



UNIVERSIDADE FEDERAL DE RIO GRANDE
INSTITUTO DE CIÊNCIAS ECONÔMICAS, ADMINISTRATIVAS E CONTÁBEIS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA APLICADA
MESTRADO EM ECONOMIA

ANALFABETISMO ISOLADO E ACESSO A PROGRAMAS SOCIAIS NO BRASIL

FERNANDA DACHI CARRETS

RIO GRANDE – RS

2017

FERNANDA DACHI CARRETS

ANALFABETISMO ISOLADO E ACESSO A PROGRAMAS SOCIAIS NO BRASIL

Dissertação apresentada ao Programa de Pós-Graduação em Economia Aplicada da Universidade Federal de Rio Grande, em cumprimento às exigências para obtenção do título de mestre em Economia Aplicada.

Orientador: Prof. Dr. Gibran da Silva Teixeira

RIO GRANDE – RS

2017

TERMO DE APROVAÇÃO

FERNANDA DACHI CARRETS

ANALFABETISMO ISOLADO E ACESSO A PROGRAMAS SOCIAIS NO BRASIL

Esta dissertação foi apresentada às 8 horas do dia 16 de fevereiro de 2017 como requisito parcial para a obtenção do título de Mestre em Economia Aplicada do Programa de Pós-Graduação em Economia Aplicada da Universidade Federal de Rio Grande. O candidato foi arguido pela Banca Examinadora composta pelos professores abaixo citados. Após deliberação, a Banca Examinadora considerou o trabalho aprovado.

Prof. Dr. Gibran da Silva Teixeira (Orientador)
FURG – PPGE

Prof. Dr. Felipe Garcia Ribeiro
PPGOM – UFPel

Prof. Dr. Gabrielito Rauter Menezes
FURG – PPGE

RIO GRANDE – RS

2017

Resumo

Este trabalho visa investigar se a situação de analfabetismo isolado no Brasil gera uma barreira ao conhecimento do Cadastro Único do Governo Federal. Para tanto, partimos do estudo de Basu e Foster (1998) que desenvolveu o conceito de analfabetismo isolado, pela primeira vez, junto com a teoria de externalidades da alfabetização. Assim sendo, os indivíduos analfabetos se apropriam do acesso às atividades que demandam o uso da escrita e leitura quando têm proximidade de indivíduos alfabetizados. Segundo o Ministério do Desenvolvimento Social e Agrário (MDSA) ainda há uma parcela de potenciais beneficiários de políticas públicas que não é coberta pelas redes de proteção social, e diversos fatores podem contribuir para a criação dessa barreira de acesso, como por exemplo o analfabetismo. Desta forma, a proposta do estudo é checar se o fenômeno de analfabetismo isolado é capaz de reforçar as indesejadas exclusões indevidas de políticas públicas no Brasil. A estratégia empírica empregou dados da Pesquisa Nacional de Amostra de Domicílios (PNAD) para o ano de 2014 combinados com o método de *Propensity Score Matching* e o algoritmo de Imbens (2015) para a seleção das variáveis que irão compor o escore de propensão de cada indivíduo da amostra ser analfabeto isolado. Os resultados obtidos, em geral, indicam que a situação de analfabetismo isolado reduz a probabilidade desses indivíduos conhecerem o Cadastro Único. Visto que os mesmos seriam potenciais beneficiários, os resultados levam a evidência de que a menor informação, ou mesmo a falta de acesso a ela, que a situação de analfabetismo isolado pode impor, afeta esses indivíduos criando um obstáculo para o conhecimento sobre seus direitos e, possivelmente, engrossando as exclusões indevidas de políticas públicas.

Palavra-chave: Analfabetismo Isolado, Cadastro Único, *Propensity Score Matching*.

Abstract

This paper aims to investigate whether the isolated illiteracy situation in Brazil generates a barrier to the knowledge of the Cadastro Único. To do so, we start with the study by Basu and Foster (1998) that developed the concept of illiteracy isolated, for the first time, together with the theory of externalities of literacy. Thus, illiterate individuals appropriate access to activities that require the use of writing and reading when they have proximity to literate individuals. According to the Ministério do Desenvolvimento Social e Agrário (MDSA) there are still a number of potentials beneficiaries of public policies that is not covered by social protection networks and several factors can contribute to the creation of such access barriers, such as illiteracy. Thus, the purpose of the study is to verify if the phenomenon of isolated illiteracy is capable of reinforcing the unwanted undue exclusions of public policies in Brazil. The empirical strategy used data from the Pesquisa Nacional por Amostra de Domicílios (PNAD) for the year 2014 combined with the Propensity Score Matching method and the Imbens (2015) algorithm to select the variables that will compose the propensity score of each individual in the sample to be illiterate. The results obtained, in general, indicate that the isolated illiteracy situation reduces the probability of these individuals to know the Cadastro Único. Since they would be potentials beneficiaries, the results lead to the evidence that less information, or even lack of access to it, that the isolated illiteracy situation can impose, affects these individuals by creating an obstacle to knowledge about their rights and, possibly, widening the undue exclusions of public policies.

Keywords: Isolated Illiteracy, Cadastro Único, Propensity Score Matching.

LISTA DE ABREVIATURAS E SIGLAS

AI	Analfabeto Isolado
AI7	Analfabeto isolado para a definição da amostra com alfabetização a partir dos 7 anos
AI10	Analfabeto isolado para a definição da amostra com alfabetização a partir dos 10 anos
AI15	Analfabeto isolado para a definição da amostra com alfabetização a partir dos 15 anos
ANI	Analfabeto Não Isolado
ATT	Efeito médio do tratamento sobre os tratados
Cadúnico	Cadastro Único
Casen	<i>Encuesta de Caracterización Socioeconómica Nacional</i>
DHS	<i>Demographic Health Survey</i>
ENIGH	<i>Encuesta Nacional de Ingresos y Gastos de los Hogares</i>
ESAM	<i>Enquête Sénégalaise Auprès des Ménages</i>
HIES	<i>Household Income and Expenditure Survey</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
MDSA	Ministério do Desenvolvimento Social e Agrário
NCAER	<i>National Council of Applied Economic Research</i>
PBF	Programa Bolsa Família
PBSM	Plano Brasil Sem Miséria
PNAD	Pesquisa Nacional por Amostra de Domicílios
PNGHS	<i>Papua New Guinea Household Survey</i>
POF	Pesquisa de Orçamentos Familiares
PSM	Propensity Score Matching
PTC	Programas de Transferência de Condicionada
SUF	<i>Subsídio Único Familiar</i>
UNESCO	Organização das Nações Unidas para a Educação, a Ciência e a Cultura

LISTA DE TABELAS

Tabela 1 - Média e desvio-padrão das variáveis para indivíduos analfabetos em todas as definições na amostra, Brasil 2014.....	32
Tabela 2 - Média e desvio-padrão das variáveis para indivíduos analfabetos do sexo feminino para todas as definições na amostra, Brasil 2014.	34
Tabela 3 - Média e desvio-padrão das variáveis para indivíduos analfabetos do sexo masculino para todas as definições na amostra, Brasil 2014.....	35
Tabela 4 - Estimativa do efeito de tratamento médio sobre os tratados: probabilidade de conhecer o Cadastro Único para indivíduos com nível de renda domiciliar per capita até ½ salário mínimo, Brasil 2014.	39
Tabela 5 - Estimativa do efeito de tratamento médio sobre os tratados: probabilidade de conhecer o Cadastro Único para indivíduos do sexo feminino com nível de renda domiciliar per capita até ½ salário mínimo, Brasil 2014.	41
Tabela 6 - Estimativa do efeito de tratamento médio sobre os tratados: probabilidade de conhecer o Cadastro Único para indivíduos do sexo masculino com nível de renda domiciliar per capita até ½ salário mínimo, Brasil 2014	42
Tabela 7 - Análise de sensibilidade dos Limites de Rosenbaum para indivíduos analfabetos, Brasil 2014.....	46
Tabela 8 - Análise de sensibilidade dos Limites de Rosenbaum para indivíduos analfabetos do sexo feminino, Brasil 2014	48
Tabela 9 - Análise de sensibilidade dos Limites de Rosenbaum para indivíduos analfabetos do sexo masculino, Brasil 2014	49

LISTA DE APÊNDICE

APÊNDICE A – Esquema de construção das variáveis de analfabetos isolado e analfabetos não isolados, Brasil 2014.	56
APÊNDICE B - Estimação do propensity score fundamentado nas características observáveis para indivíduos analfabetos AI7, Brasil 2014.	57
APÊNDICE C - Estimação do propensity score fundamentado nas características observáveis para indivíduos analfabetos AI10, Brasil 2014.	58
APÊNDICE D - Estimação do propensity score fundamentado nas características observáveis para indivíduos analfabetos AI15, Brasil 2014.	59
APÊNDICE E - Estimação do propensity score fundamentado nas características observáveis para indivíduos analfabetos AI7 do sexo feminino, Brasil 2014.	60
APÊNDICE F - Estimação do propensity score fundamentado nas características observáveis para indivíduos analfabetos AI10 do sexo feminino, Brasil 2014.	61
APÊNDICE G - Estimação do propensity score fundamentado nas características observáveis para indivíduos analfabetos AI15 do sexo feminino, Brasil 2014.	62
APÊNDICE H - Estimação do propensity score fundamentado nas características observáveis para indivíduos analfabetos AI7 do sexo masculino, Brasil 2014.	63
APÊNDICE I - Estimação do propensity score fundamentado nas características observáveis para indivíduos analfabetos AI10 do sexo masculino, Brasil 2014.	64
APÊNDICE J - Estimação do propensity score fundamentado nas características observáveis para indivíduos analfabetos AI15 do sexo masculino, Brasil 2014.	65
APÊNDICE K - Verificação da hipótese de sobreposição para todas as definições da amostra e métodos de pareamento, Brasil 2014.	66
APÊNDICE L - Verificação da hipótese de sobreposição para todas as definições da amostra para o sexo feminino e métodos de pareamento, Brasil 2014.	68
APÊNDICE M - Verificação da hipótese de sobreposição para todas as definições da amostra para o sexo masculino e métodos de pareamento, Brasil 2014.	70
APÊNDICE N - Estimativa do efeito de tratamento médio sobre os tratados: probabilidade de conhecer o Cadastro Único para indivíduos com nível de renda domiciliar per capita até 1 salário mínimo, Brasil 2014.	72
APÊNDICE O - Estimativa do efeito de tratamento médio sobre os tratados: probabilidade de conhecer o Cadastro Único para indivíduos do sexo feminino com nível de renda domiciliar per capita até 1 salário mínimo, Brasil 2014.	73

APÊNDICE P - Estimativa do efeito de tratamento médio sobre os tratados: probabilidade de conhecer o Cadastro Único para indivíduos do sexo masculino com nível de renda domiciliar per capita até 1 salário mínimo, Brasil 2014.....	74
---	----

SUMÁRIO

1.	INTRODUÇÃO	10
2.	REFERENCIAL TEÓRICO	13
2.1	Externalidades da alfabetização	13
2.2	Cadastro Único do Governo Federal	16
2.3	Focalização e direcionamento de políticas públicas	17
3.	ESTRATÉGIA EMPÍRICA E BASE DE DADOS	22
3.1	<i>Propensity Score Matching</i> (PSM)	22
3.1.1	Análise de sensibilidade (Limites de Rosenbaum)	27
3.2	Descrição dos dados e das variáveis	29
4.	ANÁLISE DOS RESULTADOS	37
4.1	Resultados do efeito médio do tratamento sobre os tratados (ATT)	37
4.2	Resultados da análise de sensibilidade	43
5.	CONSIDERAÇÕES FINAIS	50
	REFERÊNCIAS BIBLIOGRÁFICAS	52
	APÊNDICE	56

1. INTRODUÇÃO

O presente estudo investiga qual a relação do analfabetismo isolado com o acesso à informação sobre a existência do Cadastro Único do Governo Federal. Em última instância, o que se está propondo checar neste estudo é se o fenômeno de analfabetismo isolado é capaz de reforçar as indesejadas exclusões indevidas de políticas públicas. Este fato se daria em função da menor informação, ou mesmo do acesso a ela, visto que residir em um domicílio com apenas indivíduos analfabetos pode impor uma série de restrições ligadas ao acesso de informações e serviços.

O conceito de analfabetismo isolado foi apresentado, pela primeira vez, junto com a teoria de externalidades da alfabetização exposta por Basu e Foster (1998). Segundo eles, indivíduos analfabetos se apropriam do acesso às atividades que demandam o uso da escrita e leitura quando têm proximidade de indivíduos alfabetizados. Em função do conceito proposto pelos autores surgiram os termos Analfabeto Não Isolado (ANI) e Analfabeto Isolado (AI)¹. O primeiro refere-se às pessoas analfabetas que residem em um domicílio com pelo menos uma pessoa alfabetizada, e o segundo representa os analfabetos que residem em um domicílio sem pessoas alfabetizados.

De acordo com o 11º Relatório de Monitoramento Global de Educação para Todos de 2013 da UNESCO que avaliou a situação de 150 países ainda havia 774 milhões de adultos analfabetos no mundo. Entre os anos de 2005 a 2011 o Brasil estava entre os 10 países com a maior população de adultos analfabetos no mundo. Esta informação é corroborada pelos dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) de 2014, visto que a taxa de analfabetismo das pessoas de 15 anos ou mais de idade foi 8,3%, o que representa aproximadamente 13,2 milhões de pessoas.

Segundo o Ministério do Desenvolvimento Social e Agrário (MDSA) um dos grandes desafios para as políticas públicas no Brasil é alcançar a população considerada invisível, ou seja, aquelas pessoas elegíveis que não têm acesso aos serviços públicos e de proteção social. A população definida como invisível pelo MDSA, conhecida na literatura pelo termo "*hard-to-reach*", descreve aqueles subgrupos da população que não são atingidos pelas redes de proteção sociais devido a desvantagens relacionadas a fatores geográficos, sociais e econômicos

¹Os termos analfabetos isolados (AI) e analfabetos não isolados (ANI) são uma livre tradução dos termos *Proximate Literate* e *Isolated illiterate* (BASU e FOSTER, 1998) feita por Ribeiro e Souza (2013).

(FREIMUTH; METTGER, 1990). Sendo assim, as populações de difícil acesso são caracterizadas por serem socialmente invisíveis dado um conjunto de desvantagens, como o analfabetismo, e originando a criação de barreiras para o acesso aos seus direitos (SHAGHAGH; BHOPAL; SHEIKH, 2011).

Após o estudo de Basu e Foster (1998) os trabalhos subsequentes seguiram por duas correntes de pesquisa. Na primeira o intuito foi aprimorar a medida tradicional de alfabetização com a inclusão dos conceitos de externalidades, analfabetismo isolado e analfabetismo não isolado. Já na segunda corrente, na qual o presente estudo está, o objetivo é avaliar empiricamente se, de fato, existem externalidades sociais da alfabetização. Assim, a proposta é estimular o debate a respeito da relevância da alfabetização através da existência de externalidades social da alfabetização. Uma vez que é reconhecido na literatura os retornos privados da alfabetização (FERRER; GREEN; RIDDELL, 2006; De BALDINI ROCHA; PONCZECK, 2011), ainda assim, os benefícios da alfabetização podem ultrapassar os ganhos privados e gerar externalidades para as pessoas analfabetas.

Como ferramenta de análise empírica utilizamos o *Propensity Score Matching* (PSM), o qual é largamente empregado na avaliação do efeito de alguma intervenção nos mais diversos campos de estudo. Examinaremos o efeito do analfabetismo isolado sobre o conhecimento do Cadastro Único para indivíduos com renda *per capita* domiciliar de até $\frac{1}{2}$ salário mínimo. Os dados utilizados são da Pesquisa Nacional por Amostra de Domicílios (PNAD) de 2014, os quais possuem informações sobre o Cadastro Único e os programas sociais do Governo Federal.

Os resultados que obtivemos indicam que o status de analfabetismo isolado, independente da faixa de idade em que foi definida a alfabetização dos indivíduos no domicílio e gênero, atua no sentido de reduzir a probabilidade desses indivíduos conhecerem o Cadastro Único em relação a indivíduos analfabetos não isolados. Para as três definições da amostra empregadas no trabalho, e sem distinção de gênero, os coeficientes variam, em média, de -11,93 p.p. a -13,44 p.p. entre os métodos de pareamento empregados. Uma vez que esses indivíduos se encaixam no critério de renda para elegibilidade do Cadastro Único, e não tem conhecimento do mesmo, possivelmente, podem acabar fazendo parte da população considerada invisível que ainda vive fora de qualquer rede de proteção social dada a situação de analfabetismo isolado.

O presente trabalho está estruturado em mais quatro seções além desta introdução. Na segunda seção, apresentamos a revisão de literatura que sintetiza os estudos que investigam a existência de externalidades da alfabetização e a questão da focalização de políticas públicas no Brasil sobre o ponto da exclusão indevida de indivíduos. Na terceira seção, retratamos os dados e a estratégia metodológica utilizados. Já na quarta seção examinamos os resultados obtidos e, por fim, na quinta seção, apresentamos as considerações finais do estudo com o apanhado dos principais resultados e sugestões para trabalhos futuros.

2. REFERENCIAL TEÓRICO

Um dos questionamentos da literatura mais atual é direcionado ao fato de que os benefícios da alfabetização podem extrapolar os ganhos privados e gerar externalidade positivas aos indivíduos analfabetos que convivem com alfabetizados. Vinculado aos estudos da existência de tal externalidade, a seção ainda traz pesquisas de focalização de políticas públicas justificada pelo grau de exclusões indevidas. Sendo assim, a subseção 2.1 traz trabalhos voltados para os estudos das externalidades da alfabetização, a subseção 2.2 mostra uma breve apresentação do Cadastro Único e por fim, a subseção 2.3 apresenta a revisão de literatura sobre focalização de políticas públicas, e aqui cabe salientar que como o Programa Bolsa Família comportava a maior parte de beneficiários inclusos no Cadastro Único a gama de estudos sobre tal programa é mais frequente na literatura.

2.1 Externalidades da alfabetização

Numa literatura mais recente, pesquisadores passaram a admitir que os benefícios da alfabetização podem exceder os ganhos privados e gerar externalidades positivas para pessoas analfabetas que residem em domicílios com pelo menos uma pessoa alfabetizada, uma vez que a presença de uma pessoa capaz de desempenhar atividades de leitura e escrita proporciona que a alfabetização seja vista como um bem público para os membros analfabetos no domicílio (BASU; FOSTER, 1998).

Estudo como de Basu, Narayan e Ravallion (2001) avaliou, com base na equação minceriana, se o salário de um trabalhador analfabeto é afetado pela alfabetização de outros membros do domicílio em Bangladesh. Foram utilizados dados do *Household Income and Expenditure Survey* (HIES) captados entre os anos de 1995 e 1996. Os autores constataram que os rendimentos dos homens analfabetos são em média 15% maiores se eles fazem parte de um domicílio com pelo menos um dos membros alfabetizados.

Para as mulheres, os autores encontraram que uma trabalhadora analfabeta não isolado do convívio de indivíduos alfabetizados no domicílio ganha cerca do dobro do salário de uma analfabeta que vive em domicílio onde todos os outros membros também são analfabetos. E ainda, as mulheres analfabetas solteiras apresentaram maiores retornos salariais, dado o fato de residir no mesmo domicílio que pessoas

alfabetizadas, do que mulheres analfabetas casadas. Basu, Narayan e Ravallion (2001) também sugerem que os diferentes efeitos da alfabetização sobre o rendimento de homens e mulheres implicam que as mulheres podem ser mais eficientes como receptoras de externalidades de alfabetização no domicílio do que os homens.

Segundo Iversen e Palmer-Jones (2008) existe a possibilidade de que os efeitos encontrados para a externalidade da alfabetização no mercado de trabalho sejam contaminados por problemas de seleção, e características não observáveis tanto dos indivíduos analfabetos quanto dos indivíduos alfabetizados. Para isso, eles utilizaram a equação de salários minceriana e os dados do *Bangladesh Household Income and Expenditure Survey (HIES)* de 1995/6 e 2000/1 para explorar a robustez e discutir alguns resultados encontrados por Basu, Narayan e Ravallion (2001) com relação a seleção no casamento quanto ao nível educacional dos cônjuges. Segundo os autores, a seleção no casamento levanta a possibilidade de que o maior retorno salarial feminino que Basu, Narayan e Ravallion (2001) encontraram possa refletir os atributos não observados de mulheres analfabetas casadas com homens alfabetizados ao invés de compartilhamento de alfabetização domiciliar.

