,35	0,44	2,79
TOTAL	0,81	0,68	1,49	0,70	2,18

* Em reais de outubro de 2012.

Fonte: Adaptado de NTU (2016).

a faixa de renda e também com o nível de escolaridade, conforme pode-se observar nas tabelas 4 e 5, que apresentam dados referentes aos anos de 2007 e 2012.

No caso das famílias de baixa renda, o transporte urbano compõe uma parcela significativa das despesas com serviços públicos. De acordo com dados do IPEA, em 2009 a parcela das famílias correspondente aos 10% mais pobres (1º decil) comprometia em torno de 13,5% da renda com transporte, enquanto as famílias do 1º e 2º decis comprometiam em torno de 10,3% da renda. Além disso, “entre os 10% mais pobres cerca de 30% das famílias não efetuam gastos com TPU³, o que é um indicador de exclusão dessas famílias em função da falta de capacidade de pagamento pelos serviços de transporte (CARVALHO *et al.*, 2013, p. 4).” Fica evidente que a redução ou total isenção dos custos de transporte pode aumentar consideravelmente a renda disponível das famílias.

Qualquer tentativa de reduzir a pobreza deve levar em consideração as características locais das famílias de baixa renda, seus hábitos de transporte e suas necessidades específicas. O padrão de ocupação do espaço urbano e da localização espacial dos mais pobres pode

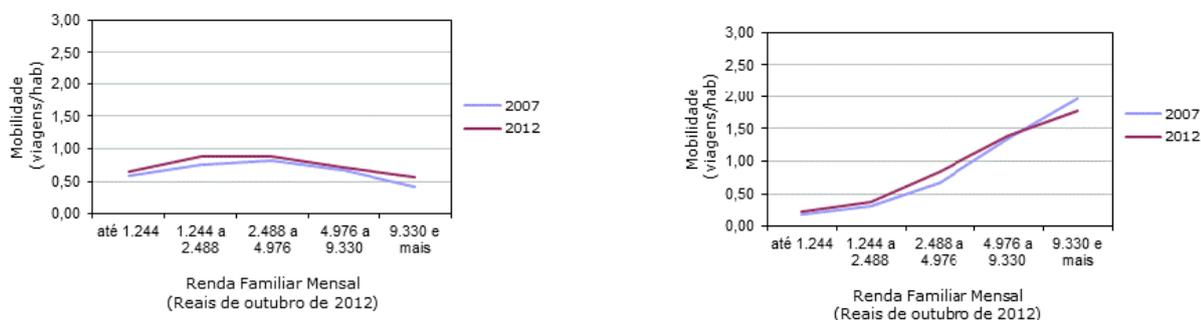
³ Transporte público urbano.

Tabela 5 – Mobilidade de acordo com escolaridade para os anos de 2007 e 2012.

2007					
Escolaridade	Índice de mobilidade (viagens diárias/habitante)				
	Coletivo	Individual	Motorizado	Não Motorizado	Total
Analfabeto/ 1º grau incompleto	0,52	0,29	0,81	0,76	1,56
1º grau completo/ 2º incompleto	0,78	0,44	1,22	0,88	2,10
2º grau completo / Superior incompleto	1,01	0,74	1,75	0,48	2,22
Superior completo	0,63	1,71	2,34	0,38	2,72
TOTAL	0,71	0,58	1,29	0,66	1,95

2012					
Escolaridade	Índice de mobilidade (viagens diárias/habitante)				
	Coletivo	Individual	Motorizado	Não Motorizado	Total
Analfabeto/ 1º grau incompleto	0,62	0,35	0,97	0,80	1,77
1º grau completo/ 2º incompleto	0,84	0,51	1,35	0,84	2,19
2º grau completo/ Superior incompleto	1,09	0,86	1,95	0,54	2,49
Superior completo	0,68	1,63	2,31	0,53	2,85
TOTAL	0,81	0,68	1,49	0,70	2,18

Fonte: Adaptado de NTU (2016).



(a) Índice de mobilidade por modo coletivo e renda familiar mensal – 2007 e 2012. Região metropolitana de São Paulo.

(b) Índice de mobilidade por modo individual e renda familiar mensal – 2007 e 2012. Região metropolitana de São Paulo.

Figura 4 – Mobilidade por modo de viagem.

variadas e políticas de transporte pró-pobre devem levar esses fatores em consideração. Por exemplo, os pobres, sobretudo as pessoas ocupadas no mercado informal, tendem a restringir deslocamentos motorizados e substituí-los por deslocamentos a pé e a maior parte de suas viagens são para escola e trabalho. Para pessoas com renda familiar de até um salário mínimo, Gomide (2003) aponta que o motivo “escola” supera o motivo “trabalho”, uma vez que essas pessoas auferem sua renda do trabalho informal ou doações. Além disso, revela: “Entre os motivos pelos quais os pobres viajam a pé prevalecem a falta de dinheiro e a disponibilidade do serviço onde moram (incluindo a baixa frequência).”

De acordo com o levantamento feito pelo Metropolitano de São Paulo (COMPANHIA DO METROPOLITANO DE SÃO PAULO, 2013), quanto maior a renda dos indivíduos, maior é a utilização de meios motorizados como forma de transporte escolhida. A faixa de renda mais baixa (até R\$1244) apresentava 50% de participação dos meios não motorizados e 50% de meios motorizados. Desta, 75% consistia de transporte coletivo, proporção que cai quanto mais alta a faixa de renda considerada (COMPANHIA DO METROPOLITANO DE SÃO PAULO, 2013). A figura 4 retrata a divisão modal (coletivo e individual) por faixa de renda para a Região Metropolitana de São Paulo.

5.4 Políticas públicas de transporte

Nas discussões referentes ao transporte, a preocupação com a sustentabilidade ambiental tem ganhado peso nos últimos anos. Já o aspecto da sustentabilidade social carece de dados e melhor avaliação dos programas existentes, de modo que se possa de fato concluir quais políticas são eficazes em promover inclusão social e o desenvolvimento. Uma questão comumente levantada é a ausência ou insuficiência de dados que permitam uma avaliação das condições de mobilidade e acesso, do impacto sobre a parcela menos favorecida da população, entre outras informações relevantes. Nas palavras de Grieco (2013), “*the literature is scant, the best practices few, the databases absent*”.

Ressalta-se a importância de, ao elaborar uma nova política, conhecer profundamente a realidade local e as necessidades dos pobres, de modo que não haja desperdício de recursos ou que os benefícios não sejam apropriados por segmentos mais favorecidos da população. Além disso, as melhores propostas implementadas com sucesso em países desenvolvidos, que combinam sustentabilidade social e ambiental com eficiência podem não ser aplicáveis à realidade brasileira.

O espectro de intervenções possíveis é muito amplo e há pouco consenso quanto ao melhor caminho. No entanto, algumas diretrizes básicas podem ser apontadas, como a priorização de meios não motorizados e pedestres. Gomide (2003) atenta para o fato de que os benefícios gerados pelas tradicionais medidas de mobilidade urbana, como a expansão da

malha viária, tendem a ser capturados pelos usuários de automóveis, os quais encontram-se em faixas de renda mais elevadas. Desse modo, priorizar o transporte coletivo e não motorizado consiste em uma medida de inclusão social e intervenções que melhorem as condições de utilização destes tendem a ter impacto direto sobre o bem-estar dessa parcela da população. Todavia, o custo de veículos não motorizados pode ser proibitivo para essas pessoas e iniciativas como programas de financiamento desse tipo de veículo pode ser uma alternativa economicamente viável (GANNON; LIU, 1997).

