



UNIVERSIDADE FEDERAL DO CEARÁ
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, ATUÁRIA E
CONTABILIDADE - FEAAC
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

FRANCISCA ZILANIA MARIANO

ENSAIOS SOBRE O ENSINO PROFISSIONALIZANTE NO CEARÁ

FORTALEZA
2016

FRANCISCA ZILANIA MARIANO

ENSAIOS SOBRE O ENSINO PROFISSIONALIZANTE NO CEARÁ

Tese apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Doutor em Economia.

Orientador: Prof. Dr. Ronaldo de Albuquerque e Arraes.

FORTALEZA
2016

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- M286e Mariano, Francisca Zilania.
ENSAIOS SOBRE O ENSINO PROFISSIONALIZANTE NO CEARÁ / Francisca Zilania Mariano. –
2016.
135 f. : il. color.
- Tese (doutorado) – Universidade Federal do Ceará, Faculdade de Economia, Administração, Atuária e Contabilidade, Programa de Pós-Graduação em Economia, Fortaleza, 2016.
Orientação: Prof. Dr. Ronaldo de Albuquerque e Arraes.
1. Desempenho no Ensino Profissionalizante. 2. Escola Pública e Privadas. 3. Mercado de Trabalho. 4. Dados Longitudinais. I. Título.

CDD 330

FRANCISCA ZILANIA MARIANO

ENSAIOS SOBRE O ENSINO PROFISSIONALIZANTE NO CEARÁ

Tese apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Doutor em Economia.

Aprovada em: _22__/_12__/_2016__.

BANCA EXAMINADORA

Prof. Dr. Ronaldo de Albuquerque Arraes (Orientador)
UFC/CAEN

Prof. Dr Edward Martins Costa

Prof. Dr. Guilherme Diniz Irfi

Prof. Ricardo Brito Soares

Prof. Leandro Oliveira Costa

AGRADECIMENTOS

Em primeiro lugar, quero agradecer ao meu Deus Altíssimo, pela inteligência, pela saúde, pela determinação e principalmente pela capacidade de lutar pelos meus objetivos e sonhos.

A toda a minha família que sempre foi minha base de sustentação, em especial ao meus pais Tobias e Zilmar, e meu esposo Marcio, por estarem ao meu lado sempre, em todos os momentos. Obrigada pelo apoio, amor e paciência. Ao meu irmão e cunhada, Tobias Filho e Vania. Aos meus lindos afilhados: Luis Guilherme, Elias, Valentina e Ronan Jr., obrigada por tornarem minha vida mais leve e feliz.

Agradeço muitíssimo ao professor Ronaldo que foi mais que um orientador presente, foi também um amigo e um mestre sem igual. Obrigada por todas as dicas, broncas e, principalmente pela confiança, elas foram e sempre serão importantes para a minha formação acadêmica e profissional. Tudo o que sou academicamente agradeço a ele.

Aos professores Edward, Guilherme, Ricardo e Leandro, não apenas por aceitar participar da banca, mas também por ter contribuído com seus conhecimentos para a elaboração deste estudo.

Agradeço a todos aqueles que contribuíram para a minha formação, especialmente os professores do Programa de Pós Graduação em Economia – CAEN.

Agradeço a todos os meus colegas de curso que compartilharam durante estes anos momentos de felicidades, tristezas e angustias. Com vocês essa caminhada se tornou mais leve. Agradeço especialmente aqueles mais próximos: Celina e Carol, são mais do que colegas, são amigas de verdade.

A Secretaria de Educação do Estado do Ceará, pelo apoio técnico e disponibilização de informações que foi primordial para a conclusão deste trabalho.

A Universidade Federal do Ceará, por me proporcionar o título de Doutora.

Enfim, a todas as pessoas que contribuíram de alguma forma com este momento tão importante na minha vida, mais uma vez obrigada a todos!