Para o Senegal, Sarr (2004) mediu o efeito do compartilhamento da alfabetização, no ambiente domiciliar, para o retorno salarial no mercado de trabalho. Os resultados foram obtidos através de equação minceriana com base nos dados da *Enquête Sénégalaise Auprès des Ménages (ESAM)* para o ano de 1994. Os autores comprovaram que há efeitos positivos da externalidade da alfabetização direcionados aos ganhos salariais de analfabetos não isolados. Segundo eles, os trabalhadores analfabetos, que vivem em uma família com pelo menos uma pessoa alfabetizada, recebem salários mais altos do que trabalhadores analfabetos que pertencem a uma família em que todos os membros são analfabetos. E além disso, os resultados sugerem que as mulheres são melhores destinatárias e geradoras da externalidade da alfabetização do que os homens.

No campo da saúde infantil, Gibson (2001) utilizou como indicador de saúde a estatura de crianças de 0 a 5 anos de idade de Papua Nova Guiné, e mediu o impacto da externalidade da alfabetização para relação altura-idade das crianças decorrente da presença de membros alfabetizados no domicílio. Segundo os autores, a altura retrata o estado nutricional das crianças, além de ser um reflexo dos hábitos e cuidados dos adultos para com as crianças. Desta forma, a alfabetização dos

indivíduos adultos no domicílio aumenta a capacidade de informação sobre cuidados relacionados a saúde das crianças. Empregando o método de mínimos quadrados ordinários a partir de dados do *Papua New Guinea Household Survey* (PNGHS) para o ano de 1996, os autores, encontraram resultados que revelam um aumento significativo na altura média das crianças que tem convívio com adultos alfabetizados em comparação com crianças que não convivem com pessoas alfabetizadas.

O trabalho de Borroah (2009) empregou um modelo *probit* bivariado para analisar os determinantes do risco de crianças de zonas rurais na Índia sofrerem de má nutrição infantil utilizando dados do *National Council of Applied Economic Research* (NCAER) no ano de 1994. Os resultados, de maneira geral, reportaram a importância da presença de mulheres alfabetizadas no domicílio sobre a redução do risco de má nutrição infantil. Nos domicílios em que a mãe era alfabetizada foram encontrados melhores índices nutricionais entre as crianças do que quando as mães não eram alfabetizadas. No entanto, quando o membro alfabetizado do domicílio era o pai o efeito da externalidade da alfabetização não se mostrou significativo.

Ainda na área de saúde, Husain e Dutta (2012) investigaram os efeitos da externalidade alfabetização entre casais no uso de métodos contraceptivos na Índia. Os autores realizaram as estimações via modelo *logit* com base nos dados levantados pelo *Demographic Health Survey* (DHS) para os anos de 2005 e 2006. O estudo investigou a hipótese de que a adoção de métodos contraceptivos entre mulheres analfabetas com parceiros alfabetizados é maior do que com parceiros analfabetos. Os resultados reportaram a existência do efeito de externalidade da alfabetização relacionada a questão do planejamento familiar de uma pessoa alfabetizada para seu parceiro analfabeto, porém, segundo os autores, o efeito é restrito a alguns grupos específicos (dado características socioculturais) e varia de acordo com o método contraceptivo adotado entre o casal.

Para o Brasil, o estudo feito por Ribeiro *et al.* (2016) utilizou regressões de Poisson e investigou o retorno social da alfabetização, as externalidades, em relação ao tabagismo para analfabetos que residem com indivíduos alfabetizados. Com base nos dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) de 2008, os autores encontram efeitos positivos da alfabetização das mulheres, ou seja, a alfabetização delas serviu como fator de proteção contra o tabagismo para indivíduos analfabetos, de ambos os sexos, residentes no mesmo domicílio.

2.2 Cadastro Único do Governo Federal

O Cadastro Único, também conhecido como CadÚnico, é instrumento de identificação e caracterização socioeconômica das famílias brasileiras de baixa renda utilizado para seleção de beneficiários e integração de programas sociais do Governo Federal. Sendo considerados de baixa renda aqueles indivíduos que possuem renda familiar *per capita* mensal de até ½ salário mínimo ou que possua renda familiar mensal de até três salários mínimos².

Criado em 2001, através do Decreto nº 3.877, de 24 de julho de 2001³, o Cadastro Único só começou a ser utilizado após a unificação dos programas de transferência de renda do Governo Federal em 2003, sendo o Programa Bolsa Família (PBF) o único a utilizar o Cadastro Único até então. Ao longo do tempo diversos programas Federais foram agregados ao CadÚnico, e a criação do Plano Brasil Sem Miséria (PBSM) em 2011 contribui para o aumento no número de programas. (DIREITO *et al*, 2016).

O Cadastro Único vem ampliando suas atividades na participação de processos de seleção de beneficiários e manutenção de políticas públicas direcionadas a população carente. Em setembro de 2016, cerca de 27,7 milhões de famílias estavam inscritas no Cadastro Único, e dessas, aproximadamente 13,5 milhões são beneficiárias do Programa Bolsa Família⁴.

A inscrição no Cadastro Único é um importante passo na construção de uma base de dados para o registro de potenciais beneficiários, e que poderá efetivar o acesso dos mesmo a diversos programas de assistência social no Brasil. Além do destaque da importância do emprego do CadÚnico como base de registro e monitoramento do Plano Brasil Sem Miséria (PSBM)⁵ dado seu poder de abrangência nacional (PAES-SOUSA, 2013).

Segundo Direito *et al*. (2016) uma das estratégias do Plano Brasil Sem Miséria é a Busca Ativa com a pretensão de localizar e incluir no Cadastro Único aqueles ainda

²Definição dada pelo Art. 4º do Decreto nº 6.135, de 26 de junho de 2007. Sendo a renda familiar mensal: a soma dos rendimentos brutos auferidos por todos os membros da família moradores de um mesmo domicílio.

³Revogado pelo Decreto nº 6.135, de 26 de junho de 2007.

⁴ Segundo dados do Ministério do Desenvolvimento Social.

⁵ Conforme Decreto nº 7.492, de 2 de junho de 2011 “o Plano Brasil Sem Miséria, tem a finalidade de superar a situação de extrema pobreza da população em todo o território nacional, por meio da integração e articulação de políticas, programas e ações”.

excluídos da rede de proteção, e assim encaminha-los aos serviços da rede de proteção social. A Busca Ativa trabalha para minimizar os erros de exclusão no cadastro e inserir as famílias de baixa renda que ainda não são cadastradas. Desde a concretização do PBSM, do período de dezembro de 2011 a janeiro de 2016, foram incluídas mais de 7 milhões de novas famílias no Cadastro Único.

Segundo o Ministério do Desenvolvimento Social e Agrário (MDSA), um dos grandes obstáculos do Brasil Sem Miséria é alcançar a população considerada invisível, ou seja, aquela parcela de pessoas que não tem acesso aos serviços públicos e de proteção social. A definição “população invisível” mencionada pelo MDSA é conhecida na literatura como “*hard-to-reach*”. O termo de “difícil acesso” é aplicado a diversos públicos tais como: pessoas com baixo nível socioeconômico, membros de minorias étnicas, pessoas com deficiência, pessoas portadoras de doenças sexualmente transmissíveis, pessoas com baixo nível de instrução e entre outros grupos que tanto o acesso como a obtenção de registros sobre eles são de difícil alcance (FREIMUTH; METTGER, 1990).

Conforme estudo de Shaghagh, Bhopal e Sheikh (2011) o termo “*hard-to-reach*” é utilizado para descrever subgrupos da população que são difíceis de atingir e incluir em programas sociais devido à sua localização ou mesmo devido a situação social e econômica. Sendo assim, as populações de difícil acesso são caracterizadas por serem socialmente invisíveis dado um conjunto de desvantagens, como o analfabetismo, e originando a criação de grandes barreiras para o acesso aos seus direitos.

2.3 Focalização e direcionamento de políticas públicas

Segundo Kerstenetzky (2009) conciliar a diminuição da pobreza e da desigualdade com o desenvolvimento através de programas de transferência de renda é uma das estratégias implantadas por países em desenvolvimento. Os programas de renda garantida dispõem de instrumentos de focalização nos mais pobres e as condicionantes com intuito de aumentar habilidades futuras dos beneficiários

Para a autora há alguns fatores determinantes na ocorrência de indivíduos elegíveis que não recebem benefícios do Programa Bolsa Família (PBF). A falta de acesso a informação, dado que muitas dessas pessoas têm baixo ou nenhum grau de instrução, muitas vezes, faz com que elas nem saibam que tem direito ao benefício.

Em outros casos, a própria pessoa prefere evitar os rótulos de beneficiária e/ou não quer viver na dependência de transferências de renda do governo. Além de problemas de erros na coleta de dados dos beneficiários, e por fim o problema de orçamento restrito para essas políticas.

O emprego de ferramentas de focalização em programas sociais de alívio à pobreza é defendido por uma questão de melhorar a eficiência da alocação dos recursos, o que representa concentrar um orçamento limitado no público-alvo. Portanto, a focalização é vista como um instrumento capaz de ampliar os impactos sobre a população mais pobre fazendo uma melhor distribuição dos recursos disponíveis, e assim tornando o programa mais efetivo (COADY; GROSH; HODDINOTT, 2004).

Tavares *et al.* (2009), a partir de dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) de 2004, avaliaram a situação da focalização do Programa Bolsa Família e qual seu impacto sobre a desigualdade de renda e dos níveis de pobreza no Brasil. Os autores encontraram uma grande disparidade no grau de focalização entre os estados, o que gerou heterogeneidade nos impactos gerados pelo programa. Segundo eles, o sucesso de focalização do programa depende do orçamento disponível e da eficiência nos processos de seleção dos beneficiários. Outro fator que influencia é a magnitude da pobreza em cada estado, assim, estados com população mais pobres tem maior facilidade em identificar o público-alvo do programa apresentando melhores índices de focalização.

Os autores também propõem avaliar se mudanças no processo de seleção dos beneficiários influenciam os indicadores de focalização do programa. Para isso, foram realizadas duas simulações de seleção: a primeira utilizou o mecanismo de seleção de beneficiários com o método de *proxy means-test*, mantendo fixo o orçamento do programa para o ano de 2004; e a segunda simulação definiu uma ampliação do programa de 2004 para 2006, mas mantendo o critério atual de seleção via renda declarada. O resultado das simulações, segundo Tavares *et al.* (2009), sugerem que a combinação do método de seleção via *proxy means-test*, haja vista, a expansão do programa aplicada na segunda simulação poderia obter melhores resultados de focalização.

Soares *et al.* (2010) estudaram a influência do Programa Bolsa Família na redução da desigualdade e da pobreza no Brasil a partir de dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) de 1999 até 2009. Para o estudo sobre a

desigualdade foi utilizada a metodologia de decomposição fatorial do coeficiente de Gini a fim de captar qual a contribuição do programa para a queda na desigualdade. E, segundo os autores, a renda proveniente do Programa Bolsa Família apresentou um efeito um tanto imprevisível, de acordo com os resultados, o benefício do Bolsa Família é responsável por 16% da queda global da desigualdade durante os anos analisados, uma vez que a renda do programa representa menos que 0,8% da renda total das famílias. Os autores afirmam que o principal motivo da renda transferida pelo Bolsa Família contribuir de modo tão determinante para a redução da desigualdade é a focalização nos mais pobres.

Sendo a focalização uma ferramenta importante na diminuição da desigualdade a presença de uma política bem focalizada é essencial para amparar a população mais carente, explicam os autores. Os resultados encontrados indicaram uma tendência de declínio dos erros de inclusão e exclusão⁶ do Programa Bolsa Família durante os anos de 2005, 2007 e 2009, assim como as curvas de incidência, para todos os anos, mostraram uma boa focalização do programa concentrando mais de 80% dos benefícios do Bolsa Família entre os 40% mais pobres. Apesar disso, Soares *et al.* (2010) ressaltam que os erros de exclusão são ainda bastante preocupantes. Ao imaginar uma família que se encontra em situação de extrema vulnerabilidade social, e que por algum motivo não está incluída entre os beneficiários, um erro de exclusão deste porte resulta no desemparo da família. Dessa forma, para os autores, no aperfeiçoamento das ferramentas de focalização, a prioridade deve ser a redução dos erros de exclusão mesmo que, em contrapartida, isso gere o aumento dos erros de inclusão.

Discutidas as questões de desigualdade e focalização, os autores partem para o efeito do Programa Bolsa Família sobre a pobreza. Os resultados reportam que a diferença entre as taxas de pobreza baseadas na renda domiciliar *per capita*, com e sem o benefício, comprovam o efeito atenuador da pobreza resultante do programa.

Barros *et al.* (2010) realizaram um estudo de focalização do Programa Bolsa Família em perspectiva comparada com diversos programas sociais em outros países e no Brasil. Os autores utilizaram curvas de incidência que indicam a porcentagem de beneficiários que pertencem a cada centésimo da distribuição de renda, desta forma,

⁶O erro de inclusão é calculado como a razão entre o número de beneficiários não pobres e o total da população beneficiária. O erro de exclusão corresponde à razão entre os não beneficiários pobres (elegíveis) e o total de pobres (elegíveis).

sendo um programa bem focalizado aquele que tem maior parcela de beneficiários nos centésimos mais pobres da população.

Em comparação internacional, países como México, Chile e Panamá adotaram Programas de Transferência de Condicionada (PTC) semelhantes ao Brasil, porém, com diferentes sistemas de seleção dos beneficiários. Segundo Barros *et al.* (2010) é fundamental fazer uma avaliação dos programas internacionais de PTC para verificar se o sistema de focalização do Bolsa Família está funcionando de forma adequada. Os autores encontraram resultados semelhantes para o desempenho da focalização do PBF e dos programas *Oportunidades* do México e do *Subsídio Único Familiar (SUF)* no Chile⁷. O *Red de Oportunidad* do Panamá apresentou um desempenho melhor que o PBF, o programa panamenho transferiu cerca de 80% dos seus recursos para os 20% mais pobres, já o programa brasileiro alcançou 80% das transferências somente entre os 40% mais pobres. Lembrando que o PTC panamenho utiliza um mecanismo de seleção baseado no método de prova de meios (*proxy means test*) que mostra a probabilidade das famílias cadastradas serem pobres levando em conta variáveis socioeconômicas.

Na comparação do grau de focalização do Programa Bolsa Família (PBF) e de outras políticas públicas do país⁸, os autores encontram resultados que mostram que o PBF estava em melhor situação de focalização que os outros programas no ano de 2003⁹. Para as curvas de incidência obtidas, no PBF cerca de 75% dos beneficiários pertencem aos 40% mais pobres, e para os outros programas, menos de 40% dos beneficiários estavam entre os 40% mais pobres. E por fim, Barros *et al.* (2010) ainda compararam dois cenários possíveis para o Bolsa Família, um com o emprego de ferramentas de focalização e outro com a universalização das transferências. Buscando identificar como seria a renda dos mais pobres caso não fossem postos em prática os instrumentos de focalização; e acabaram confirmando que a focalização

⁷Os dados utilizados pelos autores foram: Pesquisa Nacional por Amostra de Domicílios, Brasil (2008), Encuesta de Hogares, Panamá (2008), Encuesta de Caracterización Socioeconómica Nacional (Casen), Chile (2003) e Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) México (2004).

⁸Os outros programas e políticas públicas utilizados pelos autores foram: abono salarial, Fundo de Garantia de Tempo de Serviço (FGTS), seguro-desemprego, outras indenizações trabalhistas, auxílio alimentação e transporte, auxílio natalidade, maternidade e pensões e aposentadorias públicas.

⁹Como os autores usaram dados da Pesquisa de Orçamentos Familiares (POF) de 2003, avaliaram a focalização das políticas que deram origem ao PBF: Renda Mínima, Programa de Erradicação do Trabalho Infantil, Bolsa Escola, Agente Jovem, Auxílio-Gás e Auxílio Energia Elétrica.

garantiu a maior eficiência nas transferências de renda para os centésimos mais pobres da população.

Sobre a análise do desempenho do Bolsa Família e de outros programas de transferência de renda na América Latina, Soares, Ribas e Osório (2010) comprovaram que há erros de direcionamento nos programas. Com base na Pesquisa Nacional por Amostra de Domicílios (PNAD) de 2004 para o Brasil e na *Encuesta Nacional de Ingresos y Gastos de los Hogares* (ENIGH) de 2004 para o México, o Programa Bolsa Família apresentou um erro de exclusão menor em relação ao programa *Oportunidades* do México, enquanto o programa brasileiro deixava de atingir 59% dos pobres, o mexicano excluía 70% dos pobres. Por outro lado, o erro de inclusão do programa mexicano foi menor do que o brasileiro. Dos beneficiários do *Oportunidades*, 36% faziam parte dos não elegíveis, já para o Bolsa Família esse percentual era de 49% dos beneficiários. Para os autores, os resultados distintos entre os dois programas são causados pelos mecanismos de seleção dos mesmos. Enquanto o *Oportunidades* tem menor inclusão de não elegíveis ao custo de ter uma menor cobertura, o PBF tem maior cobertura, porém tem que sustentar um maior grau de vazamento de não elegíveis que recebem o benefício.

3. ESTRATÉGIA EMPÍRICA E BASE DE DADOS

Para a realização deste estudo, foi necessária a construção da variável de analfabetismo isolado com base no status de alfabetização dentro dos domicílios. Isso só foi possível através da utilização dos dados da PNAD 2014 que também trazem informações sobre o Cadastro Único e os programas sociais do Governo Federal. Além do uso da metodologia de *Propensity Score Matching* (PSM) que é capaz de estimar os efeitos do tratamento sobre o acesso a informação do Cadastro Único por parte dos indivíduos analfabetos. Sendo assim, fez-se o uso da abordagem do PSM, que será descrita na subseção 3.1; seguida pela subseção 3.1.1 que descreve a análise de sensibilidade baseada nos Limites de Rosenbaum. E por sua vez, a subseção 3.2 que apresenta a base de dados e as variáveis empregadas no estudo.

3.1 *Propensity Score Matching* (PSM)

O método indicado por Rosenbaum e Rubim (1983) visa agrupar as características observáveis das unidades em um único escore que possibilite o pareamento entre tratados e controles lidando com o problema da multidimensionalidade. A técnica é amplamente utilizada na literatura de avaliação quando se refere a estimação de efeitos causais de tratamento (ANDRADE *et al.*, 2012; TAVARES, 2016; De SOUZA; WALTENBEG, 2016)

Para a construção de modelos de avaliação que visam identificar o efeito do tratamento é preciso ter informações sobre as características observáveis dos indivíduos analisados, além do status de tratamento e seus resultados potenciais para a variável de interesse. No entanto, para indivíduos tratados não temos posse do resultado potencial caso eles não recebessem o tratamento, e é preciso encontrar uma maneira de supor esse resultado.

Surge assim o problema do contrafactual. Dado que só é possível observar o resultado potencial Y_{1t} do indivíduo tratado, e a informação sobre o resultado potencial do indivíduo tratado caso ele não sofresse intervenção não é observável, Y_{0t} . Torna-se necessária a construção do contrafactual para o grupo de tratamento, denotado como grupo de controle, formado por indivíduos que não sofreram a intervenção, mas apresentam características semelhantes as do grupo de tratamento para obtenção desse resultado potencial.