De acordo com Gannon e Liu (1997), esforços para aliviar a pobreza no meio rural são mais facilmente implementados, pois esses locais tendem a ser geograficamente isolados e a população pobre desses locais relativamente mais homogênea do que no meio urbano, o que torna as políticas mais eficazes. Emprego em obras de infraestrutura (manutenção de estradas, por exemplo) pode consistir em novas oportunidades de renda, em que há grandes incentivos à participação, uma vez que no meio rural geralmente há menos oportunidades desse tipo. Nesses locais, as intervenções diretas devem ser orientadas para a acessibilidade aos serviços básicos, buscando incluir a comunidade afetada na tomada de decisões. Uma alternativa é a redução da necessidade de deslocamento através da provisão local desses serviços (GRIECO, 2013).

No meio urbano, a população pobre tende a ser bem mais heterogênea, dificultando ou reduzindo a eficácia de intervenções diretas. Diversas modalidades de serviço competem em um mesmo espaço e intervenções podem levar a distorções no mercado e nas escolhas dos indivíduos, o que também dificulta sua avaliação. Muitas famílias de baixa renda não possuem imóvel próprio e medidas que valorizem os terrenos adjacentes (como manutenção das vias, melhoria das condições de transporte no bairro, expansão da linha de metrô, e assim por diante) podem aumentar os preços dos aluguéis, expulsando os pobres para regiões desvalorizadas. Desse modo, é possível que os benefícios sociais de uma política voltada às camadas inferiores da distribuição de renda sejam capturados pelos proprietários dos imóveis, os quais não compunham o público alvo da intervenção (GANNON; LIU, 1997).

A gratuidade ou desconto na tarifa para segmentos da população é uma política largamente utilizada no Brasil. No entanto, Gomide (2003) aponta que essas concessões não obedecem a um critério de renda e tem um efeito regressivo na distribuição de renda, uma vez que um trabalhador informal sem renda fixa paga o valor integral da tarifa, enquanto estudantes de escolas privadas recebem descontos. O autor sugere, por exemplo, que os estudantes da rede pública recebessem suas passagens de suas respectivas escolas, financiadas pela própria Secretaria de Educação, isto é, com recursos extratarifários, de modo que a isenção não onere os demais usuários do sistema.

Já Gannon e Liu entendem que os subsídios aos usuários mais pobres são ineficazes. Os autores levantam a preocupação de que as operadoras de transporte capturem o benefício sob forma de custos de operação inflados ou ainda pelos proprietários de imóveis em áreas beneficiadas por subsídios na forma de valorização dos terrenos, decorrente da melhoria de acesso a esses locais. Subsídios na forma de tarifas baixas também podem falhar em chegar à população alvo se esta simplesmente não utiliza os serviços de transporte ou não os utiliza o suficiente para experimentar uma real melhoria de bem-estar (GANNON; LIU, 1997). Devido à ineficácia dos subsídios às operadoras de transporte, recomenda-se explorar outras opções para ajudar os pobres do meio urbano, que busquem reduzir o tempo e o custo das viagens, fornecimento de estruturas mais seguras — visto que os problemas de segurança afetam os pobres de maneira desproporcional — e melhor qualidade para circulação de pedestres e veículos não motorizados.

5.4.1 Universalização do acesso

Estudos empíricos acerca do impacto da universalização do acesso ao transporte não são de conhecimento da autora, todavia, Goodman *et al.* (2014) fornece evidências de como o passe livre de ônibus para menores de 18 anos implementado em Londres afeta a mobilidade independente dos jovens. Valendo-se da abordagem de expansão das capacidades individuais de Amartya Sen, Goodman *et al.* (2014) defende que a mobilidade independente é uma capacidade absolutamente necessária à vida humana. O artigo propõe uma nova maneira de pensar a mobilidade independente — enquanto para os adultos isso comumente é entendido como ser capaz de viajar sozinho, no caso dos jovens a independência se reflete em viagens em grupo, espontâneas.

No caso dos jovens londrinos, o maior impacto nas decisões de transporte foram aquelas descritas como “desnecessárias”, isto é, não relacionadas à escola, emprego e outras obrigações, mas voltadas à convivência social e viagens exploratórias pela cidade. Isso é importante pois indica que a isenção tem impacto sobre os usuários que recebem vale transporte e, por conseguinte, podem arcar com o custo da tarifa para ir ao trabalho, mas não podem financiar as “viagens desnecessárias”, voltadas ao lazer. O problema é que tal “mobilidade desnecessária” em diversos casos constitui a principal maneira através da qual os indivíduos podem participar e fazer sentido do espaço urbano (GOODMAN *et al.*, 2014). Desse ponto de vista, os autores identificam a emergência propriedades em nível de grupo, isto é, propriedades em que o efeito total é maior do que a soma de suas partes componentes. O estudo conclui que a universalização do acesso vai além do benefício em nível individual proporcionado pela liberdade de ir e vir, o que pode ser especialmente importante entre os jovens.

From a social policy perspective, it is important to note that the magnitude of this group-level effect was crucially dependent on the universal rather than targeted nature of the entitlement. This universal entitlement resulted in near-universal coverage, which in turn extended the independent mobility of both groups [...] and individuals (GOODMAN et al., 2014, p. 15).

Outro fenômeno emergente identificado em Londres é que o fato de o ônibus ser utilizado por todos os membros de um grupo pode ter minado a percepção do transporte coletivo como sendo relacionado às classes mais baixas (GOODMAN *et al.*, 2014).

6 Resultados e discussão

Neste capítulo, serão apresentadas informações acerca da obtenção e tratamento dos dados, detalhes da implementação dos métodos de pareamento e do modelo de efeitos fixos, bem como os resultados obtidos em cada etapa.

6.1 Origem e tratamento dos dados

As características municipais utilizadas como variáveis explicativas na estimação do escore de propensão foram obtidas a partir dos dados censitários do IBGE do ano de 2000, que consiste em ano pré-tratamento para todos os municípios incluídos na análise. Foram coletadas 28 características de 5565 municípios, as quais foram filtradas e posteriormente utilizadas para estimar o escore de propensão. A tabela 12 apresenta as estatísticas descritivas dessas variáveis.

Já os níveis de emprego e salário, variáveis sobre as quais desejava-se medir o impacto da tarifa zero, foram obtidos em frequência mensal a partir do Cadastro Geral de Empregados e Desempregados (CAGED). Estes estavam disponíveis a partir de janeiro de 2004, totalizando 160 períodos até abril de 2017. A base com o estoque mensal de emprego foi montada a partir dos dados fornecidos pelo sistema do CAGED (variação mensal da quantidade de empregos formais) e dados de estoque de dezembro de 2015 fornecidos diretamente por um funcionário do Programa de Disseminação de Estatísticas do Trabalho (PDET).

A seguir, fez-se necessária uma triagem dos dados obtidos para identificação de *outliers* ou informações faltantes. Tanto da base de emprego quanto da base de salário, foram eliminados municípios que possuíam mais de 5% das observações iguais a 0. No caso dos dados de emprego, isso acarretou na eliminação de 194 municípios e no caso dos salários em 1203 municípios.

Um critério comum de identificação de observações discrepantes é o de determinar quais delas são superiores a um nível de tolerância dado por $t = Q_3 + 1,5 \times IQ$, em que Q_i é o quartil i da variável e $IQ = Q_3 - Q_1$ é a distância interquartilica, isto é, a diferença entre o terceiro e o primeiro quartil. Uma vez que buscava-se identificar os *outliers* na série de um mesmo município — e que esta apresenta tendência ao longo do tempo — foi utilizado um critério mais amplo, dado por $3 \times t$. Com respeito a *outliers* para baixo, o critério foi um pouco diferente, $Q_1 - 4,5 \times IQ$.