RESUMO

A presente tese é composta por três artigos. O primeiro, que tem como título “Avaliação do Ensino Profissionalizante com Dados Longitudinais”, visa ampliar o debate na literatura da Economia da Educação sobre o ensino público, avaliando o desempenho dos alunos de Escolas Estaduais de Ensino Profissionalizante (EEEP) nos exames do ENEM. Algumas contribuições podem ser consideradas: utilização de dados longitudinais; aplicação do algoritmo CEM e condução da avaliação através dos métodos Lasso e pós-Lasso no processo de seleção de variáveis para estimação do modelo de escore de propensão. Conclui-se que o efeito médio do tratamento se mostrou positivo e significativo em todas as áreas de conhecimento e na média geral do ENEM, destaque para Redação, cujo efeito foi de 39,9% a mais na nota para um aluno da EEEP. Ao se isolar o efeito líquido da escola, as EEEP demonstram possuir diretrizes mais eficazes em potencializar habilidades do conhecimento cognitivo relativo ao ensino regular, resultado que contribui para subsidiar futuras decisões aos formuladores de políticas educacionais. O segundo, denominado “Decomposição Quantílica Incondicional dos Diferenciais de Desempenho entre Ensino Profissionalizante e Privado”, busca contribuir com a literatura ao trazer novas evidências sobre os diferenciais de desempenho entre o ensino público e privado, analisando as diferenças nas distribuições de notas no ENEM 2014, para as cinco áreas de conhecimento e na média geral, entre alunos das escolas de ensino profissionalizante e alunos das escolas privadas. Além de propor uma nova construção para o grupo das escolas públicas, este trabalho buscou inovar ao utilizar o método de Regressão Quantílica Incondicional (RQI) proposto por Firpo *et al* (2007), o qual emprega regressões com base na função de influência recentrada (FIR) combinadas com um procedimento de reponderação inspirado em DiNardo et al (1996) e permite aplicar o método de decomposição de Oaxaca-Blinder. Dos resultados, observou-se diferenciais positivos e significativos, em todas as áreas, a favor dos alunos das escolas particulares em relação aos alunos do ensino profissional, principalmente em Redação, cujos diferenciais foram os mais elevados ao longo de toda distribuição, oscilando entre 12% e 16%, aproximadamente. Ao aplicar o método da decomposição, verificou-se que, os efeitos não observáveis são todos positivos e significativos em todos os quantis da distribuição e em todas as áreas do ENEM e são responsáveis por explicar a maior parte do diferencial das notas, com destaque para matemática, cujo efeito é crescente ao longo da curva. O efeito característico apresentou-se negativo apenas em alguns quantis da parte inferior da distribuição das notas, agindo assim, como redutor dos diferenciais de desempenho. Porém, na maioria dos casos, esse efeito foi positivo e significativo, atuando de forma a contribuir com o diferencial de notas, ou seja, os alunos das escolas privadas apresentam melhores características que contribuem para um bom desempenho, principalmente nos quantis mais elevados da distribuição. O terceiro artigo, cujo título é “Ensino Profissionalizante e Inserção no Mercado de Trabalho”, procura ampliar o debate existente na literatura da Educação sobre o ensino profissionalizante, com ênfase nos efeitos gerados no mercado de trabalho. Para tanto, busca-se trazer novas evidências sobre o papel das escolas profissionalizantes e da escolha dos cursos sobre os rendimentos e os tipos de ocupações dos jovens. A partir de uma amostra com informações longitudinais criadas pela junção das seguintes bases de dados: Censo escolar 2011 e 2012, e Relação Anual de Informações Sociais (RAIS) 2013, aplicou-se os procedimentos de Watson e Elliot (2016), os quais combinam o método da entropia, desenvolvido por