Para realizar a construção deste contrafactual é necessário a estimação de um escore de propensão, com base em características observáveis. O estimador do escore de propensão é baseado no modelo de regressão logística e mensura a probabilidade de cada indivíduo receber o tratamento ou não. A partir da escolha da função $h(x)$ e a estimação do parâmetro desconhecido γ por máxima verossimilhança. O estimador de *propensity score* segue a equação:

$$\hat{e}(x|W, X) = \frac{\exp(h(x)' \widehat{\gamma}_{ml}(W, X))}{1 + \exp(h(x)' \widehat{\gamma}_{ml}(W, X))} \quad (1)$$

A escolha das variáveis que vão formar a função $h(x)$ é realizada por meio do algoritmo de seleção proposto por Imbens (2015), o qual seleciona uma série de variáveis lineares e quadráticas, e ainda a interação entre as variáveis lineares, via teste de razão de máxima verossimilhança. Com X representando a matriz de características observáveis e W_i o indicador de tratamento. Sendo assim, se a unidade i receber o tratamento $W_i = 1$, caso contrário $W_i = 0$.

O algoritmo oferece uma maior especificação na composição do escore de propensão por permitir a inclusão tanto de variáveis lineares como a interação entre elas, e realiza a escolha das mesmas via teste de razão de máxima verossimilhança e não por arbítrio do pesquisador. Isto posto, foi realizada a escolha das variáveis para construir o escore que estima a probabilidade de participação no tratamento via algoritmo de Imbens e calculado o *propensity score* $\hat{e}(X_i)$.

Segundo Rosenbaum e Rubin (1983), com o *propensity score* já estimado o impacto do tratamento sobre a variável de interesse é obtido através do cálculo do efeito médio do tratamento sobre o tratado, representado pela expressão:

$$ATT = E\{E[Y_{1i}^{AI} | W = 1, \hat{e}(X)] - E[Y_{0i}^{ANI} | W = 0, \hat{e}(X)] | W = 1\} \quad (2)$$

Onde, no nosso caso, o primeiro termo da igualdade é estimado com base nos indivíduos analfabetos isolados e o segundo termo através do resultado médio do grupo de controle de analfabetos não isolados pareados.

Para que as estimações sejam viáveis, a validade do método depende de duas hipóteses: (1) independência condicional ou de seleção nas observáveis, a qual assume que condicional ao vetor de características observáveis o resultado é

independente do tratamento. Sendo a seleção baseada apenas em características observáveis dos indivíduos, admite-se que todas as variáveis que afetam o tratamento e o resultado são observadas na análise, dada por $Y_i \perp W_i | \hat{e}(X_i)$; (2) hipótese de suporte comum que implica que para cada valor de x exista observações de tratados e controle passíveis de comparação, na qual $0 < \hat{e}(W_i = 1 | X_i) < 1$.

Segundo Rosenbaum (2002) a hipótese de independência condicional restringe a análise do PSM, já que infere que todas as características que afetam o tratamento e o resultado são observadas no modelo. No entanto, a presença de fatores não observados pode interferir nos resultados.

Ainda assim, a estimativa do *propensity score* não é suficiente para estimar o ATT, de modo que é uma variável contínua e a probabilidade de encontrarmos dois indivíduos com o mesmo valor de *propensity score* é praticamente zero (BECKER; ICHINO, 2002). Para lidar com esse problema vários métodos foram propostos na literatura como o *Nearest Neighbor Matching*, *Radius Matching*, *Kernel Matching* e *Stratification Matching*.

O pareamento por *Nearest Neighbor Matching* consiste em combinar as observações de tratamento com cada observação de controle utilizando os valores *propensity score* mais próximos. O emprego desse método pode ocorrer com ou sem substituição. Desta forma, uma observação de controle pode ser pareada com mais de uma observação de tratamento. O efeito médio do tratamento sobre os tratados (ATT) é obtido por intermédio das diferenças de resultado entre as observações de tratamento pareadas com as observações de controle.

No método de *Radius Matching* as unidades do grupo de tratamento são pareadas com as unidades do grupo controle somente se o valor do *propensity score* estiver dentro de uma vizinhança já definida de *propensity score* das unidades de tratamento. Cabe ressaltar que quanto menor a vizinhança, maior será a qualidade dos pareamentos.

$$\begin{aligned}
 T^M &= \frac{1}{N^{AI}} \sum_{i \in AI} \left[Y_i^{AI} - \sum_{j \in ANI(i)} w_{ij} Y_j^{ANI} \right] \\
 &= \frac{1}{N^{AI}} \left[\sum_{i \in AI} Y_i^{AI} - \sum_{i \in AI} \sum_{j \in ANI(i)} w_{ij} Y_j^{ANI} \right]
 \end{aligned} \tag{3}$$

$$= \left| \frac{1}{N^{AI}} \sum_{i \in AI} Y_i^{AI} - \frac{1}{N^{AI}} \sum_{j \in ANI(i)} w_{ij} Y_j^{ANI} \right|$$

Onde N^{AI} é a amostra constituída pelos indivíduos pertencentes ao grupo de tratamento, Y_i^{AI} e Y_j^{ANI} são os resultados observados das unidades tratadas e controles, respectivamente. E $ANI(i)$ representa o conjunto de observações do grupo de controle pareadas com as observações de tratamento i a partir do valor estimado do *propensity score*. Sendo assim, o número de controles pareados à observação $i \in AI$ é definida por N_i^{ANI} e denotados os pesos $w_{ij} = \frac{1}{N_i^{ANI}}$ se $j \in ANI(i)$ ¹⁰ e $w_{ij} = 0$, caso contrário.

O *Stratification Matching* (SM) consiste no método de dividir a variação do *propensity score* em intervalos, de forma que em cada um destes intervalos os indivíduos que pertencem ao grupo de tratamento e controle possuam na média o mesmo *propensity score*. Assim, dentro de cada intervalo em que existam indivíduos dos grupos de tratamento e de controle, calcula-se a diferença no resultado médio entre os dois grupos. O ATT é obtido como uma média do efeito médio do tratamento sobre os tratados de cada bloco com os pesos sendo atribuídos pela distribuição das unidades tratadas entre os blocos.

$$T_q^S = \frac{\sum_{i \in I(q)} Y_i^{AI}}{N_q^{AI}} - \frac{\sum_{j \in I(q)} Y_j^{ANI}}{N_q^{ANI}} \quad (4)$$

Onde $I(q)$ é tido como o conjunto de unidades no bloco q , N_q^{AI} e N_q^{ANI} correspondem ao número de observações do grupo tratados e de controle no bloco q . Enquanto Y_i^{AI} e Y_j^{ANI} são os resultados observados para as unidades de tratamento e controle, respectivamente.

Com o método de *Kernel Matching* todas as unidades de tratamento são pareadas com média ponderada de todas as unidades de controle, com pesos que

¹⁰Conjunto de $ANI(i) = \min_j \|p_i - p_j\|$ para pareamento por *Nearest Neighbor Matching* e $ANI(i) = \{p_j \mid \|p_i - p_j\| \leq r\}$ para o pareamento por *Radius Matching*.

são inversamente proporcionais à distância entre o *propensity score* dos indivíduos de tratamento e de controle.

$$T^K = \frac{1}{N^{AI}} \sum_{i \in AI} \left[Y_i^{AI} - \frac{\sum_{j \in ANI(i)} Y_j^{ANI} G\left(\frac{p_j - p_i}{h_n}\right)}{\sum_{i \in ANI} G\left(\frac{p_k - p_i}{h_n}\right)} \right] \quad (5)$$

Onde $G(\cdot)$ é uma função *kernel* e h_n é o parâmetro de largura da banda *bandwidth*.

Cabe ressaltar que independentemente do método de pareamento escolhido, a estimativa do efeito médio do tratamento sobre os tratados sustenta-se na hipótese de suporte comum para todas as observações. Quando a hipótese de suporte comum não for satisfeita para o grupo de tratamento, estes indivíduos serão descartados e o ATT será estimado apenas para aqueles que possuam $0 < \hat{e}(W_i = 1|X_i) < 1$. A partir dos valores obtidos de escore de propensão, procedeu-se, nesse estudo, a estimação do efeito médio do tratamento pelos métodos de pareamento mais indicados na literatura: três vizinhos mais próximo com reposição, *Radius Matching* e *Kernel Matching* (BECKER; ICHINO, 2002; CAMERON; TRIVEDI, 2005, p. 874-878).

Cabe destacar, que uma restrição do método surge do fato de serem empregadas amostragens complexas para estimação, como a PNAD, e o modelo não prevê a inclusão do peso amostral e nem a identificação do desenho amostral complexo. Recentemente alguns estudos se dedicaram a investigação do uso da metodologia com amostras complexas, mas ainda não atingiram resultados concretos sobre a melhor forma de aplicação.

O trabalho mais recente é de Austin, Jeembere e Chiu (2016), em que procedeu diversas simulações de Monte Carlo para abordar duas questões no emprego de *Propensity Score Matching* com amostra complexa. A primeira questão abordada foi da construção do modelo de *propensity score*. Foram consideradas diferentes estimações, dependendo se o modelo de regressão foi ponderado ou não, e se os pesos amostrais foram incluídos como uma covariável adicional. Nenhum dos diferentes modelos de construção do *propensity score* alcançaram um melhor equilíbrio, desta forma os resultados foram inconsistentes quanto a qual especificação do modelo resultou em estimativas com menor viés. E a segunda questão é se os indivíduos do grupo de controle deveriam herdar o peso de amostral do indivíduo

tratado a quem eles foram pareados ou se deveriam receber o seu peso de amostragem natural. Os autores verificaram que o uso de pesos naturais tende a resultar em estimativas com menor viés em comparação com o uso de pesos herdados. Dugoff, Schuler e Stuart (2014) enfatizaram que para atingir resultados a nível populacional para o método de *Propensity Score Matching* é aconselhável considerar o desenho amostral complexo. Segundo os autores, se não há pretensões de generalizar os resultados para nível populacional não há obrigatoriedade no uso pesos amostrais.

Como forma de superar os problemas do emprego de amostras complexas, visto que ainda não há evidências concretas sobre a melhor forma de uso de pesos amostrais para o método de PSM, as estimativas do ATT foram realizadas por meio do método de *bootstrap* com 50 replicações, com o propósito de reduzir o viés no cálculo do erro-padrão do parâmetro de interesse.

Enfim, realizadas todas etapas mencionadas anteriormente na seção e salvo a hipótese de independência imposta pelo modelo, é plausível admitirmos que podem existir fatores não observáveis que interfiram na probabilidade de ser analfabeto isolado, tal como interfiram na probabilidade do indivíduo conhecer o Cadastro Único. Se supormos que a falta de estímulo e motivação para aprender a ler e escrever dos indivíduos, dentro do domicílio, os tornam analfabetos isolados, mas da mesma forma podem impactar de forma positiva a procura de assistência social do governo. Ou seja, o fato do indivíduo ser analfabeto isolado, e pertencer a classes de baixa renda, pode aumentar as chances desse indivíduo procurar o suporte nos benefícios de políticas públicas através do Cadastro Único. Para identificarmos a existência de viés de variável não observada e o quanto ela influencia no tratamento fazemos uso da análise de limites proposta por Rosenbaum (2002) que será apresentada na seção a seguir.

3.1.1 Análise de sensibilidade (Limites de Rosenbaum)

Os estudos que avaliam o efeito de alguma intervenção com estimadores de correspondência partem da hipótese de independência condicional. Se os grupos de tratados e controles diferem devido a fatores não observados, e desta forma, as variáveis não observadas podem afetar simultaneamente o tratamento e variável de resultado, a especificação do modelo de pareamento pode gerar um viés na estimação

do efeito médio do tratamento sobre os tratados. Uma estratégia para retratar este problema é através da abordagem proposta por Rosenbaum (2002). A análise de sensibilidade, os limites de Rosenbaum, não elimina a possibilidade de que fatores não observados existam, mas estabelece o quão prejudicial eles são para os resultados, a fim de contaminar as inferências de uma análise de correspondência.

A verificação da sensibilidade proposta por Rosenbaum (2002) foi aplicada em Aakvik (2001), DiPrete e Gangl (2004) e Becker e Caliendo (2007) para analisar se a inferência sobre os efeitos do tratamento pode ser alterada por fatores não observados. Desta forma, propõe uma verificação da robustez do modelo através da hipótese de viés de seleção. O método identifica dois indivíduos a partir de suas características observáveis, denominados i e j , dentro de uma distribuição logística. Para isso, é assumido que a probabilidade participação no tratamento é:

$$\tau_i = \hat{e}(W = 1|X_i = x) = F(\beta X_i + \gamma \mu_i) \quad (6)$$

Onde x_i é o vetor que contém as características observadas do indivíduo i , μ_i é o termo de erro estocástico, e γ mede o efeito de μ_i sobre a probabilidade do indivíduo ser analfabeto isolado. Evidentemente, se não houver nenhum viés de variável omitida, μ_i será zero e a probabilidade de ser tratado será determinada apenas pelo vetor de características observadas X_i . Caso contrário, constatado o viés de seleção devido a interferência de variáveis não observáveis na análise, dois indivíduos com características observáveis semelhantes terão diferentes probabilidades de receber o tratamento. Assumindo que $F(\cdot)$ é uma função com distribuição logística, a probabilidade de que o indivíduo i tem de ser tratado pode ser escrito:

$$\left(\frac{\tau_i}{1 - \tau_i} \right) = \exp(\beta X_i + \gamma \mu_i) \quad (7)$$

Logo, se fizermos a comparação de dois indivíduos pareados, i e j , que apresentam as mesmas características observáveis ($X_i = X_j$), a razão de probabilidade de receber o tratamento pode ser escrita como:

$$\frac{\frac{\tau_i}{1 - \tau_i}}{\frac{\tau_j}{1 - \tau_j}} = \frac{\tau_i(1 - \tau_j)}{\tau_j(1 - \tau_i)} = \frac{\exp(\beta X_j + \gamma \mu_j)}{\exp(\beta X_i + \gamma \mu_i)} = \exp\gamma[(\mu_i - \mu_j)] \quad (8)$$

Se confirmada a presença de viés de seleção, os dois indivíduos com características observáveis semelhantes têm diferentes chances de serem submetidos ao tratamento, pois o efeito de μ é captado por γ que será diferente de zero. Não havendo diferenças relacionadas as características não observadas, nenhum viés é encontrado ($\gamma=0$) e a razão de chances dos indivíduos pareados receberem o tratamento é 1. Em resumo, a análise de sensibilidade avalia o quanto que o efeito médio do tratamento é modificado por alterações nos valores de γ e de $\mu_i - \mu_j$. Os limites da razão de probabilidades de participação no tratamento propostos por Rosenbaum (2002) são expressos pela desigualdade:

$$\frac{1}{\Gamma} \leq \frac{\tau_i(1 - \tau_j)}{\tau_j(1 - \tau_i)} \leq \Gamma \quad (9)$$

Para os indivíduos pareados que possuem a mesma probabilidade de participar do tratamento o valor de Γ será igual a um. Caso contrário, se por exemplo $\Gamma = 2$, os indivíduos que supostamente parecem ser semelhantes em termos de características observáveis podem diferir em suas chances de receber o tratamento por um fator de 2. Os limites de Rosenbaum indicam o quanto alterações nos valores de Γ modificam a inferência sobre do efeito do tratamento.

Então, com a finalidade de verificar a gravidade do viés de fatores omitidos, foi sugerido por Rosenbaum (2002) e Caliendo e Kopeinig (2008) a mensuração da sensibilidade das estimativas do efeito causal do tratamento em relação às características não-observadas por meio dos Limites de Rosenbaum. Os limites indicam o quanto alterações dos fatores não observáveis podem afetar a robustez dos resultados. A estatística de *Mantel-Haenszel* é a mais indicada para avaliar a hipótese do viés do efeito do tratamento, uma vez que a variável de resultado é binária. A estatística testa a hipótese nula de ausência de efeito do tratamento nas estimações (BECKER; CALIENDO, 2007; AAKVIK, 2001).

3.2 Descrição dos dados e das variáveis

Os dados utilizados para a estimação são da Pesquisa Nacional por Amostra de Domicílios (PNAD) do ano de 2014, disponibilizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE). Além do suplemento da PNAD 2014 de Acesso ao

Cadastro Único para Programas Sociais do Governo Federal e a Programas de Inclusão Produtiva que possibilitou a identificação dos indivíduos que conhecem o Cadastro Único de Programas Sociais do Governo Federal através da pergunta: “Algum morador já ouviu falar sobre o Cadastro Único de Programas Sociais do Governo Federal? ”.

Para encontrarmos os indivíduos analfabetos isolados, primeiramente, identificamos cada domicílio da amostra com seus respectivos moradores, e realizamos a classificação desses indivíduos como alfabetizados ou não alfabetizados criando a variável da status de alfabetização dos indivíduos no domicílio a partir de determinada faixa de idade. A definição do status de alfabetização dos indivíduos nos domicílios foi construída com base na pergunta “sabe ler e escrever” que compõem a PNAD. Quando a resposta era “sim” o indivíduo foi caracterizado como alfabetizado, e quando era “não” os indivíduos eram considerados analfabetos.

Foram construídas três variáveis de alfabetização, a primeira definiu apenas o status de alfabetização dos indivíduos maiores 7 de anos, a segunda definiu apenas a alfabetização de indivíduos maiores de 10 anos, e a terceira, considerou apenas a alfabetização de indivíduos maiores de 15 anos. Por exemplo, na variável de alfabetização que considerou apenas a alfabetização de indivíduos maiores de 10 anos, definindo o status de alfabetização de todos os moradores do domicílio maiores de 10 anos, atribuiu-se o valor um para os indivíduos alfabetizados maiores de 10 anos e zero para os demais indivíduos dentro do domicílio.

Desta forma, através da variável de alfabetização, foi possível identificar aqueles domicílios em que para todos os moradores a variável alfabetização é igual a zero e, assim, especificar cada indivíduo desse domicílio como analfabeto isolado (tratados). Para aqueles domicílios em que pelo menos um morador foi identificado como alfabetizado, com base na idade especificada na definição da amostra, os indivíduos analfabetos foram classificados como não isolados (controles).

A variável dependente utilizada na análise empírica é binária, assume valor um para os indivíduos que conhecem o Cadastro Único e zero caso contrário. Além da variável sobre o conhecimento por parte de algum morador do domicílio da existência

do Cadastro Único do Governo Federal, as covariadas utilizadas foram divididas em três grupos¹¹:

- Atributos do indivíduo: variável de idade, *dummy* para raça, *dummy* de gênero e *dummy* para estado civil.
- Atributos do domicílio: número de moradores menores de 9 anos, número de moradores maiores de 9 anos, número de quartos do domicílio, *dummy* para existência de banheiro no domicílio.
- Acesso à infraestrutura básica e localização do domicílio: *dummy* para identificar o acesso a rede de esgoto no domicílio e variável *dummy* para indicar a zona de localização do domicílio

Especificadas as variáveis empregadas, procedemos com as estimações a nível de Brasil, separando inicialmente pela caracterização da amostra dada a definição do status de alfabetização a partir de faixa de idade nos domicílios. A primeira definição da amostra corresponde ao status de alfabetização apenas dos indivíduos maiores 7 de anos, a segunda a partir de 10 anos e a terceira a partir de 15 anos que identificamos como AI7, AI10 e AI15, respectivamente, e além de separar a amostra por gêneros.

Quanto a especificação do nível de renda, optou-se por apresentar os resultados das estimações que consideram a amostra formada por indivíduos com renda *per capita* domiciliar de até $\frac{1}{2}$ salário mínimo¹² já que são esses indivíduos que atendem ao critério de renda para elegibilidade do Cadastro Único. Foram também realizadas estimações para o nível de renda de até 1 salário mínimo domiciliar *per capita* como forma de robustez para a análise, uma vez que ainda há uma parcela de indivíduos que são incluídos nas políticas públicas mesmo apresentando renda superior ao critério de elegibilidade (BARROS *et al.*, 2010; SOARES; RIBAS; OSÓRIO, 2010; SOARES *et al.*, 2010; SOUZA *et al.*, 2013).

¹¹ As covariadas utilizadas foram escolhidas com base em estudos voltados para avaliação de políticas públicas, em especial do Programa Bolsa Família. Sendo assim, acreditamos que essas variáveis possam captar, de certa forma, as características de potenciais elegíveis aos programas sociais do Governo Federal (DE ARAÚJO; GOMES; LIMA, 2016; DE SOUZA; WALTENBEG, 2016).