A partir do critério supracitado, foram identificados 90 municípios com observações discrepantes de salário, todos com menos de 2,5% de observações problemáticas (totalizando 113 observações). Diante disso, optou-se por retirar essas 90 observações (mas manter os municípios), o que produziu um painel com observações faltantes. Dois exemplos ilustrativos são os municípios de Rio Bom, no Paraná, e Terra Alta, Pará. A tabela 6 exhibe uma parte da série de salário médio desses municípios. Em 2013, o município de Rio Bom apresentou três observações muito acima das demais, enquanto Terra Alta apresenta uma observação alta em 2010.

Tabela 6 – Exemplos de municípios com observações discrepantes de salário.

	Abr/2013	Mai/2013	Jun/2013	Jul/2013	Ago/2013	Set/2013
Rio Bom (PR)	792,19	805,22	25069,66	4371,52	4277,93	812,23
	Jan/2010	Fev/2010	Mar/2010	Abr/2010	Mai/2010	Jun/2010
Terra Alta (PA)	612,00	570,00	7751,11	613,67	690,91	654,67

Fonte: Elaboração própria com base nos dados do CAGED.

Na base de emprego, foram identificados 21 municípios com observações acima desse critério, mas suas observações foram mantidas, pois removê-las acarretaria em impacto sobre todos os períodos subsequentes, diferentemente da série de salários. Como o nível de emprego só foi fornecido para dezembro de 2015 e o estoque dos períodos subsequentes e anteriores foram obtidos somando os dados mensais de fluxo, uma única observação discrepante impacta todos os períodos subsequentes. Analogamente, removê-la não implicaria em ausência de informação para o período, mas sim que a variação do emprego no período foi 0. Após o procedimento descrito, restaram 4161 municípios.

Deve-se reconhecer que a remoção de municípios na etapa de limpeza dos dados poderia incorrer em viés das estimativas finais. Não há meios de identificar o motivo da ocorrência de observações discrepantes ou ausentes; sabe-se, por exemplo, que a obrigatoriedade de enviar os registros de contratações e demissões por parte dos estabelecimentos é recente, o que pode ter afetado os registros na base de dados. Todavia, a imputação desses dados com médias, medianas ou outro método mais sofisticado requereria um estudo mais detalhado acerca das condições em que tais observações ocorreram, além das referidas técnicas de imputação, o que considera-se além do escopo deste trabalho.

6.2 Características dos municípios tratados e não tratados

Uma vez que a série histórica só estava disponível a partir de 2004, os municípios que optaram por implementar a gratuidade em períodos precedentes não puderam ser incluídos

no modelo. Desse modo, permaneceram:

- | | |
|--------------------------------|----------------------------------|
| a) Eusébio – CE (Out 2010); | e) Paulínia – SP (Ago 2013); |
| b) Itatiaiuçu – MG (Jul 2015); | f) Pitanga – PR (Fev 2012); |
| c) Maricá – RJ (Dez 2014); | g) Porto Real – RJ (Set 2011); |
| d) Muzambinho – MG (Out 2011); | h) Silva Jardim – RJ (Fev 2014). |

Então, de um total de 13 municípios que adotam a tarifa zero no Brasil, apenas 8 puderam ser avaliados. Na tabela 7 é possível comparar as médias das características coletadas para ambos os grupos. Observa-se que os municípios tratados apresentam indicadores melhores de desenvolvimento, como maior expectativa de vida ao nascer, maior IDHM, maior proporção de municípios com acesso a água encanada, maior renda per capita e expectativa de anos de estudo, menor proporção de pobres e menor taxa de analfabetismo, entre outras. Para verificar se essas diferenças são estatisticamente significativas, realizou-se um teste de diferença de médias para cada uma dessas variáveis, a fim de identificar diferenças entre os grupos e prospectar quais poderiam ser utilizadas para a estimação do escore de propensão. A tabela 13 apresenta os resultados dos testes de Mann-Whitney-Wilcoxon¹.

De fato, os resultados do teste Mann-Whitney-Wilcoxon apresentados na tabela 13, localizada no apêndice deste trabalho, confirmam que alguns desses indicadores são significativamente distintos para os municípios tratados, como taxa de analfabetismo, proporção de pobres e IDHM. Também deve-se notar que, em abril de 2017, tanto o nível de emprego quanto o de salário médio são superiores no grupo de municípios tratados, resultado que pode ser visto na tabela 8.

6.3 Pareamento

6.3.1 Escolha das variáveis

A aplicação prática do escore de propensão de interesse para este trabalho foi a construção de uma amostra indivíduos do grupo de controle, cuja distribuição das covariáveis seja similar à dos indivíduos tratados. De acordo com Rosenbaum e Rubin (1983), algumas vantagens do pareamento utilizando escore de propensão são que: (I) a equivalência entre os grupos de controle e tratamento é relativamente simples de identificar; (II) a variância da estimativa de efeito causal é menor em amostras pareadas, uma vez que a distribuição

¹ Este é um teste não paramétrico aplicado a amostras independentes. Sua aplicação não exige supor que a distribuição das variáveis é normal e por isso foi utilizada. A estatística de teste de Mann-Whitney-Wilcoxon é o análogo não-paramétrico à estatística t (CORDER; FOREMAN, 2014).

Tabela 7 – Médias das variáveis por grupo (tratamento e controle).

Variável	Controle	Tratados
% da renda apropriada pelos 20% mais ricos	58,895	60,499
Coefficiente de Gini	0,547	0,559
Taxa de atividade (18 anos ou mais)	62,951	65,032
% empregados com carteira assinada	23,401	39,249
% ocupados com ensino fundamental completo	28,206	36,865
% ocupados no setor agropecuário	42,516	19,695
% ocupados na indústria de transformação	9,408	12,910
% ocupados no setor de serviços	31,560	41,803
% ocupados com renda de até um salário mínimo	64,595	48,339
% domicílios com água encanada	66,718	82,620
Esperança de vida ao nascer	68,409	70,035
População total	30509,392	32294,000
Proporção da população no meio urbano	0,585	0,778
Mortalidade infantil	33,279	24,851
PEA (18 anos ou mais)	12845,96	13792,25
IDHM	0,523	0,590
Taxa de fecundidade	2,870	2,485
Razão de dependência	62,272	55,070
Taxa de envelhecimento	6,459	5,676
Taxa de analfabetismo	23,572	14,460
% de adultos com fundamental completo	23,725	31,751
Expectativa de anos de estudo	8,338	8,875
Taxa de frequência líquida ao ensino básico regular	84,449	88,164
Renda <i>per capita</i> exceto renda nula	347,129	511,171
Renda proveniente do trabalho	74,214	75,037
Renda <i>per capita</i> do quinto mais rico	998,428	1523,796
% de pobres	41,079	25,666
% vulneráveis à pobreza	63,919	49,450
Renda <i>per capita</i> dos pobres	75,302	83,764

Fonte: Elaboração própria a partir dos dados do censo de 2000.

Tabela 8 – Médias das variáveis de resultado em abril de 2017 por grupo.

Grupo	Emprego	Salário
Controle	8936,94	1299,29
Tratados	10101,15	1431,67

Fonte: Elaboração própria a partir dos dados do CAGED.

de X é mais parecida entre tratados e controle, se comparada com amostras aleatórias; (III) ajustes no modelo em amostras pareadas geralmente são mais robustos se comparados com amostras aleatórias.

Retomando o que foi discutido no capítulo 3, as variáveis incluídas na estimação do escore de propensão deveriam ser aquelas que pudessem influenciar a decisão de tratamento (adoção da política) dos municípios, assim como aquelas que poderiam provocar diferenças sistemáticas na variável de resultado dos dois grupos (CALIENDO; KOPEINIG, 2008).

A gratuidade no transporte coletivo de ônibus oferecida a todos cidadãos como política de transporte não consiste em um programa do governo, com regras sistemáticas e critérios claros de elegibilidade. Desse modo, não é implementado de forma homogênea nos diferentes municípios que a adotaram. O processo de decisão que levou ao tratamento envolve decisões políticas e contextos locais que estão fora do escopo deste trabalho e seriam melhor identificadas em estudos de caso. Ante esse obstáculo, buscou-se prever a participação com base em características que poderiam ser medidas para todos os municípios e que acredita-se que tenham influência na decisão de adotar a política.

Que características distintivas podem ser observadas nos municípios que implementam a tarifa zero? Ou melhor, o que os leva a tomar essa decisão? Não foram encontrados trabalhos que busquem responder a essas perguntas. No entanto, é possível conjecturar algumas circunstâncias sob as quais tal decisão política pode vir a ganhar força.

Vale lembrar que a tarifa zero é uma reivindicação típica de movimentos sociais, mas com pouca inserção no debate política oficial. A Associação Nacional das Empresas de Transporte Urbano realizou uma pesquisa de opinião acerca de diferentes questões que permeiam o tema das políticas de transporte. De acordo com o documento, “foram entrevistados deputados federais, senadores, representantes dos Poderes Executivos Federal, Estaduais e Municipais, acadêmicos e especialistas em transporte, economistas e representantes de entidades ligadas aos setores da indústria, comércio e serviços” (NTU, 2016). O nível de rejeição da proposta de tarifa zero é alto: 71,8 % dos entrevistados do congresso nacional e 85,0% dos “influenciadores de opinião” (economistas, representantes da indústria, entre outros) rejeitam a proposta.

A justificativa da proposta de tarifa zero no âmbito do movimento Passe livre é calcada no combate à exclusão social. No capítulo 5, discorreu-se sobre a relação da exclusão social com o tema de transporte. Em particular, discutiu-se que a condição de exclusão social vem acompanhada de falta de representatividade nos processo de decisão política. Então, é plausível inferir que, quando tal proposta chega às vias de fato, as classes inferiores dispõem de algum nível de representatividade, o que não parece ser possível em um

contexto de desigualdade extrema, ou de níveis muito altos de pobreza absoluta. Em síntese, presume-se que municípios com indicadores de desenvolvimento muito ruins não consistem em um ambiente favorável para o florescimento de uma proposta dessa natureza.

Inicialmente, as características coletadas como candidatas a covariáveis foram as seguintes:

- | | |
|---|---|
| 1. % da renda apropriada pelos 20% mais ricos; | 15. PEA (18 anos ou mais); |
| 2. coeficiente de Gini; | 16. IDHM; |
| 3. taxa de atividade (18 anos ou mais); | 17. taxa de fecundidade; |
| 4. % empregados com carteira assinada; | 18. razão de dependência ³ ; |
| 5. % ocupados com ensino fundamental completo; | 19. taxa de envelhecimento; |
| 6. % ocupados no setor agropecuário; | 20. taxa de analfabetismo (maiores de 18 anos); |
| 7. % ocupados na indústria de transformação; | 21. % de maiores de 18 anos com fundamental completo; |
| 8. % ocupados no setor de serviços; | 22. expectativa de anos de estudo; |
| 9. % ocupados com renda de até um salário mínimo; | 23. taxa de frequência líquida ao ensino básico regular; |
| 10. % domicílios com água encanada ; | 24. renda <i>per capita</i> exceto renda nula; |
| 11. esperança de vida ao nascer; | 25. renda <i>per capita</i> média do quinto mais rico; |
| 12. população total; | 26. % pobres (renda abaixo de R\$ 140,00); |
| 13. proporção da população urbana ² ; | 27. % vulneráveis à pobreza (renda abaixo de R\$ 255,00); |
| 14. mortalidade infantil; | 28. renda <i>per capita</i> dos pobres. |

Essas características foram obtidas a partir dos dados do censo de 2000, que consiste em um ano pré tratamento para todos os municípios selecionados para o estudo. Assim, garantiu-se que a fonte dos dados foi consistente, medindo as características de todos os municípios a partir do mesmo critério, evitando parte das fontes de viés relacionadas à

² Obtida dividindo a população residente no meio urbano pela população total.

³ A razão de dependência é a razão entre o número de pessoas com 14 anos ou menos e de 65 anos ou mais de idade e o número de pessoas com idade de 15 a 64 anos.

origem dos dados identificadas no trabalho de Heckman, Ichimura e Todd (1997), conforme argumentado no capítulo 3.

Embora poderia-se esperar que outras características influenciassem a decisão de participação, como por exemplo o tamanho da população total do município, nem sempre o modelo de previsão incorria em estimativas melhores. O tamanho da População Economicamente Ativa (PEA - maiores de 18 anos), a população total do município e a razão dos 10% mais ricos para os 40% mais pobres proporcionaram separação completa dos grupos de tratamento e controle, o que não é desejável:

[...] it should also be clear that ‘too good’ data is not helpful either. If $P(X) = 0$ or $P(X) = 1$ for some values of X , then we cannot use matching conditional on those X values to estimate a treatment effect, because persons with such characteristics either always or never receive treatment (CALIENDO; KOPEINIG, 2008, p. 6).

Em outras palavras, essas características fariam violar a hipótese de suporte comum (3.2) pois a probabilidade de participação dadas as características observáveis deve ser estritamente positiva para ambos os grupos. Então, essas variáveis não poderiam ser utilizadas para estimar o escore de propensão.

Tabela 9 – Modelo logit para previsão de participação no programa

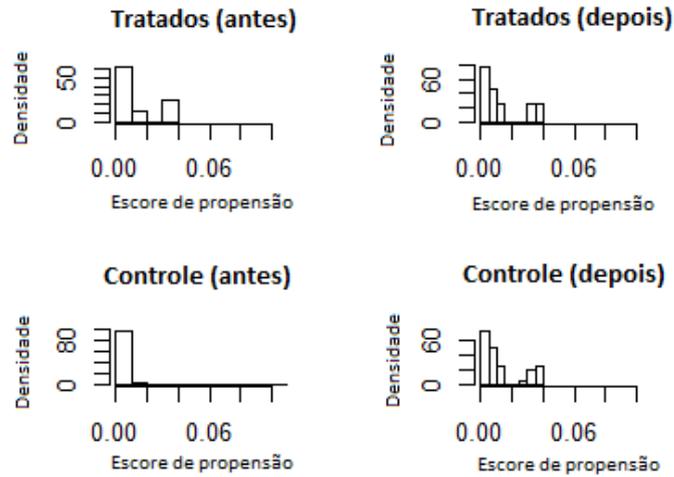
	<i>Dependent variable:</i>	
	Tratamento	p-valor
Coefficiente de Gini	18,061	0,167
Esperança de vida ao nascer	-0,538*	0,085
Mortalidade infantil	-0,131	0,158
IDHM	18,432	0,328
% ocupados com fundamental completo	-0,181**	0,049
% ocupados no setor agropecuário	-0,106***	0,006
% ocupados na indústria de transformação	-0,034	0,433
Renda <i>per capita</i> exceto renda nula	0,019	0,245
Renda <i>per capita</i> do quinto mais rico	-0,006	0,284
Constante	23,023	0,282
Observações	4161	
Log Likelihood	-49,855	
Crit. Inf. Akaike	119,710	

Nota:

* p<0,1; ** p<0,05; *** p<0,01

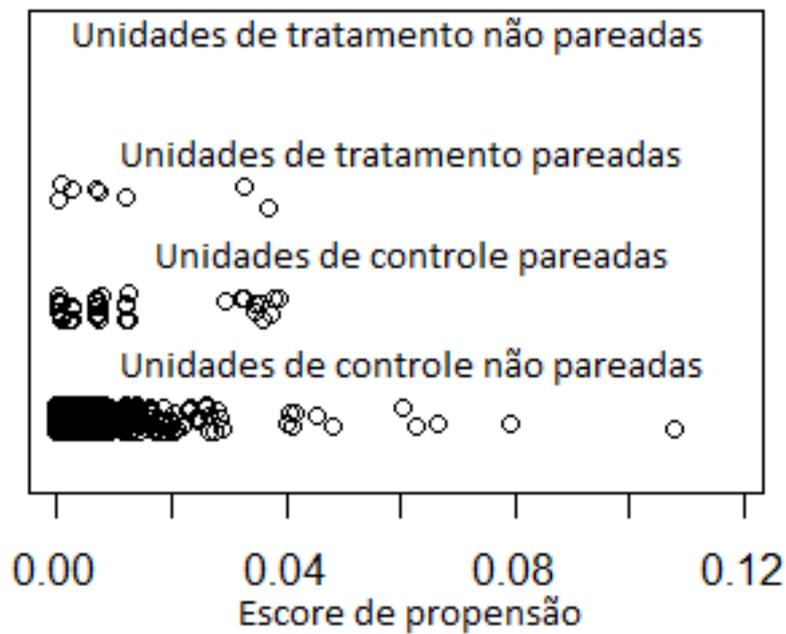
Fonte: Elaboração própria.

Figura 5 – Comparação dos histogramas dos escores de propensão antes e depois do pareamento.



Fonte: Elaboração própria.

Figura 6 – Gráfico de dispersão das unidades tratadas e de controle, pareadas e não pareadas. Todas as unidades de tratamento foram pareadas.



Fonte: Elaboração própria.

Não há consenso quanto à melhor estratégia para escolher as variáveis incluídas no modelo. Se, por um lado, a omissão de variáveis importantes pode resultar em viés considerável, conforme o estudo de Heckman, Ichimura e Todd (1997), por outro, há razões para evitar modelos com variáveis demais. São duas as principais justificativas, a saber, incluir muitas variáveis no modelo de participação pode reduzir a região de suporte comum e, além disso, embora a inclusão de variáveis não significativas não torne as estimativas enviesadas ou inconsistentes, pode aumentar sua variância (BRYSON; DORSETT; PURDON, 2002).

Uma vez removidas as variáveis que proporcionavam separação completa, prosseguiu-se re-estimando o modelo logit e retirando, uma a uma, as variáveis menos significativas estatisticamente. Ao parear os municípios, uma inspeção visual da dispersão dos escores de propensão foi utilizada como indicação da qualidade do pareamento, seguindo orientação de Bryson, Dorsett e Purdon (2002). Variáveis foram removidas a cada iteração até que, ao escolher os vizinhos mais próximos, a distância máxima entre algum par fosse menor do que 0,05. Os resultados dos modelos descartados não serão apresentados.

A título de ilustração, a tabela 9 apresenta os resultados do modelo logit final para previsão de participação. Para a realização do pareamento foi utilizado um pacote estatístico específico (MatchIt), o qual omite o resultado da regressão utilizada para estimativa do escore de propensão. Deve-se ressaltar que, apesar de o coeficiente de certas variáveis não ser estatisticamente significativo, elas foram mantidas no modelo a fim de preservar ainda um conjunto amplo de características municipais, conforme as recomendações de Heckman, Ichimura e Todd (1997) abordadas no capítulo 3. As covariáveis remanescentes foram:

- a) Coeficiente de Gini;
- b) Esperança de vida ao nascer;
- c) Mortalidade infantil;
- d) IDHM (Índice de Desenvolvimento Humano dos municípios);
- e) % ocupados com ensino fundamental completo;
- f) % ocupados no setor agropecuário;
- g) % ocupados na indústria de transformação;
- h) Renda *per capita* exceto renda nula; e
- i) Renda *per capita* do quinto mais rico.

6.3.2 Algoritmo de pareamento

Utilizando o critério do vizinho mais próximo, o escore de propensão foi utilizado para selecionar um total de 5 indivíduos do grupo de controle correspondentes a cada indivíduo tratado (sobreamostragem), resultando em uma amostra de 48 municípios no total. A tabela 14, localizada no apêndice deste trabalho, apresenta os municípios selecionados para compor o grupo de controle. A sobreamostragem envolve um *tradeoff* entre viés e variância. Quanto maior o número de indivíduos selecionados, maior é a variância, uma vez que aumenta a informação utilizada na construção do contrafactual, ao passo que se o número é muito pequeno, incorre-se em maior viés decorrente do pareamento ser pior (CALIENDO; KOPEINIG, 2008). Como o número de municípios que adotaram a gratuidade no transporte coletivo é pequeno, o pareamento 1 para 1 forneceria uma amostra muito reduzida para a posterior estimação do efeito causal.

Conforme argumentado no capítulo 3, a estimativa de impacto só se aplica à região de suporte comum. Desse modo, torna-se importante verificar se há sobreposição das distribuições dos escores de propensão, de maneira semelhante à situação hipotética representada na figura 1. Há diversas maneiras de fazê-lo, mas Caliendo e Kopeinig (2008) argumentam que a inspeção visual da distribuição é a maneira mais direta, embora haja métodos mais precisos e sofisticados para realizar a tarefa.

A figura 6 exibe o gráfico de dispersão dos escores de propensão. Ela exibe as distâncias dos escores de propensão de tratados e não tratados, assim como os municípios selecionados para compor a amostra. Nota-se que nenhum município tratado deixou de ser pareado e é também relevante apontar a aparente proximidade dos municípios de controle pareados com os tratados, o que evidencia sucesso no balanceamento da amostra.

É possível também observar a existência de suporte comum na figura 5, a qual exibe a versão empírica do que está representado na figura 1. A figura 5 exibe os histogramas das distribuições dos escores de propensão. No lado esquerdo, encontra-se a distribuição dos escores de propensão para municípios de tratamento e de controle, considerando toda a amostra, antes do pareamento. Conforme esperado, a distribuição para o grupo de controle está concentrada mais perto de 0, o que seria esperado da probabilidade esperada de participação de municípios que não adotaram a política. No lado direito, observa-se o resultado do pareamento, isto é, os histogramas das distribuições dos escores de propensão estimados para tratados e o novo grupo de controle. Comparando a figura superior e a inferior, observa-se que as regiões em que há probabilidade estimada de participação do programa positiva são similares para os dois grupos, o que é indício da validade da hipótese de suporte comum.

Ademais, de acordo com os resultados da tabela 15, localizada no apêndice deste trabalho,

pode-se afirmar que o processo de pareamento foi bem sucedido em aproximar as médias das características observáveis do novo grupo de controle com relação às características do grupo de tratados. Ela indica que os grupos de tratamento e controle não têm mais diferenças significativas nas médias das covariáveis, mesmo aquelas não incluídas no modelo logit. Então, no que diz respeito ao perfil observado desses municípios, os grupos são (quase) idênticos. Nota-se, novamente na figura 5, que a nova distribuição de escores de propensão da amostra (pós pareamento) é muito similar àquela do grupo de tratamento.