¹²O salário mínimo em 2014 era R\$ 724,00, logo a renda analisada nesse estudo compreende a faixa de renda domiciliar per capita de R\$ 362,00, ou seja, $\frac{1}{2}$ salário mínimo.

Para que fosse possível avaliar o impacto do analfabetismo isolado para aqueles indivíduos que efetivamente são o público-alvo do programa, o critério de renda do Cadastro Único foi respeitado nas estimações. Acredita-se que ao construir uma amostra com indivíduos que de fato atendem ao critério de renda para participação do Cadastro Único, e que desta forma seriam potenciais beneficiários das políticas públicas que compõem o programa, seria estruturado um grupo mais equilibrado em relação as características observáveis.

Na Tabela 1 são apresentadas as médias e desvios-padrão de todas as variáveis utilizadas separando por analfabetos isolados e analfabetos não isolados para todas as definições de amostra empregadas no estudo. O número de observações na amostra da PNAD 2014 para indivíduos analfabetos, após a exclusão daquelas observações que possuíam alguma informação perdida para ao menos uma das variáveis, é de 22971 observações para a faixa de renda *per capita* domiciliar de $\frac{1}{2}$ salário mínimo.

Do total de analfabetos, foram encontradas 1368 observações para analfabetos isolados quando a situação de alfabetização considerada no domicílio é só a partir dos 7 anos. Para domicílios em que a condição de alfabetização foi definida apenas para maiores de 10 anos, encontramos que 1510 observações são de analfabetos isolados. E para a amostra que contabilizou a alfabetização apenas de maiores de 15 anos no domicílio o número de analfabetos isolados é de 2177 observações.

Tabela 1 - Média e desvio-padrão das variáveis para indivíduos analfabetos em todas as definições na amostra, Brasil 2014.

Variável	AI7		AI10		AI15	
	Analfabetos isolados (Tratado)	Analfabetos não isolados (Controle)	Analfabetos isolados (Tratado)	Analfabetos não isolados (Controle)	Analfabetos isolados (Tratado)	Analfabetos não isolados (Controle)
Variável de resultado						
Cadastro Único (conhecer=1)	0,5054 (0,5001)	0,6464 (0,4780)	0,5147 (0,4999)	0,6467 (0,4779)	0,5246 (0,4995)	0,6499 (0,4770)
Variáveis para construção do propensity score						
Características do indivíduo						
Idade do indivíduo	41,0564 (21,4867)	16,3412 (21,6861)	40,1606 (21,5925)	16,2416 (21,6579)	38,2288 (21,8612)	15,6776 (21,4146)
Dummy para raça (branco=1)	0,1988 (0,3992)	0,2767 (0,4473)	0,1973 (0,3981)	0,2773 (0,4476)	0,1969 (0,3977)	0,2799 (0,4489)
Dummy para gênero (homem=1)	0,5408 (0,4985)	0,5188 (0,4996)	0,5363 (0,4988)	0,5189 (0,4996)	0,5054 (0,5000)	0,5216 (0,4995)

Dummy de casado (casado=1)	0,4295 (0,4951)	0,2073 (0,4053)	0,4367 (0,4961)	0,2053 (0,4039)	0,4430 (0,4968)	0,1972 (0,3979)
Atributos do domicílio						
Número de quartos	1,4524 (0,6348)	2,0919 (0,8129)	1,4847 (0,6508)	2,0939 (0,8131)	1,7046 (0,7588)	2,0904 (0,8148)
Dummy para possuir banheiro (banheiro=1)	0,7816 (0,4132)	0,9011 (0,2984)	0,7836 (0,4119)	0,9017 (0,2975)	0,7980 (0,4015)	0,9040 (0,2945)
Número de pessoas maiores de 9 anos	2,7650 (1,6275)	4,9469 (1,9592)	2,9140 (1,6730)	4,9508 (1,9618)	3,4507 (1,8790)	4,9596 (1,9677)
Número de pessoas menores de 9 anos	0,7151 (1,1525)	1,7049 (1,2176)	0,8701 (1,2523)	1,7005 (1,2167)	1,0127 (1,3079)	1,7121 (1,2095)
Acesso à infraestrutura básica e localização do domicílio						
Dummy para zona urbana (urbana=1)	0,5278 (0,4994)	0,6878 (0,4634)	0,5383 (0,4986)	0,6881 (0,4632)	0,5273 (0,4993)	0,6940 (0,4608)
Dummy para possuir esgoto (esgoto=1)	0,1908 (0,39314)	0,3237 (0,4679)	0,1986 (0,3991)	0,3241 (0,4680)	0,1950 (0,3963)	0,3285 (0,4696)
Observações	1368	21603	1510	21461	2177	20794

Fonte: Elaboração própria a partir de dados da PNAD 2014. Nota: desvio-padrão entre parênteses.

Na tabela 2, encontram-se as médias e os desvios-padrão de todas as variáveis analisadas para indivíduos analfabetos do sexo feminino de todas as definições da amostra, e separadas por analfabetos isolados e analfabetos não isolados. O número de observações da PNAD de 2014 para a amostra de indivíduos analfabetos do sexo feminino, após a exclusão daquelas unidades que tinham alguma informação inexistente para ao menos uma das variáveis, é de 11022 observações para a faixa de renda per capita domiciliar de $\frac{1}{2}$ salário mínimo.

Do total de analfabetos do sexo feminino, foram encontradas 628 observações na situação de analfabetismo isolada quando a situação de alfabetização considerada no domicílio é só a partir dos 7 anos. Para domicílios que a condição de alfabetização foi definida apenas para maiores de 10 anos, encontramos que 700 observações são de analfabetos isolados. E para a amostra que contabilizou a alfabetização apenas de maiores de 15 anos no domicílio o número de analfabetos isolados é de 1077 observações.

Tabela 2 - Média e desvio-padrão das variáveis para indivíduos analfabetos do sexo feminino para todas as definições na amostra, Brasil 2014.

Variável	AI7		AI10		AI15	
	Analfabetos isolados (Tratado)	Analfabetos não isolados (Controle)	Analfabetos isolados (Tratado)	Analfabetos não isolados (Controle)	Analfabetos isolados (Tratado)	Analfabetos não isolados (Controle)
Variável de resultado						
Cadastro Único (conhecer=1)	0,5007 (0,5003)	0,6434 (0,4790)	0,5063 (0,5003)	0,6440 (0,4788)	0,5225 (0,4997)	0,6474 (0,4777)
Variáveis para construção do <i>propensity score</i>						
Características do indivíduo						
Idade do indivíduo	42,8960 (21,4101)	16,2341 (22,1368)	41,8882 (21,4297)	16,1175 (22,1098)	39,5889 (21,6167)	15,3912 (21,8108)
Dummy para raça (branco=1)	0,2251 (0,4180)	0,2872 (0,4524)	0,2263 (0,4187)	0,2875 (0,4526)	0,2230 (0,4164)	0,2902 (0,4539)
Dummy de casado (casado=1)	0,4771 (0,4998)	0,1693 (0,3750)	0,4809 (0,4999)	0,1669 (0,3729)	0,4552 (0,4982)	0,1578 (0,3646)
Atributos do domicílio						
Número de quartos	1,4426 (0,6198)	2,0950 (0,8093)	1,4728 (0,63119)	2,0975 (0,8097)	1,7130 (0,7506)	2,0952 (,8117)
Dummy para possuir banheiro (banheiro=1)	0,8173 (0,3867)	0,9077 (0,2894)	0,8161 (0,3876)	0,9084 (0,2884)	0,8304 (0,3754)	0,9103 (0,2856)
Número de pessoas maiores de 9 anos	2,7543 (1,5152)	4,9369 (1,9553)	2,9066 (1,5734)	4,9417 (1,957)	3,4516 (1,7980)	4,9596 (1,9629)
Número de pessoas menores de 9 anos	0,73858 (1,1528)	1,7178 (1,2101)	0,8981 (1,2474)	1,7137 (1,2094)	1,0211 (1,2758)	1,7312 (1,2025)
Acesso à infraestrutura básica e localização do domicílio						
Dummy para zona (urbana=1)	0,5889 (0,4924)	0,7094 (0,4540)	0,5940 (0,4914)	0,7098 (0,4538)	0,5778 (0,49412)	0,71604 (0,4509)
Dummy para possuir esgoto (esgoto=1)	0,2362 (0,4250)	0,3396 (0,4736)	0,2418 (0,4285)	0,3400 (0,4737)	0,2313 (0,4218)	0,34489 (0,4753)
Observações	628	10394	700	10322	1077	9945

Fonte: Elaboração própria a partir de dados da PNAD 2014. Nota: desvio-padrão entre parênteses.

Na tabela 3, reportamos as médias e os desvios-padrão das variáveis para os indivíduos analfabetos do sexo masculino de todas as definições de amostra separadas por analfabetos isolados e analfabetos não isolados. O número de observações da amostra da PNAD 2014 para indivíduos analfabetos do sexo masculino, após a exclusão daquelas unidades que tinham alguma informação inexistente para ao menos uma das variáveis, é de 11949 observações para a faixa de renda *per capita* domiciliar de $\frac{1}{2}$ salário mínimo.

Do total de analfabetos do sexo masculino, foram encontradas 740 observações para analfabetos isolados quando a situação de alfabetização considerada no domicílio é só a partir dos 7 anos. Para domicílios que a condição de alfabetização foi definida apenas para maiores de 10 anos, encontramos que 810 observações estão na situação de analfabetismo isolado. E para a amostra que contabilizou a alfabetização apenas de maiores de 15 anos no domicílio o número de analfabetos isolados é de 1100 observações.

Tabela 3 - Média e desvio-padrão das variáveis para indivíduos analfabetos do sexo masculino para todas as definições na amostra, Brasil 2014.

Variável	AI7		AI10		AI15	
	Analfabetos isolados (Tratado)	Analfabetos não isolados (Controle)	Analfabetos isolados (Tratado)	Analfabetos não isolados (Controle)	Analfabetos isolados (Tratado)	Analfabetos não isolados (Controle)
Variável de resultado						
Cadastro Único (conhecer=1)	0,5093 (0,5002)	0,6492 (0,4772)	0,5220 (0,4998)	0,6492 (0,4772)	0,5266 (0,4995)	0,6521 (0,4763)
Variáveis para construção do propensity score						
Características do indivíduo						
Idade do indivíduo	39,4946 (21,4422)	16,4405 (21,2601)	38,6674 (21,6344)	16,3567 (21,2307)	36,8981 (22,0264)	15,9403 (21,0423)
Dummy para raça (branco=1)	0,1764 (0,3814)	0,2669 (0,4423)	0,1723 (0,3779)	0,2678 (0,4428)	0,17132 (0,3769)	0,2704 (,4442)
Dummy de casado (casado=1)	0,3890 (0,4878)	0,2424 (0,4286)	0,3985 (0,4898)	0,2408 (0,4276)	0,4310 (0,4954)	0,2333 (0,4229)
Atributos do domicílio						
Número de quartos	1,4608 (0,6475)	2,0891 (0,8162)	1,4950 (0,6675)	2,0905 (0,8164)	1,6963 (0,7670)	2,0860 (0,8177)
Dummy para possuir banheiro (banheiro=1)	0,7513 (0,4325)	0,8950 (0,3065)	0,7555 (0,4300)	0,8956 (0,3057)	0,7664 (0,4232)	0,8982 (0,3023)
Número de pessoas maiores de 9 anos	2,7740 (1,7180)	4,9562 (1,9628)	2,9205 (1,7555)	4,9591 (1,9657)	3,4499 (1,9558)	4,9597 (1,9722)
Número de pessoas menores de 9 anos	0,6951 (1,1526)	1,6930 (1,2246)	0,84596 (1,2568)	1,6882 (1,2233)	1,0045 (1,3391)	1,6946 (1,2157)
Acesso à infraestrutura básica e localização do domicílio						
Dummy para zona (urbana=1)	0,4759 (0,4997)	0,6677 (0,4710)	0,4902 (0,5002)	0,6679 (0,4709)	0,4779 (0,4997)	0,6739 (0,4687)
Dummy para possuir esgoto (esgoto=1)	0,1524 (0,3596)	0,3090 (0,46210)	0,1613 (0,36809)	0,3093 (0,4622)	0,1596 (0,3664)	0,3134 (0,4639)
Observações	740	11209	810	11139	1100	10849

Fonte: Elaboração própria a partir de dados da PNAD 2014. Nota: desvio-padrão entre parênteses.

Conforme os dados apresentados para as diferentes definições da amostra, com e sem distinção de gênero, pode-se concluir por meio da análise realizada, que os indivíduos analfabetos isolados apresentam, em média, um menor conhecimento do Cadastro Único em relação aos analfabetos não isolados. Cerca de 50% dos indivíduos analfabetos isolados e 35% dos indivíduos analfabetos não isolados não sabem da existência do Cadastro Único. As informações da tabela 1 reportam também as características dos indivíduos da amostra geral, e indicam que os indivíduos analfabetos isolados, em sua maioria, têm em torno de 41 anos, são não brancos, do sexo masculino e não são casados. Decompondo os atributos do domicílio, observa-se que quase 20% das residências de indivíduos analfabetos isolados não possuem banheiro. Constata-se também a precariedade em relação ao serviço de saneamento básico. De acordo com os dados, em média, 19% dos domicílios com indivíduos do grupo de tratamento e 32% dos domicílios com indivíduos do grupo de controle apresentam rede de esgoto. Verifica-se ainda que os indivíduos tratados residem em domicílios, em média, com quantidade inferior de quartos e menor número de pessoas adultas e crianças do que os indivíduos do grupo de controle.

4. ANÁLISE DOS RESULTADOS

Os resultados serão analisados em duas subseções. Na subseção 4.1 serão apresentados os resultados do efeito médio do tratamento sobre os tratados (ATT) de todas as definições da amostra empregadas no trabalho, além das estimações para as amostras com distinção de gênero. E na seção 4.2 realiza-se a análise de sensibilidade para as estimativas do ATT sobre os resultados de interesse, utilizando o método dos Limites de Rosenbaum (2002).

4.1 Resultados do efeito médio do tratamento sobre os tratados (ATT)

Optou-se por apresentar os resultados dos modelos que consideram apenas a amostra formada por indivíduos que tem renda domiciliar *per capita* de até $\frac{1}{2}$ salário mínimo, visto que são esses indivíduos que atendem ao critério de renda para elegibilidade do Cadastro Único do Governo Federal. A estimativa do efeito médio do tratamento sobre os tratados (ATT) para a amostra de indivíduos com renda domiciliar *per capita* de até 1 salário mínimo é indicada, no apêndice, apenas como forma de robustez. Mas cabe ressaltar, que as estimativas do ATT com extrapolação do critério de renda do programa apresentaram coeficientes menores de efeito do tratamento, e nem todos se mostram significativos.

Posto isso, os resultados para o modelo *logit* que calculou a probabilidade de ser analfabeto isolado com renda domiciliar *per capita* de até $\frac{1}{2}$ salário mínimo para o ano de 2014, nas diferentes definições da amostra, encontram-se no apêndice. E com base nessa probabilidade foi realizado o pareamento entre os grupos. Cada indivíduo tratado (analfabeto isolado) foi pareado com um indivíduo do grupo de controle (analfabeto não isolado) que apresentou igual probabilidade de receber o tratamento. Tornando possível a comparação de indivíduos que diferem apenas do fato de ser analfabeto isolado ou não, mas que compartilham de características observáveis muito próximas.

Após a estimação dos resultados do modelo *logit* é relevante verificar o balanceamento entre os grupos de tratamento e controle depois de realizado o pareamento, pois a qualidade desse balanceamento é de extrema importância para a obtenção de estimativas corretas do ATT. Nesse sentido, no apêndice, exibem os gráficos com a sobreposição do *propensity score* dos grupos de tratados e controles

após pareamento para todas as definições de amostra e métodos de pareamento. Em geral, os grupos amostrais analisados apresentaram um comportamento similar em termos de características observáveis após o pareamento para ambos os grupos. E entre os métodos de pareamentos utilizados, destacamos que o pareamento pelos três vizinhos mais próximos foi o que apresentou um melhor equilíbrio entre os grupos em todas as amostras, dado que registrou uma melhor sobreposição entre tratados e controles.

Feitas as estimativas do *propensity score* pelo modelo *logit* e verificado o balanceamento entre os grupos de tratamento e controle, procedeu-se à estimação do efeito médio do tratamento sobre os tratados pelos três métodos de pareamento considerados, cujos resultados serão apresentados.

Na tabela 4 apresentamos as estimações do ATT para o Brasil separando os resultados por faixa etária de definição da alfabetização no domicílio. Os coeficientes mostram que para observações que contabilizam a condição de alfabetização no domicílio a partir dos 7 anos de idade, conforme os três métodos de pareamento, há evidências de que, em média, os indivíduos analfabetos isolados têm probabilidade menor de ter acesso a informação da existência do Cadastro Único em relação aos seus pares do grupo de controle na amostra. Esses resultados são estatisticamente significativos e mostram que a diferença entre o acesso a informação sobre o programa entre analfabetos isolados e analfabetos não isolados varia de -12,42 pontos percentuais para o pareamento por raios, a -13,17 pontos percentuais, de acordo com o pareamento pelos três vizinhos mais próximos.

Os resultados para a amostra em que se estabeleceu a alfabetização a partir de 10 anos no domicílio, com base nos três métodos de pareamento, indicam que, em média, os indivíduos analfabetos isolados têm probabilidade menor de conhecer o Cadastro Único em relação aos indivíduos que residem com pelo menos uma pessoa alfabetizada no domicílio. Os coeficientes foram estatisticamente significativos e mostram o efeito médio do tratamento sobre os tratados variando de -12,71 pontos percentuais pelo pareamento por raios, a -13,44 pontos percentuais conforme o método de pareamento pelos três vizinhos mais próximos.

E ainda na tabela 4, os resultados para as observações em que se ponderou a alfabetização no domicílio a partir dos 15 anos também apresentaram efeitos negativos em relação ao acesso a informação de analfabetos isolados. De acordo com os métodos de pareamento empregados na análise, em média, indivíduos analfabetos

isolados têm menor probabilidade de acesso a informação sobre o Cadastro Único em relação a indivíduos analfabetos não isolados da amostra. Esses resultados são estatisticamente significativos e mostram que a diferença entre tratados e controles varia de -11,93 pontos percentuais com pareamento por raios, a -12,00 pontos percentuais através do pareamento pelos três vizinhos mais próximos.

De forma geral, os resultados apresentados, independente da faixa etária em que foi estabelecida a alfabetização para o domicílio e método de pareamento empregado, mostram que o efeito médio do tratamento sobre os tratados (ATT) apresentou coeficientes negativos e estatisticamente significativos para o acesso a informação sobre o Cadastro Único de analfabetos isolados. Desta forma, indivíduos analfabetos isolados com renda de até $\frac{1}{2}$ salário mínimo domiciliar *per capita* têm probabilidade reduzida para o acesso a informação sobre a existência do Cadastro Único em relação a indivíduos analfabeto não isolados da amostra.

Tabela 4 - Estimativa do efeito de tratamento médio sobre os tratados: probabilidade de conhecer o Cadastro Único para indivíduos com nível de renda domiciliar per capita até $\frac{1}{2}$ salário mínimo, Brasil 2014.

Método de pareamento	AI	ANI	ATT	Erro-padrão (Bootstrap)
AI7				
Vizinho mais próximo	0,5113	0,6430	-0,1317**	0,0393514
<i>Kernel Matching</i>	0,5113	0,6376	-0,1262**	0,029112
<i>Raius Matching</i>	0,5113	0,6355	-0,1242**	0,0305779
AI10				
Vizinho mais próximo	0,5154	0,6498	-0,1344**	0,0286887
<i>Kernel Matching</i>	0,5154	0,6463	-0,1308**	0,0265029
<i>Raius Matching</i>	0,5154	0,6425	-0,1271**	0,0328615
AI15				
Vizinho mais próximo	0,5312	0,6513	-0,1200**	0,0277122
<i>Kernel Matching</i>	0,5312	0,6510	-0,1197**	0,0230082
<i>Raius Matching</i>	0,5312	0,6505	-0,1193**	0,0219944

Fonte: Elaboração própria com dados da PNAD 2014. Nota: P-valor obtido pelo erro-padrão estimado por bootstrap com 50 repetições; **Significativo a 1% e * Significativo a 5%.