6.4 Estimativa de impacto

Em linha com o que foi apresentado no capítulo 4, o impacto sobre as variáveis de resultado foi determinado através de um modelo de efeitos fixos (de tempo e individuais), cuja especificação básica foi:

$$Y_{it} = v_i + \lambda_t + DD \cdot P_{it} + \varepsilon_{it}. \quad (4.21 \text{ revisitada})$$

As características municipais observáveis X_{it} consideradas relevantes para o modelo foram utilizadas no procedimento de pareamento para seleção da amostra e, por isso, estão ausentes na equação acima. A implementação do modelo de efeitos fixos foi realizada com o intuito de reduzir a ocorrência de viés decorrente de eventuais variáveis não observáveis que pudessem afetar a seleção para o tratamento (neste caso, a opção por adotar a política sob avaliação) e não foram incluídas mais covariáveis X_{it} .

Tabela 10 – Resultados antes do pareamento. Aqui, os modelos utilizados foram de efeitos fixos de tempo, incluindo a variável de tratamento como regressor.

	<i>Variável dependente:</i>	
	Salário Médio (1)	Emprego (2)
Tratamento (T_i)	158,31*** (6,901)	2634 (2523,407)
Interação (P_{it})	40,52*** (12,161)	3147 (4446,881)
Observações	667247	667360
R ²	0,001	0,00001
R ² ajustado	0,001	-0,0002

Nota: * p<0,1; ** p<0,05; *** p<0,01

Fonte: Elaboração própria.

Primeiramente, almejou-se estimar o impacto sobre as variáveis de resultado incluindo todos os municípios do estudo, isto é, os municípios que restaram após a eliminação dos *outliers*, desconsiderando o processo de pareamento descrito na seção anterior. Estimar o modelo de efeitos fixos nessas condições, no entanto, implicaria em estimar 4161 parâmetros v_i e 160 parâmetros λ_t e a memória do computador utilizado não foi suficiente para realizar este feito. Optou-se então por estimar a versão do modelo apenas com efeitos fixos de tempo, tornando-se necessário incluir a variável indicadora de tratamento, T_i , cujos resultados obtidos podem ser vistos na tabela 10.

Conforme a discussão teórica apresentada no capítulo 4, a variável de interação P_{it} é uma variável *dummy* que acusa períodos pós tratamento para os municípios tratados e seu coeficiente é a estimativa de impacto relevante. É possível notar que há evidência de impacto médio positivo de R\$150,31 sobre o salário médio mensal. A evidência de impacto sobre o nível de emprego, apesar de positiva, não é estatisticamente significativa.

Todavia, buscou-se, ao longo deste trabalho, apresentar razões pelas quais pode ser precipitado aceitar esses resultados sem maior investigação. Primeiramente, há características observáveis, como indicadores de educação (taxa de analfabetismo e expectativa de anos de estudo), que diferem significativamente entre os dois grupos e, portanto, podem afetar a variável de resultado, gerando ruído na medida de impacto decorrente da política em avaliação. Além disso, a não inclusão de efeitos fixos individuais pode incorrer em viés causado pela seleção de características não observáveis. Em outras palavras, nada é feito para assegurar a hipótese de independência condicional apresentada no capítulo 2, crucial para a aproximação da situação contrafactual ($\mathbb{E}(Y(0)|T = 1)$) a partir do que é observado no grupo de controle ($\mathbb{E}(Y(0)|T = 0)$).

Após a seleção da amostra através do pareamento, utilizou-se o modelo de painel de efeitos fixos de tempo e individuais para estimativa de impacto. Foram encontrados resultados positivos e estatisticamente significativos para o salário médio e o emprego, os quais são apresentados na tabela 11. O impacto médio encontrado no salário médio mensal foi de R\$ 74,67 e no emprego de 1608 vagas, ambos com p-valor inferior a 0,001.

6.5 Discussão

A seção anterior apresenta evidências de que a gratuidade no transporte coletivo tem um impacto positivo sobre os níveis de emprego e salário, resultados corroboram a discussão apresentada no capítulo 5. Naquela ocasião, buscou-se evidenciar a importância do sistema de transporte para a promoção do crescimento econômico e que este, ainda que indiretamente, proporciona aumento do bem-estar de todas as camadas de renda.

Tabela 11 – Estimação do impacto para a amostra

	<i>Variável dependente:</i>	
	Emprego (1)	Salário Médio (2)
Interação	1608*** p < 0,001	74,67*** p < 0,001
Observações	7680	7680
R ²	0,023	0,013
R ² ajustado	-0,004	-0,014
Estatística F (df = 1; 7472)	177,979***	98,422***

Nota: * p<0,1; ** p<0,05; *** p<0,01

Fonte: Elaboração própria.

Sob a ótica do desenvolvimento como liberdade de Sen (2000), os resultados apresentados seriam apenas acessórios ao fato de que o acesso universal ao sistema de transporte, sem tarifa, garante a liberdade de ir e vir das pessoas. Ainda que não trate do transporte de bens e cargas, o acesso à cidade por si só reduz o isolamento que os mais pobres enfrentam no meio urbano. Portanto, além do impacto indireto via crescimento do salário e do emprego, também tem impacto direto sobre o bem-estar dos usuários.

É possível que o impacto no salário seja decorrente de um repasse do valor antes gasto pelas firmas contratantes com vale transporte ao trabalhador ou ainda o reflexo indireto do maior dinamismo econômico trazido pela mobilidade facilitada das pessoas, o que também deve ter contribuído para o aumento relativo no nível de emprego nos municípios em questão. Além disso, o resultado corrobora a afirmação de Gomide (2003) de que o acesso livre à cidade também proporciona acesso, ao menos em teoria, a oportunidades de emprego. Sob a concepção de Sen (2000), o processo de desenvolvimento confere a capacidade de agir sobre seu próprio destino. Pode-se então dizer que, mediante o acesso a novas oportunidades econômicas geograficamente mais distantes, a isenção da tarifa de ônibus promoveu a expansão das liberdades e, portanto, contribuiu ao desenvolvimento econômico dos municípios analisados.

Embora não tenha sido possível segmentar os resultados por faixas de renda, o padrão de mobilidade dos mais pobres indica que a tarifa zero é uma política que os atinge diretamente, uma vez que as pessoas de faixas de renda inferiores utilizam mais o transporte

coletivo do que os indivíduos mais abastados. Nesse sentido, a adoção da gratuidade universal já nasce inclusiva, no sentido de que elimina uma enorme barreira — embora não a única — para a utilização do serviço, a tarifa. Gomide (2003) e Gannon e Liu (1997) fazem menção ao fato de que os extremamente pobres não utilizam sistema de transporte motorizado, conforme ilustrado na figura 4. Com os dados disponíveis, não é possível aferir se as pessoas antes completamente excluídas da utilização do serviço usufruem de sua nova liberdade.

No capítulo 5, relatou-se que o sistema de transporte pode ter impacto direto ou indireto sobre o bem-estar. Este trabalho limitou-se a medi-lo sobre indicadores de impacto indireto, relacionados ao crescimento econômico. O aumento da renda disponível é o principal meio através do qual o indivíduo pode buscar satisfazer suas necessidades. Embora se possa conjecturar sobre os possíveis impactos diretos sobre o bem-estar e sobre a liberdade de ir e vir, à luz do trabalho de Goodman *et al.* (2014), reitera-se que o benefício de acesso às oportunidades econômicas e aos serviços básicos é determinante para a realidade brasileira.

Nesse âmbito, ao pensar em desenvolvimento econômico e inclusão, deve-se considerar que o acesso ao transporte seja pensado em conjunto com o desenvolvimento de outras capacidades humanas (não basta poder se deslocar à entrevista de emprego se o indivíduo não tiver condições de competir pela vaga). Conforme discutido anteriormente, o mercado de trabalho teve papel crucial na redução da pobreza antes do período de recessão e as políticas de transporte coletivo podem ser formuladas em conjunto com outros meios de desenvolvimento de capacidades para inserção no mercado de trabalho, para potencializar os resultados positivos observados no mercado formal.

7 Conclusão

Este trabalho buscou colaborar para a discussão das políticas de transporte no Brasil, em particular a gratuidade universal do transporte coletivo urbano. O principal objetivo do trabalho foi mensurar o impacto da referida política sobre as variáveis de emprego e salário médio dos municípios brasileiros, a partir de dados censitários de 2000 e dados do mercado de trabalho obtidos no CAGED.

Foi apresentado o modelo de resultados potenciais de Neyman-Rubin, ponto de partida para a definição do que se entende por impacto e também para a compreensão de contexto em que emerge o viés de seleção. Os métodos implementados neste trabalho, o pareamento com escore de propensão e o método de diferenças em diferenças (com sua extensão para o modelo de efeitos fixos), despontam para contornar o viés de seleção e mimetizar a situação de seleção aleatória. A validade desses métodos dependem de diversas hipóteses, sobretudo a não interferência das unidades (2.4), independência condicional às variáveis observáveis (3.1), existência de suporte comum (3.2) e a tendência comum (4.2).

O pareamento com escore de propensão permitiu selecionar uma amostra dos municípios não tratados similar ao grupo de tratamento. Em seguida, foi implementado um modelo de painel com efeitos fixos de tempo e individuais para estimar o impacto.

Os resultados apontam que a tarifa zero proporcionou um aumento médio do nível de emprego de 1608 postos de trabalho e um aumento médio no salário mensal de R\$74,67, corroborando a hipótese de que o transporte pode ser um meio de aumentar o bem-estar das pessoas via aumento da renda e das oportunidades de emprego.

Deve-se reiterar que, em virtude do método utilizado, a estimativa de impacto obtida só é válida na região de suporte comum, conforme descrito em (HECKMAN; ICHIMURA; TODD, 1997), e o resultado não pode ser extrapolado para a população inteira.

Para futuros trabalhos, sugere-se (a) uma investigação acerca do impacto distributivo dessa política e se ela de fato atinge as camadas inferiores na distribuição de renda; (b) o levantamento dos custos associados a essa política nos diferentes municípios e também as saídas encontradas em nível local para financiá-la; (c) estudos de caso para identificação de possíveis resultados específicos de cada localidade, como em Goodman *et al.* (2014); e (d) avaliação de impacto sobre outros indicadores de desenvolvimento, como em saúde e educação.

Referências

- ANGRIST, J. D.; PISCHKE, J.-S. *Mostly harmless econometrics: An empiricist's companion*. Princeton, New Jersey: Princeton University Press, 2008.
- ASHENFELTER, O. Estimating the effect of training programs on earnings. *The Review of Economics and Statistics*, p. 47–57, 1978.
- BARROS, R. P. d.; HENRIQUES, R.; MENDONÇA, R. A estabilidade inaceitável: desigualdade e pobreza no brasil. Instituto de Pesquisa Econômica Aplicada (Ipea), 2001.
- BOX-STEFFENSMEIER, J. M.; BRADY, H. E.; COLLIER, D. *The Oxford Handbook of Political Methodology (Oxford Handbooks of Political Science)*. New York, NY: Oxford University Press, 2008.
- BRYSON, A.; DORSETT, R.; PURDON, S. The use of propensity score matching in the evaluation of active labour market policies. 2002.
- CALIENDO, M.; KOPEINIG, S. Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, v. 22, n. 1, p. 31–72, 2008.
- CAMERON, A. C.; TRIVEDI, P. K. *Microeconometrics: methods and applications*. New York, NY: Cambridge university press, 2005.
- CARVALHO, C. H. R. d.; GOMIDE, A.; PEREIRA, R. H. M.; MATION, L. F.; BALBIM, R.; NETO, V. C. L.; GALINDO, E. P.; KRAUSE, C.; GUEDES, E. P. *Tarifação e financiamento do transporte público urbano*. Brasília, 2013.
- COMPANHIA DO METROPOLITANO DE SÃO PAULO. *Síntese das informações de pesquisa domiciliar 2012*. São Paulo, 2013.
- CORDER, G. W.; FOREMAN, D. I. *Nonparametric statistics: A step-by-step approach*. Hoboken, New Jersey: John Wiley & Sons, 2014.
- DAVIDSON, R.; MACKINNON, J. G. *Econometric theory and methods*. New York, NY: Oxford University Press, 2004. v. 5.
- FREEDMAN, D. A. On regression adjustments in experiments with several treatments. *The annals of applied statistics*, p. 176–196, 2008.
- GANNON, C. A.; LIU, Z. *Poverty and transport*. Washington, DC, 1997.
- GOMIDE, A. Á. *Transporte urbano e inclusão social: Elementos para políticas públicas*. Brasília, 2003.
- GOODMAN, A.; JONES, A.; ROBERTS, H.; STEINBACH, R.; GREEN, J. ‘we can all just get on a bus and go’: Rethinking independent mobility in the context of the universal provision of free bus travel to young londoners. *Mobilities*, v. 9, n. 2, p. 275–293, 2014.
- GRIECO, M. *Transport, the poor and moving towards low-carbon societies*. [S.l.], 2013.

- GRIECO, M. Social sustainability and urban mobility: shifting to a socially responsible pro-poor perspective. *Social Responsibility Journal*, v. 11, n. 1, p. 82–97, 2015.
- HECKMAN, J. J.; ICHIMURA, H.; TODD, P. E. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The review of economic studies*, v. 64, n. 4, p. 605–654, 1997.
- HOLLAND, P. W. Statistics and causal inference. *Journal of the American statistical Association*, v. 81, n. 396, p. 945–960, 1986.
- IMBENS, G. W.; WOOLDRIDGE, J. M. Recent developments in the econometrics of program evaluation. *Journal of economic literature*, v. 47, n. 1, p. 5–86, 2009.
- KENNAN, J. The elusive effects of minimum wages. *Journal of Economic Literature*, v. 33, n. 4, p. 1950–1965, 1995.
- KHANDKER, S. R.; KOOLWAL, G. B.; SAMAD, H. A. *Handbook on impact evaluation: quantitative methods and practices*. Washington, DC: World Bank Publications, 2009.
- LECHNER, M. A note on endogenous control variables in causal studies. *Statistics & Probability Letters*, v. 78, n. 2, p. 190–195, 2008.
- LECHNER, M. *et al.* The estimation of causal effects by difference-in-difference methods. *Foundations and Trends in Econometrics*, v. 4, n. 3, p. 165–224, 2011.
- MPL. *Movimento Passe Livre*. 2017. <<http://tarifazero.org/mpl/>>. Acesso em: 14 de junho de 2017.
- NTU. *Transporte público como direito social. E agora?* Brasília, 2016.
- ORLEY, A.; DAVID, C. Using the longitudinal structure of earnings to estimate the effect of training programmes. *Review of Economics and Statistics*, v. 67, n. 4, p. 648–60, 1985.
- PEIXOTO, B.; PINTO, C. X.; LIMA, L.; FOGUEL, M.; BARROS, R. P. D. *Avaliação Econômica de Projetos Sociais*. São Paulo: Dinâmica Gráfica e Editora, 2012.
- PEREIRA, R. H. M.; CARVALHO, C. H. R. d.; SOUZA, P. H. G.; CAMARANO, A. A. Population aging and the effects of free public transport on the fares in the metropolitan region of são paulo. *Revista Brasileira de Estudos de População*, v. 32, n. 1, p. 101–120, 2015.
- ROSENBAUM, P. R.; RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, p. 41–55, 1983.
- SEN, A. *Desenvolvimento como liberdade*. São Paulo: Companhia das letras, 2000.
- SKOUFIAS, E.; NAKAMURA, S.; GUKOVAS, R. *Salvaguardas contra a reversão dos ganhos sociais durante a crise econômica no brasil*. [S.l.], 2017. v. 1, n. 112896.