Verificado o efeito redutor do analfabetismo isolado sobre acesso a informação da existência do Cadastro Único dos indivíduos analfabetos isolados, agora buscamos verificar qual é o efeito de acordo com o gênero dos indivíduos. Novamente as estimações foram divididas para diferentes faixas etárias de definição da alfabetização no domicílio, e agora, também para observações do sexo feminino e masculino.

Como observado na tabela 5, aqueles domicílios em que a alfabetização foi definida apenas para indivíduos maiores de 7 anos, das estimações do ATT para mulheres analfabetas da amostra apenas o coeficiente de -9,44 pontos percentuais, estimado pelo método de pareamento dos três vizinhos mais próximos, se mostrou significativo. O resultado indicou que o efeito do tratamento diminui a probabilidade de acesso a informação sobre a existência do Cadastro Único de mulheres analfabetas isoladas em comparação com mulheres analfabetas não isoladas da amostra.

Os resultados para as observações em que apurou a alfabetização do domicílio a partir dos 10 anos, com base nos três métodos de pareamento, mostraram que o acesso a informação da existência do Cadastro Único é menor entre indivíduos analfabetos isolados do sexo feminino em relação aos seus pares do grupo de controle. Os resultados encontrados foram significativos e indicam que a diferença entre o acesso a informação de indivíduos analfabetos isolados e analfabetos não isolados do sexo feminino varia, em média, de -12,20 pontos percentuais, para o pareamento por raios, a -16,76 pontos percentuais, para pareamento dos três vizinhos mais próximos.

Ainda na tabela 5, os coeficientes encontrados para as observações do sexo feminino, em domicílios que contabilizaram a condição de alfabetização apenas para maiores de 15 anos, e conforme todos os métodos de pareamento, mostram que o efeito do tratamento diminui a probabilidade de acesso a informação sobre a existência do Cadastro Único. Os resultados foram estatisticamente significativos e indicam que para as observações do sexo feminino a diferença entre o acesso a informação do programa de analfabetos isolados e não isolados varia, em média, de -9,60 pontos percentuais no método de pareamento pelos três vizinhos mais próximos, a -13,03 pontos percentuais, no método de pareamento de *Kernel*.

Tabela 5 - Estimativa do efeito de tratamento médio sobre os tratados: probabilidade de conhecer o Cadastro Único para indivíduos do sexo feminino com nível de renda domiciliar per capita até ½ salário mínimo, Brasil 2014.

Método de pareamento	AI	ANI	ATT	Erro-padrão (Bootstrap)
AI7				
Vizinho mais próximo	0,5114	0,5932	-0,0817	0,0597352
<i>Kernel Matching</i>	0,5114	0,6058	-0,0944*	0,0471796
<i>Raius Matching</i>	0,5114	0,5949	-0,0834	0,0453697
AI10				
Vizinho mais próximo	0,5108	0,6784	-0,1676**	0,0565698
<i>Kernel Matching</i>	0,5108	0,6332	-0,1224**	0,040984
<i>Raius Matching</i>	0,5108	0,6328	-0,1220*	0,0520075
AI15				
Vizinho mais próximo	0,5195	0,6156	-0,0960*	0,0447101
<i>Kernel Matching</i>	0,5195	0,6498	-0,1303**	0,0351121
<i>Raius Matching</i>	0,5195	0,6403	-0,1207**	0,0358849

Fonte: Elaboração própria com dados da PNAD 2014. Nota: P-valor obtido pelo erro-padrão estimado por bootstrap com 50 repetições; **Significativo a 1% e * Significativo a 5%.

Partindo para as estimações da amostra de indivíduos do sexo masculino, a tabela 6 apresenta os resultados para todas as faixas etárias a partir das quais foram computadas a alfabetização no domicílio e os métodos de pareamento utilizados. De acordo com os três métodos de pareamento, há evidências que o acesso a informação sobre a existência do Cadastro Único é menor entre indivíduos analfabetos isolados do sexo masculino, em domicílios que a alfabetização foi computada a partir dos 7 anos, em relação aos seus pares na amostra. Todos os resultados foram significativos e mostram que o efeito médio do tratamento sobre os tratados varia, em média, de -12,93 pontos percentuais pelo método de pareamento por raios, a -14,91 pontos percentuais no método de pareamento dos três vizinhos mais próximos.

Quando a alfabetização estabelecida no domicílio é a partir dos 10 anos, os resultados denotam que o acesso a informação sobre o programa é menor entre analfabetos isolados do sexo masculino quando comparados com seu contrafactual da amostra. O efeito médio do tratamento sobre os tratados se mostrou significativo para todos os métodos de pareamento, e identificou que a diferença entre analfabetos

isolados e não isolados varia, em média, de -12,21 pontos percentuais por pareamento com os três vizinhos mais próximos, a -13,8 pontos percentuais para o pareamento por método de *Kernel*.

E por fim, ainda na tabela 6, encontramos resultados que indicam que o acesso a informação da existência do Cadastro Único é menor para indivíduos analfabetos isolados do sexo masculino da amostra, em domicílios que a alfabetização foi ponderada a partir dos 15 anos, em relação aos seus pares para todos os métodos de pareamento utilizados. Os coeficientes foram significativos e o efeito médio do tratamento sobre os tratados varia, em média, de -10,83 pontos percentuais para o método dos três vizinhos mais próximos, a -11,55 pontos percentuais para o método de pareamento por *Kernel*.

Tabela 6 - Estimativa do efeito de tratamento médio sobre os tratados: probabilidade de conhecer o Cadastro Único para indivíduos do sexo masculino com nível de renda domiciliar per capita até ½ salário mínimo, Brasil 2014

Método de pareamento	AI	ANI	ATT	Erro-padrão (Bootstrap)
AI7				
Vizinho mais próximo	0,5217	0,6708	-0,1491*	0,0590773
<i>Kernel Matching</i>	0,5217	0,6518	-0,1301*	0,0509396
<i>Raius Matching</i>	0,5217	0,6510	-0,1293*	0,0591306
AI10				
Vizinho mais próximo	0,5276	0,6497	-0,1221*	0,0539451
<i>Kernel Matching</i>	0,5276	0,6618	-0,1342**	0,040695
<i>Raius Matching</i>	0,5276	0,6658	-0,1382**	0,0417947
AI15				
Vizinho mais próximo	0,5360	0,6443	-0,1083**	0,0419485
<i>Kernel Matching</i>	0,5360	0,6515	-0,1155**	0,0282415
<i>Raius Matching</i>	0,5360	0,6471	-0,1111**	0,0306797

Fonte: Elaboração própria com dados da PNAD 2014. Nota: P-valor obtido pelo erro-padrão estimado por bootstrap com 50 repetições; **Significativo a 1% e * Significativo a 5%.

Para ambos os gêneros, em geral, podemos concluir que o efeito do tratamento de ser analfabeto isolado afeta negativamente na probabilidade de conhecer o Cadastro Único. Para os homens, destacamos o coeficiente de -14,91 pontos

percentuais na probabilidade de conhecer o Cadastro Único para analfabetos isolados em relação a analfabetos não isolados, na amostra de domicílios que a alfabetização foi computada a partir dos 7 anos, com pareamento pelos três vizinhos mais próximo. E para mulheres, o coeficiente de -16,76 pontos percentuais na probabilidade de conhecer o Cadastro Único para analfabetos isolados em relação a analfabetos não isolados nas observações em que se estabeleceu a alfabetização a partir de 10 anos no domicílio, com pareamento pelos três vizinhos mais próximo. Além disso, a maioria dos coeficientes absolutos mostram que há um efeito médio semelhante sobre indivíduos do sexo feminino e masculino no sentido de redução do acesso a informação devido a situação de analfabetismo isolado. Para o sexo feminino foi encontrado o maior valor absoluto nos coeficientes de ATT, e para o sexo masculino o total de coeficientes de ATT significativos foi maior.

A comparação destes resultados com os outros trabalhos presentes na literatura sobre o tema, apresentados na seção de externalidades da alfabetização nesse trabalho, fica prejudicada em razão da estratégia de análise ser diferente, e além do tema de políticas públicas também não ter sido abordado nessa linha de pesquisa. Neste trabalho, estimamos o efeito da ausência dessa externalidade no acesso a informação sobre políticas públicas, e no restante dos trabalhos é verificado a presença dessa externalidade.

4.2 Resultados da análise de sensibilidade

Os resultados encontrados na subseção 4.1 levam em consideração a hipótese de independência condicional, mas ainda assim a presença de variáveis não observáveis na especificação do modelo pode resultar em um viés na estimação do efeito médio do tratamento sobre os tratados. Como forma de quantificar esse viés foi feita a análise de sensibilidade dos resultados, através do estudo dos Limites de Rosenbaum. A análise permitiu avaliar a sensibilidade dos resultados estimados em relação a possíveis falhas na suposição de independência condicional decorrentes de variáveis não observadas que afetam o status de tratamento e a variável de resultado, simultaneamente, e que podem gerar um viés de variável omitida nas estimações. Em virtude de ambas as variáveis, de tratamento e resultado, serem binárias o teste utilizado foi de *mhbounds*.

A estatística de teste *Mantel-Haenszel* verifica a hipótese nula de ausência do efeito do tratamento. O teste determina dois limites que podem ser interpretados da seguinte maneira: a estatística Q_{mh+} representa o limite para o caso de sobrestimação do efeito do tratamento e Q_{mh-} aponta o limite para o caso de subestimação do efeito do tratamento. O fator Γ mede a diferença de probabilidade de receber o tratamento entre os indivíduos com características observáveis semelhantes, para verificar qualquer alteração nas estimativas devido a variáveis omitidas. Assim, indica o grau em que os indivíduos do grupo de tratamento e controle, que aparentemente são semelhantes em termos de características observáveis e dentro do suporte comum, diferem ou não em suas probabilidades de ser tratado.

O teste leva em consideração um intervalo com os níveis críticos do fator Γ e os níveis de significância da hipótese de sobrestimação e subestimação. A consistência dos resultados é feita ao elevar a variação do Γ até o momento em que o nível de significância (p_{mh+} e p_{mh-}) é afetado. Desta forma, foi calculado o intervalo de nível crítico com as respectivas probabilidades de não-rejeição da hipótese nula de sobrestimação e subestimação dos resultados dentro de um intervalo de 1 a 2 para o fator Γ .

Apresentamos, na tabela 7, os resultados para a análise de sensibilidade referente as amostras de todas as faixas de idade utilizadas para definição da alfabetização dos indivíduos no domicílio e para os diferentes métodos de pareamento. O efeito do tratamento parece ser robusto a uma possível presença de viés de seleção para os resultados das diferentes definições da amostra utilizadas em relação aos valores críticos de sobrestimação e subestimação. Para a amostra de indivíduos analfabetos todos os valores da estatística Q_{mh+} de *Mantel-Haenszel*, tendo como hipótese nula a sobrestimação do efeito do tratamento, foram estatisticamente significativos sendo possível rejeitar a hipótese nula de ausência de tratamento ao nível de pelo menos 1% de significância estatística para todos os valores de níveis críticos.

De forma, geral os resultados mostram que há apenas alguns momentos de alteração além do intervalo de significância de 10%. Isso ocorre quando para a estatística Q_{mh-} de *Mantel-Haenszel*, tendo como hipótese nula a subestimação do efeito do tratamento, o fator Γ assume valores entre 1,4 e 1,8 para alguns métodos de pareamento, mas acaba voltando a ter significância. Esta probabilidade está

associada a uma possível subestimação do modelo no que tange ao efeito do tratamento, mas acreditamos que isso não prejudica a análise uma vez que foram alterações leves e, ainda assim, na maioria dos valores foi possível rejeitar a hipótese nula. Em geral, para os demais valores de Γ os resultados da análise de sensibilidade indicam que as inferências a respeito do efeito do tratamento são válidas ao não ser observada influência de fatores não observáveis.

Tabela 7 - Análise de sensibilidade dos Limites de Rosenbaum para indivíduos analfabetos, Brasil 2014

Vizinho mais próximo					Kernel Matching				Radius matching			
AI7												
Γ	Q_mh+	Q_mh-	p_mh+	p_mh-	Q_mh+	Q_mh-	p_mh+	p_mh-	Q_mh+	Q_mh-	p_mh+	p_mh-
1	4,37661	4,37661	6,0e-06	6,0e-06	9,06825	9,06825	0	0	9,06825	9,06825	0	0
1.2	6,2805	2,48555	1,7e-10	0,006468	12,1493	6,05347	0	7,1e-10	12,1493	6,05347	0	7,1e-10
1.4	7,90057	0,890589	1,4e-15	0,186575	14,8228	3,53909	0	0,000201	14,8228	3,53909	0	0,000201
1.6	9,31444	0,392702	0	0,34727	17,2021	1,37605	0	0,084404	17,2021	1,37605	0	0,084404
1.8	10,5711	1,61009	0	0,053689	19,3582	0,465403	0	0,320821	19,3582	0,465403	0	0,320821
2	11,7037	2,70053	0	0,003461	21,3388	2,16909	0	0,015038	21,3388	2,16909	0	0,015038
AI10												
1	4,76162	4,76162	9,6e-07	9,6e-07	9,28512	9,28512	0	0	9,28512	9,28512	0	0
1.2	6,70787	2,82984	9,9e-12	0,002329	12,5211	6,11585	0	4,8e-10	12,5211	6,11585	0	4,8e-10
1.4	8,36518	1,20145	0	0,114789	15,3268	3,47062	0	0,00026	15,3268	3,47062	0	0,00026
1.6	9,81253	0,112276	0	0,455302	17,8217	1,19371	0	0,116296	17,8217	1,19371	0	0,116296
1.8	11,0997	1,35434	0	0,087814	20,0812	0,751958	0	0,226038	20,0812	0,751958	0	0,226038
2	12,2605	2,46671	0	0,006818	22,1557	2,54673	0	0,005437	22,1557	2,54673	0	0,005437
AI15												
1	6,8564	6,8564	3,5e-12	3,5e-12	10,4994	10,4994	0	0	10,4994	10,4994	0	0
1.2	9,66389	4,06616	0	0,000024	14,497	6,57033	0	2,5e-11	14,497	6,57033	0	2,5e-11
1.4	12,0536	1,71377	0	0,043286	17,9507	3,28174	0	0,000516	17,9507	3,28174	0	0,000516
1.6	14,1391	0,255907	0	0,399011	21,0123	0,445421	0	0,328008	21,0123	0,445421	0	0,328008
1.8	15,9928	2,05116	0	0,020126	23,7773	2,00779	0	0,022333	23,7773	2,00779	0	0,022333
2	17,6637	3,6587	0	0,000127	26,3097	4,24862	0	0,000011	26,3097	4,24862	0	0,000011

Fonte: Elaboração própria a partir dos dados da PNAD 2014. Nota: Q_mh+ = estatística de *Mantel-Haenszel* tendo como hipótese nula a sobrestimação do efeito do tratamento; Q_mh- = estatística de *Mantel-Haenszel* tendo como hipótese nula a subestimação do efeito do tratamento; p+ = probabilidade de não rejeição da hipótese de sobrestimação; p- = probabilidade de não rejeição da hipótese de subestimação.

Na tabela 8 são apresentados os resultados da análise de sensibilidade para indivíduos analfabetos do sexo feminino para a amostra com todas as faixas de idade utilizadas para definição da alfabetização dos indivíduos no domicílio e métodos de pareamento. Conforme podemos observar para todos os valores de Γ foi possível rejeitar a hipótese nula de sobrestimação dos resultados do ATT, ao nível de significância estatística de 1%, devido a influência de variáveis omitidas. Para valores de Γ referentes a hipótese de subestimação do efeito do tratamento nem todos os testes conseguiram rejeitar a hipótese nula de ausência do efeito do tratamento, em alguns momentos o nível de significância ultrapassa o nível de 10%, mas voltava a ganhar significância. Porém, ainda assim, foi rejeitada a hipótese nula que não há efeito do tratamento na maioria dos valores de Γ . Então, para a amostra de indivíduos do sexo feminino foi possível concluir que para a maior parte da análise, o efeito do tratamento encontrado não sofre influência de fatores não observados.

Os valores presentes na tabela 9 indicam os resultados da análise de sensibilidade para indivíduos analfabetos do sexo masculino em todas as definições de amostra empregadas no estudo. Conforme os valores do fator Γ encontrados, foi possível rejeitar a hipótese nula de sobrestimação de que não há efeito do tratamento para todos as faixas etárias e métodos de pareamento. Para a hipótese de subestimação nem todos os valores de Γ rejeitaram a hipótese nula, entre o intervalo de 1,4 e 1,8 nível de significância ultrapassou o valor de 10%. Mas, cabe ressaltar, que para a maior parte dos resultados foi possível rejeitar a hipótese nula de ausência do efeito do tratamento. Logo, na maioria dos valores de níveis críticos, não há perturbação na análise motivada por fatores não observados para os indivíduos da amostra de sexo masculino.

Tabela 8 - Análise de sensibilidade dos Limites de Rosenbaum para indivíduos analfabetos do sexo feminino, Brasil 2014

	Vizinho mais próximo				Kernel Matching				Radius matching			
	AI7											
Γ	Q_mh+	Q_mh-	p_mh+	p_mh-	Q_mh+	Q_mh-	p_mh+	p_mh-	Q_mh+	Q_mh-	p_mh+	p_mh-
1	1,98226	1,98226	0,023725	0,023725	6,06657	6,06657	6,5e-10	6,5e-10	6,06657	6,06657	6,5e-10	6,5e-10
1.2	3,33018	0,64105	0,000434	0,260745	8,18317	3,99465	1,1e-16	0,000032	8,18317	3,99465	1,1e-16	0,000032
1.4	4,47462	0,356348	3,8e-06	0,36079	10,0189	2,26563	0	0,011737	10,0189	2,26563	0	0,011737
1.6	5,47168	1,33832	2,2e-08	0,090396	11,6519	0,777548	0	0,218418	11,6519	0,777548	0	0,218418
1.8	6,35653	2,20586	1,0e-10	0,013697	13,1312	0,442491	0	0,329067	13,1312	0,442491	0	0,329067
2	7,153	2,98375	4,2e-13	0,001424	14,4898	1,61539	0	0,053113	14,4898	1,61539	0	0,053113
	AI10											
1	3,94012	3,94012	,000041	0,000041	6,79187	6,79187	5,5e-12	5,5e-12	6,79187	6,79187	5,5e-12	5,5e-12
1.2	5,31366	2,58097	5,4e-08	0,004926	9,13355	4,49862	0	3,4e-06	9,13355	4,49862	0	3,4e-06
1.4	6,48397	1,43583	4,5e-11	0,075526	11,1634	2,58445	0	0,004877	11,1634	2,58445	0	0,004877
1.6	7,50728	0,446266	3,0e-14	0,327703	12,9683	0,93692	0	0,1744	12,9683	0,93692	0	0,1744
1.8	8,41834	0,29044	0	0,38574	14,6025	0,432036	0	0,332858	14,6025	0,432036	0	0,332858
2	9,24081	1,07011	0	0,142286	16,1028	1,73029	0	0,041789	16,1028	1,73029	0	0,041789
	AI15											
1	4,64655	4,64655	1,7e-06	1,7e-06	7,96919	7,96919	7,8e-16	7,8e-16	7,96919	7,96919	7,8e-16	7,8e-16
1.2	6,63437	2,6714	1,6e-11	0,003777	10,803	5,18714	0	1,1e-07	10,803	5,18714	0	1,1e-07
1.4	8,32525	1,00523	0	0,157394	13,2533	2,86075	0	0,002113	13,2533	2,86075	0	0,002113
1.6	9,80042	0,343956	0	0,36544	15,4268	0,856119	0	0,195966	15,4268	0,856119	0	0,195966
1.8	11,1111	1,61579	0	0,05307	17,3909	0,84286	0	0,199653	17,3909	0,84286	0	0,199653
2	12,2922	2,75474	0	0,002937	19,1906	2,42407	0	0,007674	19,1906	2,42407	0	0,007674

Fonte: Elaboração própria a partir dos dados da PNAD 2014. Nota: Q_mh+ = estatística de *Mantel-Haenszel* tendo como hipótese nula a sobrestimação do efeito do tratamento; Q_mh- = estatística de *Mantel-Haenszel* tendo como hipótese nula a subestimação do efeito do tratamento; p+ = probabilidade de não rejeição da hipótese de sobrestimação; p- = probabilidade de não rejeição da hipótese de subestimação.

Tabela 9 - Análise de sensibilidade dos Limites de Rosenbaum para indivíduos analfabetos do sexo masculino, Brasil 2014

	Vizinho mais próximo				Kernel Matching				Radius matching			
AI7												
Γ	Q_mh+	Q_mh-	p_mh+	p_mh-	Q_mh+	Q_mh-	p_mh+	p_mh-	Q_mh+	Q_mh-	p_mh+	p_mh-
1	3,33679	3,33679	,000424	0,000424	6,03918	6,03918	7,7e-10	7,7e-10	6,03918	6,03918	7,7e-10	7,7e-10
1.2	4,67955	2,00576	1,4e-06	0,022441	8,21048	3,91247	1,1e-16	0,000046	8,21048	3,91247	1,1e-16	0,000046
1.4	5,82185	0,882871	2,9e-09	0,188653	10,0928	2,13678	0	0,016308	10,0928	2,13678	0	0,016308
1.6	6,81917	0,04913	4,6e-12	0,519592	11,7666	0,607772	0	0,271669	11,7666	0,607772	0	0,271669
1.8	7,70585	0,806837	6,6e-15	0,20988	13,2824	0,651387	0	0,257398	13,2824	0,651387	0	0,257398
2	8,50523	1,57326	0	0,057829	14,6742	1,85762	0	0,031611	14,6742	1,85762	0	0,031611
AI10												
1	2,71533	2,71533	,00331	,00331	6,23209	6,23209	2,3e-10	2,3e-10	6,23209	6,23209	2,3e-10	2,3e-10
1.2	4,11023	1,33016	,00002	,091734	8,57019	3,93893	0	0,000041	8,57019	3,93893	0	0,000041
1.4	5,29653	0,160811	5,9e-08	,436121	10,5945	2,02212	0	0,021582	10,5945	2,02212	0	0,021582
1.6	6,33176	0,719167	1,2e-10	,236019	12,3926	0,37018	0	0,355624	12,3926	0,37018	0	0,355624
1.8	7,25185	1,61245	2,1e-13	,053432	14,0194	1,00478	0	0,157502	14,0194	1,00478	0	0,157502
2	8,08121	2,41316	3,3e-16	,007907	15,5118	2,30949	0	0,010458	15,5118	2,30949	0	0,010458
AI15												
1	4,77744	4,77744	8,9e-07	8,9e-07	7,17147	7,17147	3,7e-13	3,7e-13	7,17147	7,17147	3,7e-13	3,7e-13
1.2	6,74704	2,82121	7,5e-12	0,002392	9,96435	4,42651	0	4,8e-06	9,96435	4,42651	0	4,8e-06
1.4	8,42307	1,17141	0	0,120717	12,3771	2,12847	0	0,016649	12,3771	2,12847	0	0,016649
1.6	9,88591	0,162439	0	0,43548	14,516	0,146033	0	0,441948	14,516	0,146033	0	0,441948
1.8	11,1862	1,42135	0	0,077607	16,4477	1,5347	0	0,062429	16,4477	1,5347	0	0,062429
2	12,3584	2,54863	0	0,005407	18,2171	3,10175	0	0,000962	18,2171	3,10175	0	0,000962

Fonte: Elaboração própria a partir dos dados da PNAD 2014. Nota: Q_mh+= estatística de *Mantel-Haenszel* tendo como hipótese nula a sobrestimação do efeito do tratamento; Q_mh-= estatística de *Mantel-Haenszel* tendo como hipótese nula a subestimação do efeito do tratamento; p+ = probabilidade de não rejeição da hipótese de sobrestimação; p- = probabilidade de não rejeição da hipótese de subestimação.

5. CONSIDERAÇÕES FINAIS

O presente estudo teve como proposta avaliar se há relação entre o analfabetismo isolado e o acesso a informação sobre o Cadastro Único para indivíduos com renda domiciliar per capita de até $\frac{1}{2}$ salário mínimo, usando os dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) de 2014 e o método de *Propensity Score Matching* (PSM). O estimador do PSM foi aplicado para diferentes definições da amostra e métodos de pareamento.

O tema de externalidades da alfabetização ganhou espaço após pesquisa de Basu e Foster (1998). Para os autores, uma pessoa capaz de desempenhar atividades de leitura e escrita proporciona que a alfabetização seja vista como um bem público para os membros analfabetos na residência. No entanto, analfabetos isolados não teriam acesso à informação dada a ausência dessa externalidade dentro do domicílio, e estariam em uma situação de desvantagem em relação a analfabetos não isolados.

Dada a importância da implantação de políticas públicas como estratégia dos governos no combate à pobreza e à desigualdade social (KERSTENETZKY, 2009). A pesquisa avaliou o desafio de direcionar o acesso de programas sociais aos indivíduos que realmente demandam desses benefícios, com base na focalização de políticas públicas justificadas pelo grau de exclusão de indivíduos dada a menor informação, ou mesmo a falta de acesso a ela, visto que residir em um domicílio com apenas indivíduos analfabetos pode impor uma série de restrições ligadas ao acesso de informações e serviços.

Os resultados confirmaram que a ausência de retorno social domiciliar da alfabetização, dado o fato de não residir com nenhum indivíduo alfabetizado, foi determinante na redução do acesso a informação sobre a existência do Cadastro Único. De maneira geral, os maiores coeficientes encontrados entre as três definições da amostra com e sem distinção de gênero para a diferença entre analfabetos isolados e não isolados foram: sem a distinção de gênero de -13,44 p.p., para indivíduos do sexo feminino de -16,76 p.p. e para indivíduos do sexo masculino de -14,91 p.p., todos pelo método de pareamento dos três vizinhos mais próximos. Visto que os mesmos seriam potenciais beneficiários, por se encaixarem no critério de renda para elegibilidade no Cadastro Único, os resultados levam a evidência de que a menor informação, ou mesmo a falta de acesso a ela, que a situação de analfabetismos isolado pode impor, afeta esses indivíduos criando um obstáculo para o conhecimento

sobre seus direitos e, possivelmente, deixando-os de fora da rede de proteção social do governo.

Evidenciada a importância do retorno social da alfabetização para o acesso a informação, uma vez que a falta de um indivíduo alfabetizado restringe o acesso às atividades que requerem o uso de leitura e escrita no domicílio. Fica a sugestão para pesquisas futuras a análise da influência relacionada a posição dos membros nos domicílios para externalidade social da alfabetização, assim como a verificação do efeito da externalidade em níveis mais abrangentes de contato dos indivíduos.

REFERÊNCIAS BIBLIOGRÁFICAS

AAKVIK, Arild. Bounding a matching estimator: the case of a Norwegian training program. **Oxford bulletin of economics and statistics**, v. 63, n. 1, p. 115-143, 2001.

ANDRADE, Mônica Viegas et al. Income transfer policies and the impacts on the immunization of children: the Bolsa Família Program. **Cadernos de saúde pública**, v. 28, n. 7, p. 1347-1358, 2012

AUSTIN, Peter C.; JEMBERE, Nathaniel; CHIU, Maria. Propensity score matching and complex surveys. **Statistical Methods in Medical Research**, July 26, 2016.

DE BARROS, Ricardo Paes et al. A focalização do Programa Bolsa Família em perspectiva comparada. In: Jorge Abrahão de Castro, Lúcia Modesto. Brasília (Org.) **Bolsa família 2003-2010: avanços e desafios**. Brasília: IPEA, 2 v. (366p.), 2010.

BASU, Kaushik; FOSTER, James E. On measuring literacy. **The Economic Journal**, v. 108, n. 451, p. 1733-1749, 1998.

BASU, Kaushik; NARAYANC, Ambar; RAVALLION, Martin. Is literacy shared within households? Theory and evidence for Bangladesh. **Labour Economics**, v. 8, p. 649-665, 2001.

BECKER, Sascha O.; ICHINO, Andrea. Estimation of average treatment effects based on propensity scores. **The stata journal**, v. 2, n. 4, p. 358-377, 2002.

BECKER, Sascha O.; CALIENDO, Marco. Sensitivity analysis for average treatment effects. **The Stata Journal**. V. 7, N. 1, pp. 71–83, 2007.

BOROOAH, Vani. *The Role of Maternal Literacy in Reducing the Risk of Child Malnutrition in India*. In: Gender and Discrimination: Health, Nutritional Status and Role of Women in India. (Eds: Pal, Manoranjan, Bharati, Premananda, Ghosh, Bholanath and Vasulu, TS), **Oxford University Press**, pp. 141-162, 2009.

BRASIL. Decreto nº 3.877, de 24 de julho de 2001. Institui o Cadastramento Único para Programas Sociais do Governo Federal. Brasília – DF. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto/2001/D3877.htm

_____. Decreto nº 6.135, de 26 de junho de 2007. Dispõe sobre o Cadastro Único para Programas Sociais do Governo Federal e dá outras providências. Brasília – DF. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2007/Decreto/D6135.htm#art14

CALIENDO, Marco; KOPEINIG, Sabine. Some practical guidance for the implementation of propensity score matching. **Journal of economic surveys**, v. 22, n. 1, p. 31-72, 2008.

CAMERON, A. Colin; TRIVEDI, Pravin K. **Microeconometrics: methods and applications**. Cambridge university press, 2005.

COADY, David; GROSH, Margaret E.; HODDINOTT, John. **Targeting of transfers in developing countries: Review of lessons and experience**. World Bank Publications, 2004.

DE ARAÚJO, Aracy Alves; GOMES, Marília Fernandes Maciel; LIMA, João Eustáquio. Influência do Programa Bolsa Família na redução do trabalho infantil: evidências para o Nordeste brasileiro. **Revista Econômica do Nordeste**, v. 45, n. 3, p. 33-45, 2016.

DE BALDINI ROCHA, Maúna Soares; PONCZEK, Vladimir. The effects of adult literacy on earnings and employment. **Economics of Education Review**, v. 30, n. 4, p. 755-764, 2011.

DE SOUZA, Luana Passos; WALTENBEG, Fábio Domingues. Bolsa Família e assimetrias de gênero: reforço ou mitigação?. **Revista Brasileira de Estudos de População**, v. 33, n. 3, p. 517-539, 2016.

DIPRETE, Thomas A.; GANGL, Markus. Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments. **Sociological methodology**, v. 34, n. 1, p. 271-310, 2004.

DIREITO, Denise do Carmo, et al. **O Cadastro Único como instrumento de articulação de políticas sociais**. No. 327. International Policy Centre for Inclusive Growth, 2016.

DUGOFF, Eva H.; SCHULER, Megan; STUART, Elizabeth A. Generalizing observational study results: applying propensity score methods to complex surveys. **Health services research**, v. 49, n. 1, p. 284-303, 2014.

FERRER, Ana; GREEN, David A.; RIDDELL, W. Craig. The effect of literacy on immigrant earnings. **Journal of Human Resources**, v. 41, n. 2, p. 380-410, 2006.

FREIMUTH, Vicki S.; METTGER, Wendy. Is there a hard-to-reach audience? **Public health reports**, v. 105, n. 3, p. 232, 1990.

GIBSON, J. (2001). Literacy and Intra household Externalities. **World Development**, vol. 29, p.p. 155-166, 2001.

IMBENS, Guido W. Matching methods in practice: Three examples. **Journal of Human Resources**, v. 50, n. 2, p. 373-419, 2015.

IVERSEN, Vegard; PALMER-JONES, Richard. Literacy sharing, assortative mating, or what? Labour market advantages and proximate illiteracy revisited. **The Journal of Development Studies**, v. 44, n. 6, p. 797-838, 2008.

KERSTENETZKY, Celia Lessa. Redistribuição e desenvolvimento? A economia política do programa bolsa família. **Dados-Revista de Ciências Sociais**, v. 52, n. 1, p. 53-83, 2009.

HUSAIN, Zakir; DUTTA, Mousumi. **Proximate illiteracy and modern contraceptive use in India: Analysis of DHS data**. MPRA Paper No. 39951, July 2012.

PAES-SOUSA, Rômulo et al. **Plano Brasil Sem Miséria: Incremento e mudança na política de proteção e promoção social no Brasil**. IPC-IG Working Paper, n. 113, 2013.

RIBEIRO, Felipe Garcia; SOUZA, André Portela Fernandes de. Uma investigação e aplicação das novas medidas de alfabetização no Brasil. **Nova Economia**, v. 23, n. 3, p. 623-656, 2013.

RIBEIRO, Felipe Garcia et al. Retorno social domiciliar da alfabetização: análise da associação com o tabagismo de analfabetos corresidentes no Brasil. **Revista Panamericana de Salud Pública**, v. 39, n. 6, p. 316-321, 2016.

ROSENBAUM, Paul R.; RUBIN, Donald B. The central role of the propensity score in observational studies for causal effects. **Biometrika**, p. 41-55, 1983.

ROSENBAUM, Paul R. Observational studies. In: **Observational Studies**. Springer New York, 2002.

SARR, Leopold R. **The impact of family literacy on the earnings of illiterates: Evidence from Senegal**. Cornell Food and Nutrition Policy Program Working Paper No. 159, 2004.

SHAGHAGHI, Abdolreza; BHOPAL, Raj S.; SHEIKH, Aziz. Approaches to Recruiting 'Hard-To-Reach' Populations into Re-search: A Review of the Literature. **Health Promotion**, v. 1, n. 2, p. 01-09, 2011

SOARES, Fábio Veras; RIBAS, Rafael Perez; OSÓRIO, Rafael Guerreiro. Evaluating the impact of Brazil's Bolsa Família: Cash transfer programs in comparative perspective. **Latin American Research Review**, v. 45, n. 2, p. 173-190, 2010.

SOARES, Sergei; SOUZA, Pedro Herculano G. Ferreira de; OSÓRIO, Rafael Guerreiro e SILVEIRA, Fernando Gaiger. Os impactos do benefício do programa bolsa família sobre a desigualdade e a pobreza. In: Jorge Abrahão de Castro, Lúcia Modesto (Org.) **Bolsa família 2003-2010: avanços e desafios**. Brasília: Ipea, 2 v. (366p.), 2010.

SOUZA, André Portela Fernandes de et al. **Uma investigação sobre a focalização do Programa Bolsa Família e seus determinantes imediatos**. *Working Paper*, Nov 2013.

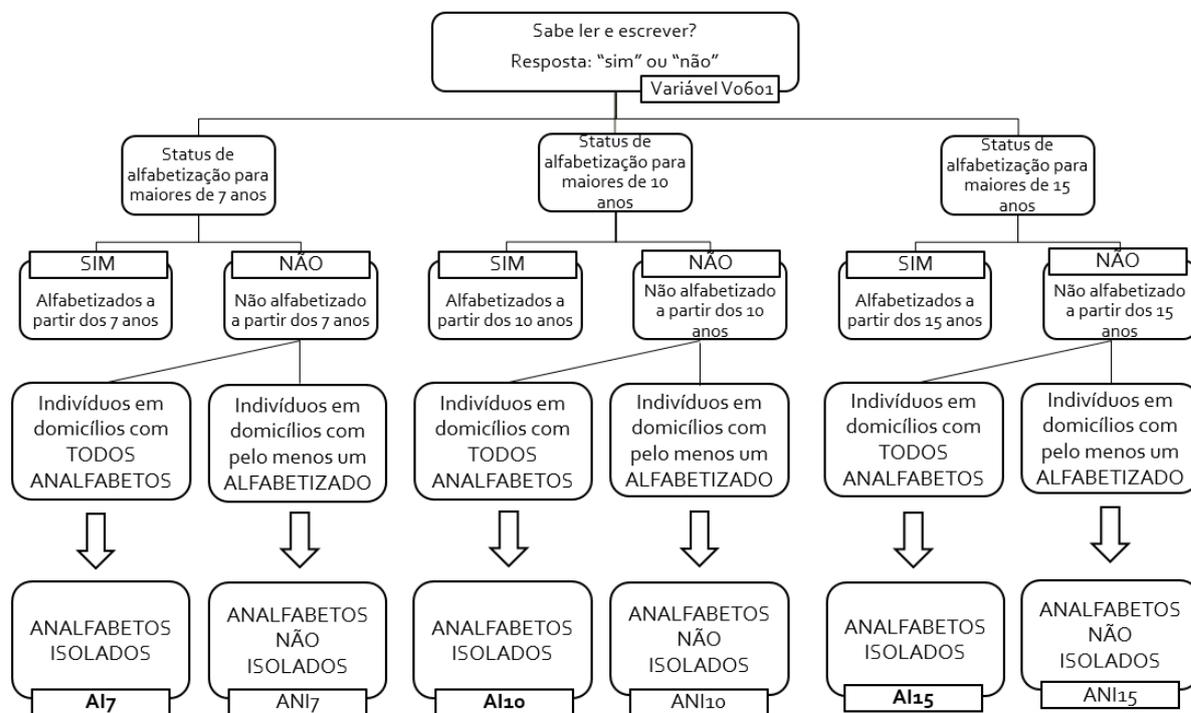
TAVARES, Priscilla Albuquerque et al. Uma avaliação do Programa Bolsa Família: focalização e impacto na distribuição de renda e pobreza. **Pesquisa e Planejamento Econômico**, v. 39, n. 1, 2009.

TAVARES, Priscilla Albuquerque. Efeito do Programa Bolsa Família sobre a oferta de trabalho das mães. **Economia e sociedade**, v. 19, n. 3, p. 613-635, 2016.

UNESCO. Relatório de Monitoramento Global de Educação para Todos. Publicado em 2014 pela Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Título original: ***Teaching and Learning: Achieving Quality for All***; *EFA Global Monitoring Report*, 2013)

APÊNDICE

APÊNDICE A – Esquema de construção das variáveis de analfabetos isolado e analfabetos não isolados, Brasil 2014.



Fonte: Elaboração própria a partir dos dados da PNAD 2014.

APÊNDICE B - Estimação do *propensity score* fundamentado nas características observáveis para indivíduos analfabetos AI7, Brasil 2014.

Variáveis	Coeficiente	Erro-padrão	P-valor
Nível máximo de escolaridade	0,511112	0,2214559	0,021
Número de pessoas maiores de 9 anos no domicílio	-0,6240813	0,2396831	0,009
Esgoto	-0,1673037	0,3972039	0,674
Número de quartos	-1,709211	0,3203394	0,000
Urbano	0,0006261	0,1605078	0,997
Número de pessoas menores de 9 anos no domicílio	0,3324764	0,1998053	0,096
Banheiro	-0,5487954	0,4029631	0,173
Casado	-1,759861	0,3412345	0,000
Idade	0,094624	0,0147844	0,000
Cor	-1,616935	0,3548621	0,000
Variáveis de interação			
Idade ²	-0,0007672	0,0001318	0,000
Npessoas9*urbano	-0,4473248	0,0989956	0,000
Max_escolaridade ²	-,1964366	0,0342359	0,000
Nquartos*npessoas	0,3405819	0,0652858	0,000
Npessoas*max_escolaridade	-0,1917856	0,0334276	0,000
Idae*max_escolaridade	-0,0096776	0,0020239	0,000
Banheiro*nquartos	0,4708062	0,2072508	0,023
Cor*npessoas9	0,6415049	0,1326248	0,000
Cor*idade	0,0249842	0,0066358	0,000
Npessoas ²	0,0180272	0,0241644	0,456
Casado*npessoas	0,4090122	0,1007405	0,000
Idade*npessoas	-0,009681	0,0026486	0,000
Npessoas9*nquartos	-0,2286397	0,0855323	0,008
Esgoto*max_escolaridade	0,2813201	0,1040993	0,007
Banheiro*max_escolaridade	-0,284839	0,1063969	0,007
Esgoto*npessoas	-0,2423904	0,1231997	0,049
_cons	3,473738	0,6927246	0,000

Fonte: Elaboração própria a partir dos dados da PNAD 2014.