APÊNDICE A – Estatísticas Descritivas

Tabela 12 – Estatísticas descritivas das variáveis do censo de 2000.

Variável	Min	Max	Média	Mediana
% da renda apropriada pelos 20% mais ricos	38,0	91,4	58,9	58,8
Coefficiente de Gini	0,3	0,9	0,5	0,6
Taxa de atividade	18,4	95,1	63,0	63,0
% empregados com carteira assinada	0,1	78,2	23,4	19,9
% ocupados com ensino fundamental completo	0,6	79,3	28,2	27,0
% ocupados no setor agropecuário	0,1	94,9	42,5	44,1
% ocupados na indústria de transformação	0,1	69,2	9,4	6,7
% ocupados no setor de serviços	3,3	79,1	31,6	30,9
% ocupados com renda de até um salário mínimo	8,4	98,0	64,6	66,4
% domicílios com água encanada	0,0	100,0	66,7	76,4
Esperança de vida ao nascer	57,5	77,2	68,4	69,0
População total	795	10437203	30512	10339
Proporção da população no meio urbano	0,0	1,0	0,6	0,6
Mortalidade infantil	11,2	96,4	33,3	29,4
PEA (18 anos ou mais)	263	5049775	12847	3877
IDHM	0,2	0,8	0,5	0,5
Taxa de fecundidade	1,6	7,8	2,9	2,7
Razão de dependência	39,5	129,5	62,3	59,2
Taxa de envelhecimento	0,6	15,6	6,5	6,4
Taxa de analfabetismo (18 anos ou mais)	1,0	63,0	23,6	19,5
% de adultos com fundamental completo	0,8	70,1	23,7	22,6
Expectativa de anos de estudo	2,3	13,0	8,3	8,4
Taxa de frequência líquida ao ensino básico regular	17,4	100,0	84,5	85,7
Renda <i>per capita</i> exceto renda nula	75,0	1759,8	347,4	319,0
Renda proveniente dos rendimentos do trabalho	28,3	97,0	74,2	75,6
Renda <i>per capita</i> do quinto mais rico	182,5	6064,7	999,2	893,6
% de pobres	0,7	90,8	41,1	38,5
% de vulneráveis à pobreza	5,2	99,0	63,9	65,7
Renda <i>per capita</i> dos pobres	12,9	123,2	75,3	76,6

Fonte: Elaboração própria a partir dos dados do censo de 2000.

Tabela 13 – Teste de Mann-Whitney-Wilcoxon

	Estatística	p-valor
% da renda apropriada pelos 20% mais ricos	17760,00	0,33
Coefficiente de Gini	19324,00	0,52
Taxa de atividade	17743,50	0,32
% empregados com carteira assinada	10507,00	0,01**
% ocupados com ensino fundamental completo	12855,00	0,04**
% ocupados no setor agropecuário	35482,00	0,00***
% ocupados na indústria de transformação	14024,50	0,07*
% ocupados no setor de serviços	10061,00	0,01**
% ocupados com renda de até um salário mínimo	32837,00	0,02**
% domicílios com água encanada	15888,00	0,16
Esperança de vida ao nascer	17621,50	0,31
População total	10797,00	0,01**
Proporção da população no meio urbano	11877,00	0,02**
Mortalidade infantil	30620,50	0,06*
PEA (18 anos ou mais)	10042,50	0,01**
IDHM	14482,50	0,09*
Taxa de fecundidade	28984,50	0,14
Razão de dependência	29573,00	0,11
Taxa de envelhecimento	28283,50	0,18
Taxa de analfabetismo (18 anos ou mais)	30814,00	0,06**
% de adultos com fundamental completo	12796,50	0,04**
Expectativa de anos de estudo	18776,50	0,45**
Taxa de frequência líquida ao ensino básico regular	14167,00	0,08**
Renda <i>per capita</i> exceto renda nula	12630,00	0,03**
Renda proveniente dos rendimentos do trabalho	22382,50	0,97
Renda <i>per capita</i> do quinto mais rico	12097,00	0,03**
% de pobres	30904,50	0,06**
% de vulneráveis à pobreza	30854,50	0,06**
Renda <i>per capita</i> dos pobres	15501,00	0,14

Fonte: Elaboração própria.

Tabela 14 – Amostra resultante do pareamento (grupo de tratados e controle).

UF	Município	Grupo (T_i)
AL	União dos Palmares	0.00
AP	Santana	0.00
BA	Macarani	0.00
BA	Paramirim	0.00
BA	Retirolândia	0.00
CE	Tabuleiro do Norte	0.00
GO	Águas Lindas de Goiás	0.00
GO	São Luís de Montes Belos	0.00
MG	Itabira	0.00
MT	Glória D'Oeste	0.00
MT	União do Sul	0.00
PE	Vitória de Santo Antão	0.00
PI	Esperantina	0.00
PR	Cianorte	0.00
PR	Diamante D'Oeste	0.00
PR	Mangueirinha	0.00
RJ	Angra dos Reis	0.00
RJ	Cordeiro	0.00
RJ	Engenheiro Paulo de Frontin	0.00
RJ	Guapimirim	0.00
RJ	Japeri	0.00
RJ	Petrópolis	0.00
RJ	Rio Claro	0.00
RJ	Tanguá	0.00
RJ	Teresópolis	0.00
RO	Ariquemes	0.00
RO	Espigão D'Oeste	0.00
RO	Machadinho D'Oeste	0.00
RS	Canela	0.00
RS	Capão da Canoa	0.00
RS	Ciríaco	0.00
RS	Pedro Osório	0.00
SC	Camboriú	0.00
SC	Luiz Alves	0.00
SP	Águas de Lindóia	0.00
SP	Aramina	0.00
SP	Ilhabela	0.00
SP	São Sebastião	0.00
SP	Vargem	0.00
SP	Vera Cruz	0.00

Fonte: Elaboração própria.

Tabela 15 – Teste de Mann-Whitney-Wilcoxon após o pareamento.

	Estatística	p-valor
% da renda apropriada pelos 20% mais ricos	147,00	0,73
Coefficiente de Gini	146,00	0,71
Taxa de atividade	159,00	0,99
% empregados com carteira assinada	147,00	0,73
% ocupados com ensino fundamental completo	142,00	0,63
% ocupados no setor agropecuário	161,00	0,99
% ocupados na indústria de transformação	163,50	0,93
% ocupados no setor de serviços	169,50	0,80
% ocupados com renda de até um salário mínimo	149,00	0,78
% domicílios com água encanada	135,00	0,51
Esperança de vida ao nascer	156,50	0,93
População total	156,00	0,92
Proporção da população no meio urbano	150,50	0,80
Mortalidade infantil	170,00	0,79
PEA (18 anos ou mais)	156,00	0,92
IDHM	148,50	0,76
Taxa de fecundidade	191,00	0,40
Razão de dependência	170,00	0,79
Taxa de envelhecimento	171,50	0,76
Taxa de analfabetismo (18 anos ou mais)	170,50	0,78
% de adultos com fundamental completo	144,00	0,67
Expectativa de anos de estudo	146,00	0,71
Taxa de frequência líquida ao ensino básico regular	107,00	0,15
Renda <i>per capita</i> exceto renda nula	155,00	0,90
Renda proveniente dos rendimentos do trabalho	185,50	0,49
Renda <i>per capita</i> do quinto mais rico	160,00	1,00
% de pobres	167,00	0,86
% de vulneráveis à pobreza	162,50	0,96
Renda per capita dos pobres	164,00	0,92

Fonte: Elaboração própria.