Nota: Regressão *logit*. Variável dependente binária: um, se o indivíduo é analfabeto isolado, e zero, caso contrário.

APÊNDICE C - Estimação do *propensity score* fundamentado nas características observáveis para indivíduos analfabetos AI10, Brasil 2014.

Variáveis	Coefficiente	Erro-padrão	P-valor
Nível máximo de escolaridade	0,7662385	0,2231333	0,001
Número de pessoas maiores de 9 anos no domicílio	-0,6089807	0,1592864	0,000
Esgoto	-0,7627844	0,1882075	0,000
Número de pessoas menores de 9 anos no domicílio	0,6642128	0,1975437	0,001
Idade	0,1251214	0,013256	0,000
Número de quartos	-2,103794	0,2847748	0,000
Banheiro	-1,554692	0,4368245	0,000
Cor	-1,594683	0,3523665	0,000
Urbano	-0,3604577	0,4099933	0,379
Casado	-0,9752195	0,1703616	0,000
Variáveis de interação			
Idade ²	-0,0010695	0,0001351	0,000
Max_escolaridade ²	-0,2537327	0,0359303	0,000
Nquartos*npessoas	0,3801973	0,0615534	0,000
Banheiro*nquartos	0,9003419	0,2065195	0,000
Npessoas*max_escolaridade	-0,2169555	0,0337958	0,000
Idade*max_escolaridade	-0,0110415	0,0020496	0,000
Casado*npessoas9	0,5271291	0,1126828	0,000
Idade*npessoas	-0,0072402	0,0020478	0,000
Cor*npessoas9	0,5657691	0,1272353	0,000
Cor*idade	0,0250642	0,0066203	0,000
Urbano*npessoas	-0,1615658	0,0697914	0,021
Npessoas9*nquartos	-0,2169119	0,0852227	0,011
Urbano*banheiro	0,7528712	0,3361237	0,025
Esgoto*max_escolaridade	0,2371619	0,1007342	0,019
Banheiro*max_escolaridade	-0,2316836	0,1120229	0,039
_cons	3,448254	0,6613584	0,000

Fonte: Elaboração própria a partir dos dados da PNAD 2014.

Nota: Regressão *logit*. Variável dependente binária: um, se o indivíduo é analfabeto isolado, e zero, caso contrário.

APÊNDICE D - Estimação do *propensity score* fundamentado nas características observáveis para indivíduos analfabetos AI15, Brasil 2014.

Variáveis	Coefficiente	Erro-padrão	P-valor
Máxima escolaridade	-1,019155	0,0610748	0,000
Idade	0,1257973	0,0101887	0,000
Esgoto	-0,7730197	0,2271606	0,001
Número de quartos do domicílio	0,1251684	0,1079541	0,246
Número de pessoas maiores de 9 anos no domicílio	0,213616	0,125603	0,089
Homem	-0,0802118	0,2068923	0,698
Urbano	0,4797622	0,2368778	0,043
Cor	-0,7332105	0,3230388	0,023
Banheiro	-10,046283	0,3313604	0,002
Casado	-10,93333	0,2290614	0,000
Variáveis de interação			
Idade ²	-0,0011567	0,0000905	0,000
Urbano*npessoas	-0,1688696	0,0483331	0,000
Urbano*max_escolaridade	0,1429516	0,0366457	0,000
Idade*npessoas	-0,0104213	0,0013727	0,000
Casado*npessoas	0,3749189	0,0522599	0,000
Idade*max_escolaridade	0,0021826	0,0007256	0,003
Casado*cor	0,6123665	0,1749851	0,000
Cor*npessoas	0,2022492	0,0514791	0,000
Banheiro*npessoas	0,1578954	0,0556299	0,005
Esgoto*npessoas	0,1247387	0,0542143	0,021
Urbano*nquartos	-0,2589015	0,103405	0,012
Banheiro*idade	0,0119592	0,0044415	0,007
Cor*nquartos	-0,2402885	0,1265596	0,058
Cor*homem	-0,3863532	0,1701856	0,023
Homem*max_escolaridade	-0,0906053	0,0349512	0,010
Homem*npessoas	0,0785423	0,0395066	0,047
Npessoas ²	-0,0249382	0,0090961	0,006
Nquartos*max_escolaridade	0,0449765	0,0217743	0,039
Casado*homem	-0,2924769	0,1415399	0,039
Cor*urbano	0,2754969	0,1746184	0,115
_cons	0,6286974	0,4709269	0,182

Fonte: Elaboração própria a partir dos dados da PNAD 2014.

Nota: Regressão *logit*. Variável dependente binária: um, se o indivíduo é analfabeto isolado, e zero, caso contrário.

APÊNDICE E - Estimação do *propensity score* fundamentado nas características observáveis para indivíduos analfabetos AI7 do sexo feminino, Brasil 2014.

Variáveis	Coefficiente	Erro-padrão	P-valor
Nível máximo de escolaridade	0,203899	0,2862092	0,476
Número de pessoas maiores de 9 anos no domicílio	-0,6671211	0,2364271	0,005
Esgoto	-1,524105	0,3955708	0,000
Número de quartos	-1,254,022	0,4070184	0,002
Idade	0,0522921	0,0127719	0,000
Urbano	0,9252483	0,4143716	0,026
Banheiro	-0,3794261	0,2112524	0,072
Variáveis de interação			
Max_escolaridade ²	-0,1852045	0,0467348	0,000
Urbano*npessoas	-0,3511795	0,1090391	0,001
Npessoas ²	0,0777538	0,0238333	0,001
Npessoas*max_escolaridade	-0,1574519	0,0462711	0,001
Idade*max_escolaridade	-0,0104251	0,0028277	0,000
Idade ²	-0,0005346	0,0001692	0,002
Nquartos ²	0,2687181	0,0957892	0,005
Esgoto*max_escolaridade	0,3412853	0,1333635	0,010
Idade*esgoto	0,0151847	0,0075244	0,044
_cons	331,764	0,7144137	0,000

Fonte: Elaboração própria a partir dos dados da PNAD 2014.

Nota: Regressão *logit*. Variável dependente binária: um, se o indivíduo é analfabeto isolado, e zero, caso contrário.

APÊNDICE F - Estimação do *propensity score* fundamentado nas características observáveis para indivíduos analfabetos AI10 do sexo feminino, Brasil 2014.

Variáveis	Coefficiente	Erro-padrão	P-valor
Nível máximo de escolaridade	0,5386643	0,2958195	0,069
Número de pessoas maiores de 9 anos no domicílio	-0,775892	0,2953325	0,009
Esgoto	-0,799616	0,2655714	0,003
Número de pessoas menores de 9 anos no domicílio	1,102723	0,3373304	0,001
Idade	0,1251532	0,0209979	0,000
Banheiro	-2,326454	0,675914	0,001
Número de quartos	-2,106,545	0,5217344	0,000
Urbano	0,2643037	0,646429	0,683
Casado	0,5627404	0,6851184	0,411
Variáveis de interação			
Idade ²	-0,0008487	0,0002365	0,000
Max_escolaridade ²	-0,2347416	0,0508312	0,000
Urbano*npessoas	-0,3303732	0,1100395	0,003
Idade*max_escolaridade	-0,0122284	0,0029425	0,000
Npessoas*max_escolaridade	-0,2099781	0,0506261	0,000
Nquartos ²	0,0434848	0,1381455	0,753
Nquartos*banheiro	0,6835801	0,3319947	0,039
Casado*idade	-0,0183008	0,0127332	0,151
Esgoto*max_escolaridade	0,2766487	0,133686	0,039
Nquartos*npessoas	0,4462026	0,1232932	0,000
Nquartos*npessoas9	-0,4052809	0,1362599	0,003
Idade*npessoas	-0,0074887	0,0030288	0,013
Banheiro*npessoas9	0,3448521	0,1680884	0,040
Urbano*banheiro	0,8859264	0,5279849	0,093
Casado*npessoas9	0,2957242	0,1772406	0,095
_cons	3,279287	1,005588	0,001

Fonte: Elaboração própria a partir dos dados da PNAD 2014.

Nota: Regressão *logit*. Variável dependente binária: um, se o indivíduo é analfabeto isolado, e zero, caso contrário.

APÊNDICE G - Estimação do *propensity score* fundamentado nas características observáveis para indivíduos analfabetos AI15 do sexo feminino, Brasil 2014.

Variáveis	Coefficiente	Erro-padrão	P-valor
Nível máximo de escolaridade	-0,9155706	0,0699648	0,000
Idade	0,1333196	0,0114757	0,000
Urbano	0,660351	0,3185628	0,038
Número de pessoas maiores de 9 anos no domicílio	-0,0758361	0,1053521	0,472
Número de quartos	0,6080329	0,1381613	0,000
Esgoto	-0,5532374	0,1709718	0,001
Número de pessoas menores de 9 anos no domicílio	0,650662	0,1700318	0,000
Variáveis de interação			
Idade ²	-0,0011045	0,0001275	0,000
Idade*npessoas	-0,0116249	0,0018463	0,000
Npessoas*max_escolaridade	-0,1044348	0,0250678	0,000
Npessoas*idade	0,0121121	0,0031385	0,000
Urbano*npessoas	-0,0304064	0,0944132	0,747
Npessoas*max_escolaridade	0,0523322	0,0156321	0,001
Urbano*max_escolaridade	0,1410059	0,0514555	0,006
Npessoas*nquartos	-0,1427748	0,0513607	0,005
Urbano*nquartos	-0,4320804	0,1527372	0,005
Npessoas*esgoto	0,2341026	0,1054421	0,026
Npessoas*urbano	-0,2403091	0,1265135	0,058
_cons	-0,930977	0,4171659	0,026

Fonte: Elaboração própria a partir dos dados da PNAD 2014.

Nota: Regressão *logit*. Variável dependente binária: um, se o indivíduo é analfabeto isolado, e zero, caso contrário.

APÊNDICE H - Estimação do *propensity score* fundamentado nas características observáveis para indivíduos analfabetos AI7 do sexo masculino, Brasil 2014.

Variáveis	Coefficiente	Erro-padrão	P-valor
Nível máximo de escolaridade	0,7577758	0,3153591	0,016
Número de pessoas maiores de 9 anos no domicílio	-0,5344015	0,1626032	0,001
Esgoto	-0,6385742	0,2170089	0,003
Urbano	-0,8372709	0,4632807	0,071
Casado	-2,620589	0,4705448	0,000
Número de quartos	-2,008852	0,4203631	0,000
Idade	0,0632614	0,0178902	0,000
Número de pessoas menores de 9 anos no domicílio	-0,7537907	0,2987404	0,012
Cor	-1,994011	0,5000419	0,000
Banheiro	-1,441441	0,6048043	0,017
Variáveis de interação			
Idade ²	-0,0006455	0,0001953	0,001
Max_escolaridade ²	-0,2248055	0,0506127	0,000
Npessoas9*urbano	-0,4692084	0,1409194	0,001
Nquartos*npessoas	0,2712117	0,0715914	0,000
Npessoas*max_escolaridade	-0,2225892	0,0458267	0,000
Idade*max_escolaridade	-0,0149882	0,004153	0,000
Cor*npessoas9	0,7900339	0,1777529	0,000
Cor*idade	0,0252071	0,0092505	0,006
Banheiro*max_escolaridade	-0,3566379	0,1382816	0,010
Banheiro*nquartos	0,7910754	0,273027	0,004
Casado*npessoas9	0,4709483	0,1535348	0,002
Banheiro*casado	0,9839162	0,372776	0,008
Casado*max_escolaridade	0,3882092	0,1743387	0,026
Npessoas9*npessoas	0,0948775	0,0439916	0,031
Urbano*banheiro	0,7988487	0,4701818	0,089
_cons	5,288949	0,9045503	0,000

Fonte: Elaboração própria a partir dos dados da PNAD 2014.

Nota: Regressão *logit*. Variável dependente binária: um, se o indivíduo é analfabeto isolado, e zero, caso contrário.

APÊNDICE I - Estimação do *propensity score* fundamentado nas características observáveis para indivíduos analfabetos AI10 do sexo masculino, Brasil 2014.

Variáveis	Coefficiente	Erro-padrão	P-valor
Nível máximo de escolaridade	0,6635573	0,2779413	0,017
Número de pessoas maiores de 9 anos no domicílio	-0,3579334	0,1810476	0,048
Esgoto	-1,288098	0,4776047	0,007
Número de pessoas menores de 9 anos no domicílio	0,1785832	0,1381648	0,196
Cor	-1,51558	0,3520059	0,000
Número de quartos	-1,242664	0,3022362	0,000
Urbano	-0,6815057	0,2140218	0,001
Idade	0,1368668	0,0193904	0,000
Casado	-2,715389	0,4285624	0,000
Variáveis de interação			
Idade ²	-0,0011076	0,0001898	0,000
Max_escolaridade ²	-0,2725651	0,0506912	0,000
Npessoas*max_escolaridade	-0,2161111	0,0449135	0,000
Idade*max_escolaridade	-0,0173566	0,0041861	0,000
Nquartos*npessoas	0,22617	0,0627263	0,000
Casado*npessoas9	0,5292075	0,1631234	0,001
Idade*npessoas	-0,0048459	0,0028224	0,086
Casado*max_escolaridade	0,4235692	0,1812841	0,019
Casado*cor	1,158879	0,3898387	0,003
Cor*npessoas9	0,4336535	0,1452309	0,003
Esgoto*nquartos	0,5381371	0,2771668	0,052
Casado*urbano	0,537335	0,2948704	0,068
_cons	2,594095	0,7586127	0,001

Fonte: Elaboração própria a partir dos dados da PNAD 2014.

Nota: Regressão *logit*. Variável dependente binária: um, se o indivíduo é analfabeto isolado, e zero, caso contrário.

APÊNDICE J - Estimação do *propensity score* fundamentado nas características observáveis para indivíduos analfabetos AI15 do sexo masculino, Brasil 2014.

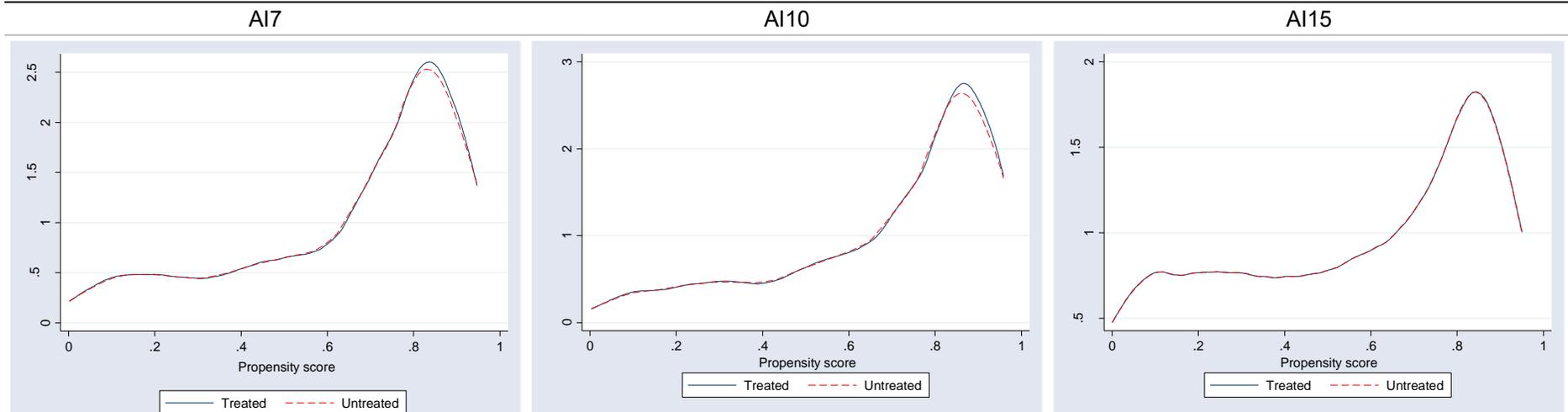
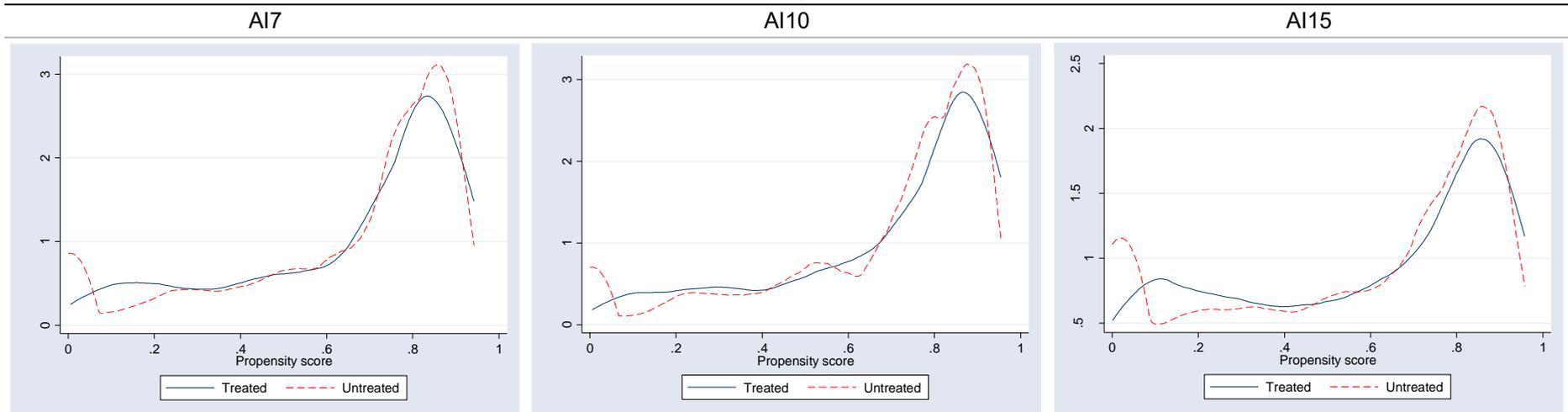
Variáveis	Coefficiente	Erro-padrão	P-valor
Nível máximo de escolaridade	-0,981405	0,0408056	0,000
Idade	0,1042129	0,0122318	0,000
Urbano	-0,2171242	0,2978342	0,466
Número de quartos	0,3042053	0,0988464	0,002
Cor	-0,8778675	0,4039193	0,030
Casado	-0,2585622	0,3593285	0,000
Esgoto	-0,6606231	0,2098167	0,002
Número de pessoas maiores de 9 anos no domicílio	0,0022989	0,0500376	0,963
Variáveis de interação			
Idade ²	-0,0008551	0,0001241	0,000
Casado*max_escolaridade	0,1731343	0,0521506	0,001
Casado*cor	0,7853936	0,2544565	0,002
Casado*urbano	0,5329366	0,1994034	0,008
Esgoto*max_escolaridade	0,1418379	0,0571074	0,013
Urbano*nquartos	-0,1921003	0,129883	0,139
Casado*npessoas	0,2946596	0,0795602	0,000
Idade*npessoas	-0,0067498	0,001832	0,000
Cor*npessoas	0,1991585	0,0747678	0,008
Cor*nquartos	-0,351687	0,1952798	0,072
_cons	0,8834396	0,3345504	0,008

Fonte: Elaboração própria a partir dos dados da PNAD 2014

Nota: Regressão *logit*. Variável dependente binária: um, se o indivíduo é analfabeto isolado, e zero, caso contrário.

APÊNDICE K - Verificação da hipótese de sobreposição para todas as definições da amostra e métodos de pareamento, Brasil 2014.

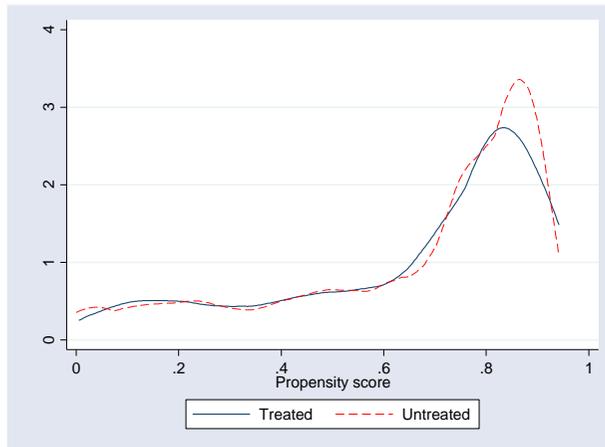
Pareamento pelos três vizinhos mais próximos

Pareamento por *Kernel*

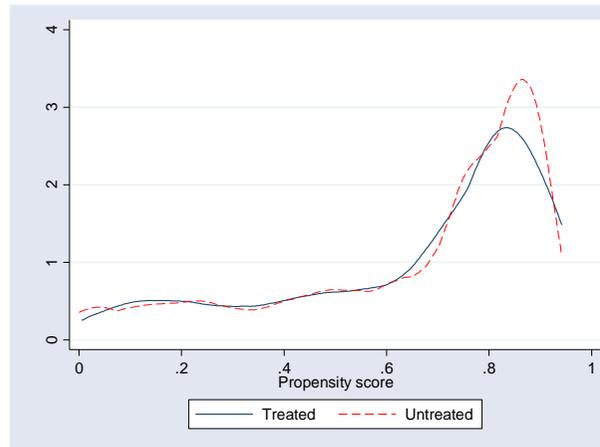
Continua...

Pareamento por Raio

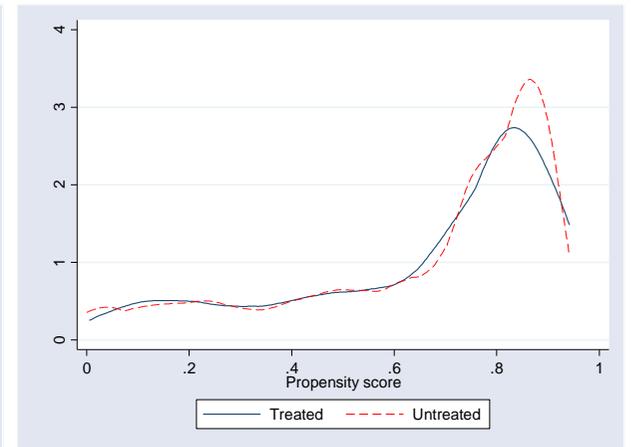
AI7



AI10



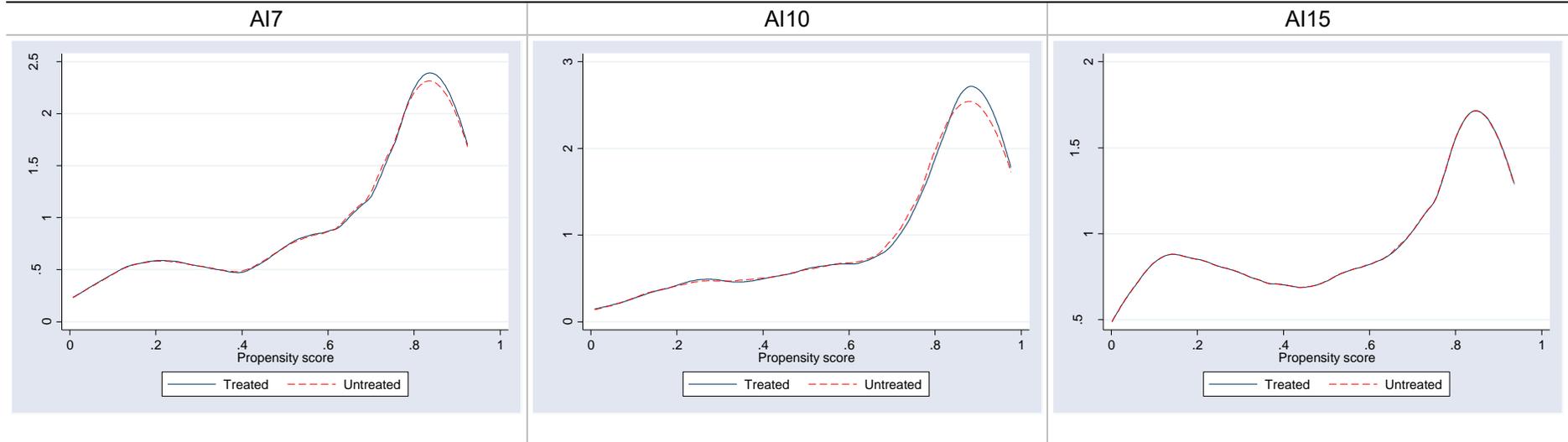
AI15



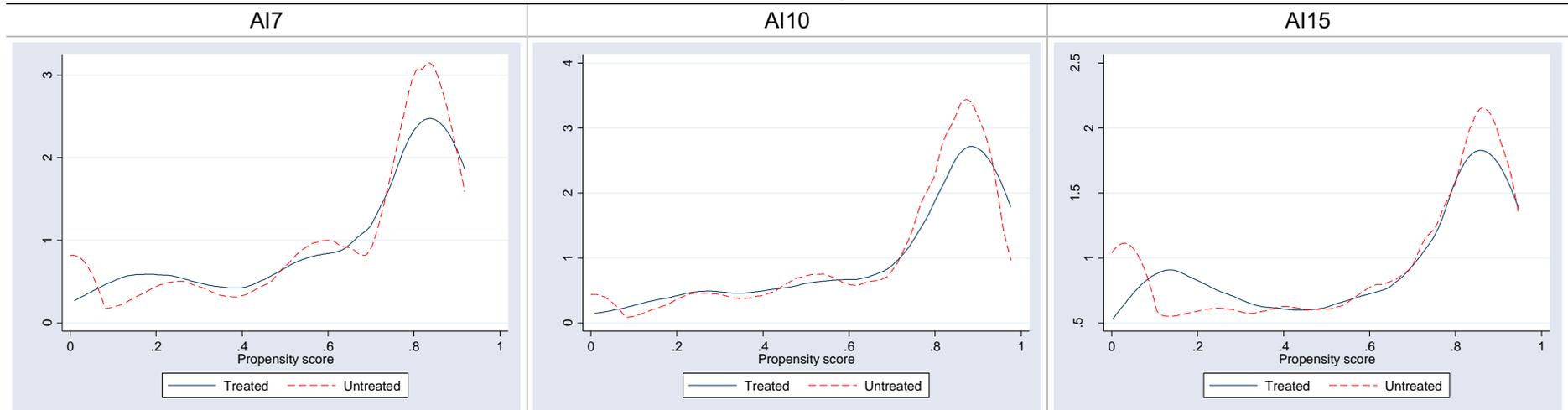
Fonte: Elaboração própria a partir de dados da PNAD 2014.

APÊNDICE L - Verificação da hipótese de sobreposição para todas as definições da amostra para o sexo feminino e métodos de pareamento, Brasil 2014.

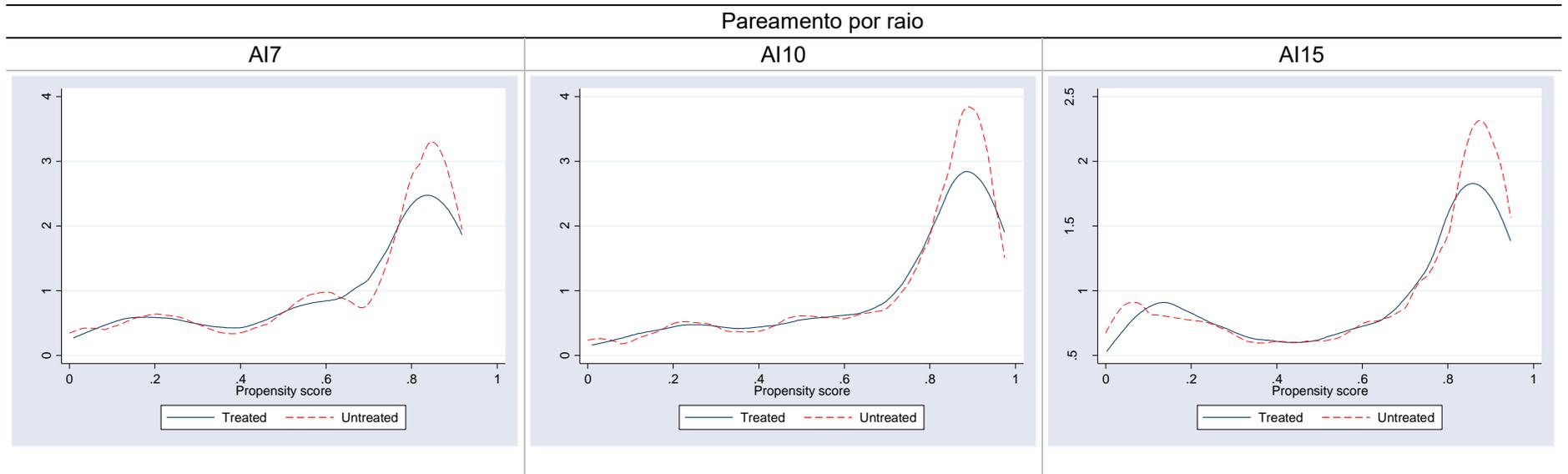
Pareamento pelos três vizinhos com reposição



Pareamento por Kernel



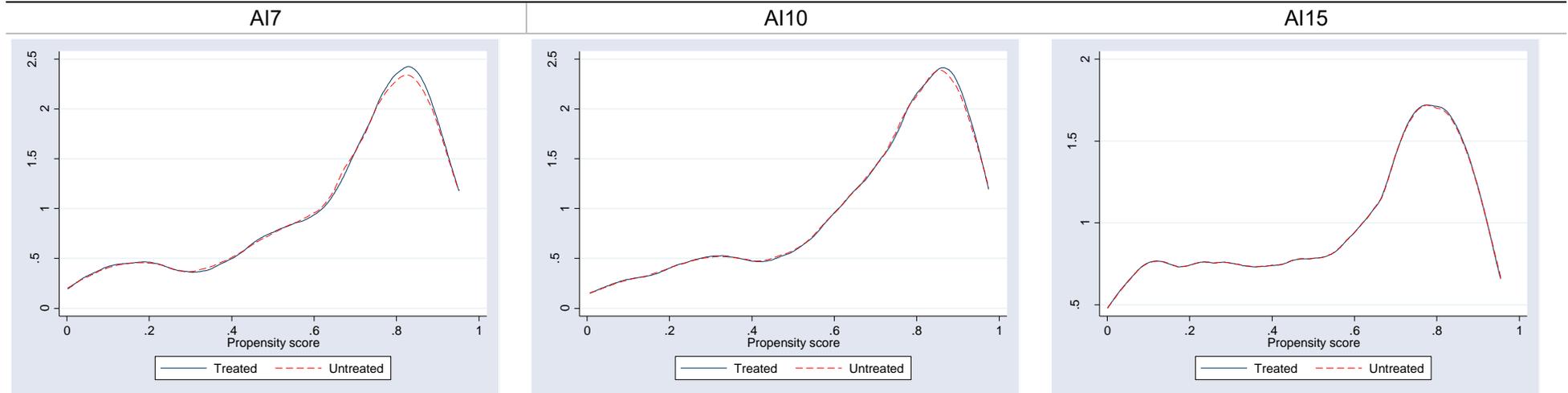
Continua...



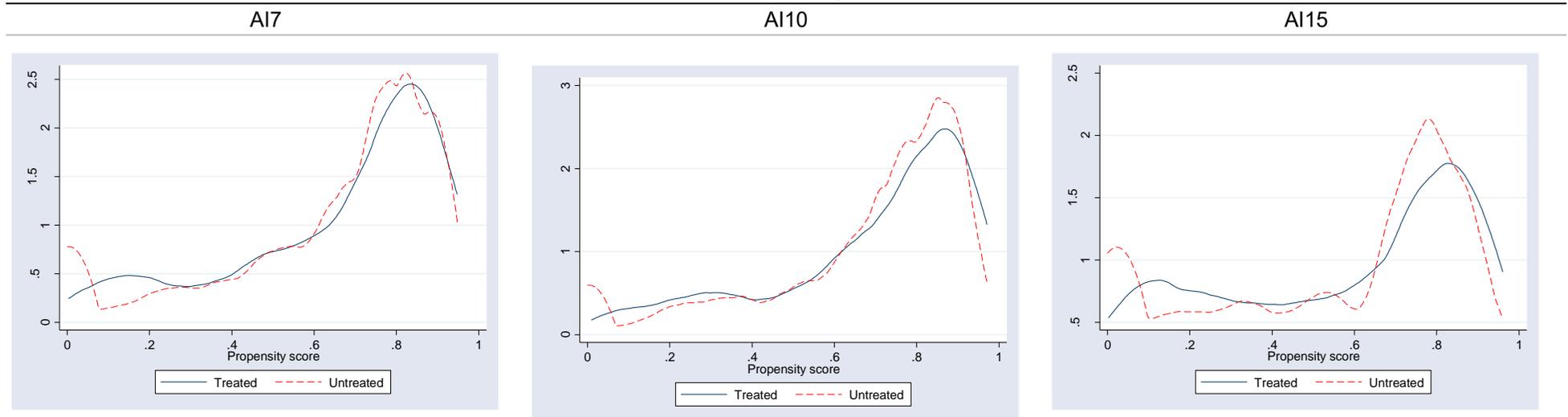
Fonte: Elaboração própria a partir de dados da PNAD 2014.

APÊNDICE M - Verificação da hipótese de sobreposição para todas as definições da amostra para o sexo masculino e métodos de pareamento, Brasil 2014.

Pareamento pelos três vizinhos com reposição



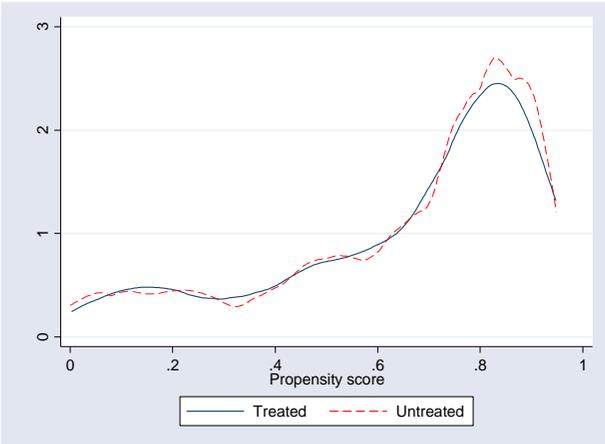
Pareamento por Kernel



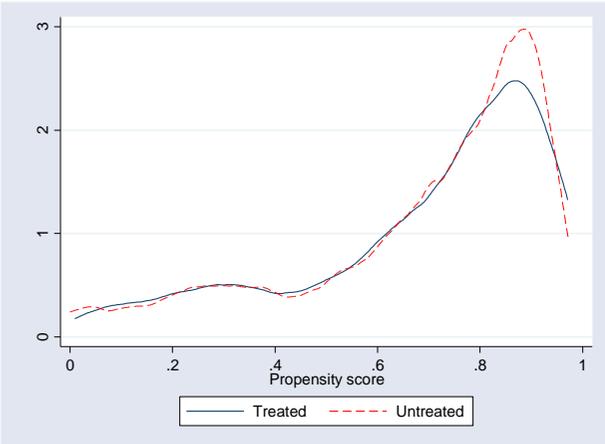
Continua...

Pareamento por raio

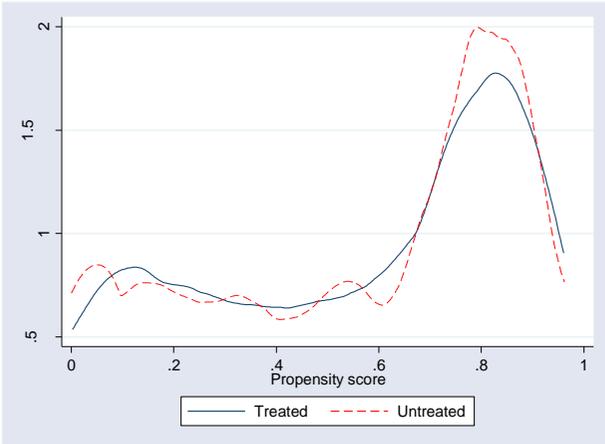
AI7



AI10



AI15



Fonte: Elaboração própria a partir de dados da PNAD 2014.

APÊNDICE N - Estimativa do efeito de tratamento médio sobre os tratados: probabilidade de conhecer o Cadastro Único para indivíduos com nível de renda domiciliar per capita até 1 salário mínimo, Brasil 2014.

Método de pareamento	AI	ANI	ATT	Erro-padrão (Bootstrap)
AI7				
Vizinho mais próximo	0,4794	0,5421	-0,06264*	0,0247416
<i>Kernel Matching</i>	0,4794	0,56155	-0,0820**	0,0225872
<i>Raius Matching</i>	0,4794	0,5528	-0,0733 **	0,0228745
AI10				
Vizinho mais próximo	0,4814	0,5720	-0,0906**	0,0297158
<i>Kernel Matching</i>	0,4814	0,5735	-0,0920 **	0,020592
<i>Raius Matching</i>	0,4814	0,5700	-0,0886 **	0,0222686
AI15				
Vizinho mais próximo	0,4883	0,5541	-0,0658*	0,032177
<i>Kernel Matching</i>	0,4883	0,5865	-0,0982 **	0,0161766
<i>Raius Matching</i>	0,4898	0,5800	-0,0902**	0,0200114

Fonte: Elaboração própria com dados da PNAD 2014. Nota: P-valor obtido pelo erro-padrão estimado por *bootstrap* com 50 repetições; **Significativo a 1% e * Significativo a 5%.

APÊNDICE O - Estimativa do efeito de tratamento médio sobre os tratados: probabilidade de conhecer o Cadastro Único para indivíduos do sexo feminino com nível de renda domiciliar per capita até 1 salário mínimo, Brasil 2014.

Método de pareamento	AI	ANI	ATT	Erro-padrão (Bootstrap)
AI7				
Vizinho mais próximo	0,4719	0,5495	-0,0776	0,0446506
<i>Kernel Matching</i>	0,4719	0,5620	-0,0901*	0,0368931
<i>Raius Matching</i>	0,4719	0,5594	-0,0875*	0,0382515
AI10				
Vizinho mais próximo	0,4700	0,5696	-0,0995 *	0,0391929
<i>Kernel Matching</i>	0,4700	0,5740	-0,1039**	0,030349
<i>Raius Matching</i>	0,4700	0,5780	-0,1080**	0,0334389
AI15				
Vizinho mais próximo	0,4878	0,5967	-0,1088**	0,0265443
<i>Kernel Matching</i>	0,4878	0,5966	-0,1088**	0,0292681
<i>Raius Matching</i>	0,4878	0,5953	-0,1074	0,0233253

Fonte: Elaboração própria com dados da PNAD 2014. Nota: P-valor obtido pelo erro-padrão estimado por *bootstrap* com 50 repetições; **Significativo a 1% e * Significativo a 5%.

APÊNDICE P - Estimativa do efeito de tratamento médio sobre os tratados: probabilidade de conhecer o Cadastro Único para indivíduos do sexo masculino com nível de renda domiciliar per capita até 1 salário mínimo, Brasil 2014.

Método de pareamento	AI	ANI	ATT	Erro-padrão (Bootstrap)
AI7				
Vizinho mais próximo	0,4881	0,5608	-0,0726	0,0445958
<i>Kernel Matching</i>	0,4881	0,5681	-0,0799*	0,0312277
<i>Raius Matching</i>	0,4881	0,5649	-0,0767	0,0316134
AI10				
Vizinho mais próximo	0,4878	0,4995	-0,0116	0,0400019
<i>Kernel Matching</i>	0,4878	0,5675	-0,0797*	0,0330575
<i>Raius Matching</i>	0,4878	0,5563	-0,0684	0,0340711
AI15				
Vizinho mais próximo	0,4939	0,5528	-0,0589	0,0340987
<i>Kernel Matching</i>	0,4939	0,5764	-0,0825**	0,0228978
<i>Raius Matching</i>	0,4939	0,5630	-0,0691	0,028147

Fonte: Elaboração própria com dados da PNAD 2014. Nota: P-valor obtido pelo erro-padrão estimado por *bootstrap* com 50 repetições; **Significativo a 1% e * Significativo a 5%.