Hainmueller (2012), com o pareamento por escore de propensão – PEP, para a variável de tratamento binária, e a metodologia do pareamento por escore de propensão com múltiplos tratamentos, a qual é abordada por Rosenbaum e Rubin (1983a), Imbens (2000), Lechner (2001), para a variável de tratamento categórica. Dos resultados, observou-se que o efeito do ensino profissionalizante não foi significativo sobre a remuneração e sobre a categoria socioeconômica 1 dos alunos concludentes do ensino médio, porém, na ocupação do tipo 2, o efeito foi positivo e significativo, com diferencial de mais de 100% comparado ao grupo das escolas regulares e a partir da ocupação nível 3, os efeitos foram significativos e negativos. Considerando os tipos de cursos, verificou-se que os efeitos sobre as remunerações se mostraram positivas e significativas somente para os cursos nas áreas de saúde e gestão, cujo efeito foi 12% e 8%, respectivamente, indicando que os alunos que fizeram estes cursos durante o ensino médio apresentam rendimentos superiores no mercado de trabalho aos demais alunos. Sobre os níveis de ocupações, destacaram-se as áreas de saúde, gestão e informação, os quais possuem maiores proporções de alunos presentes nas ocupações que exigem mais escolaridade e proporcionam maiores retornos financeiros, 1 e 2, enquanto que apresentam menores proporções nas categorias 3, 4 e 5.

Palavras chave: Desempenho no Ensino Profissionalizante, Escola Pública e Privadas, Mercado de Trabalho, Dados Longitudinais.

ABSTRACT

This thesis comprises three articles. The first, titled "Evaluation of Vocational Education with Longitudinal Data", aims to broaden the debate in the literature of the Economy of Education on public education, by assessing the performance of the students of State Schools of Vocational Education (EEEE) of Ceará in the five areas of the ENEM examination of 2014. Some contributions may be considered: methodological procedures for obtaining and using longitudinal data; Application of the CEM algorithm and conduction of the evaluation through the Lasso and post-Lasso methods in the process of variable selection for estimation of the propensity score matching (PSM) model. It is concluded that the average treatment effect is positive and significant in each area of the exam and in the general average of the ENEM, particularly in Writing, whose effect is 39.9% higher in the grade of an EEEP student. By isolating the school net effect, EEEP demonstrate to possess more effective guidelines in potentializing cognitive knowledge abilities relative to regular schools, under the hypothesis that the variables not considered in the analysis are equivalent between the two types of school. These results contribute to supporting future decisions of educational policy makers. The second article, "Unconditional Quantile Decomposition of Performance Differences between Vocational and Private Education," embodies new evidences on performance differentials between students of public vocational and private schools in the distribution of grades in the ENEM examination. In addition to proposing a new construction for the group of public schools, the empirical application is distinguished by the use of the Unconditional Quantile Regression method based on the recentered influence function (RIF) with a reweighting procedure, besides allowing the application of the Oaxaca-Blinder decomposition. Positive and significant differences are found in all areas of the exam in favor of private schools' students, especially in Writing, whose differentials are the highest throughout the distribution, ranging from 12% to 16%. By decomposing the effects, it is found that the unobservable ones are positive and significant in all the quantiles of the distribution and in all areas of the ENEM, besides explaining most of the differentials of the grades, especially in mathematics, whose effect grows with the quantile of the distribution (glassceiling). Since the characteristic effect turns out to be negative only in the bottom of the grade distribution, it is concluded that private schools' students present better characteristics that contribute to explain their higher grades in the exam. The third article, "Vocational Education and Labor Market Insertion", contributes to the scarce debate in the Brazilian literature about the connection between vocational vis a vis regular education and the labor market. In this sense, new evidences are drawn about the youngsters' choice for specific specializations available in the EEEP and their positions in the labor market. Based upon a sample with longitudinal information created by the junction of three databases - School Census/2011/2012 and The Annual Relation of Social Information (RAIS/2013) - the methodology that combines the entropy method with PSM

procedure is applied for the binary treatment variable, and the PSM with multiple treatments for the categorical treatment variable. In order to allow for statistical treatment, due to the excessive amount of 596 occupations listed in the Brazilian Classification of Occupations (CBO, 2002), the Jannuzzi procedure (2001, 2003) is applied to regroup them in only five socioeconomic occupational categories. It is concluded that the effect of vocational training is not significant on the earnings of high school students specialized in occupational category of Health. On the other hand, the effects are positive and significant in Management occupational category, with a differential of more than 100% compared to the group of regular schools, while the effects in the category of Information are significant and negative. It is also verified that the effects on earnings were positive and significant only for students who took courses in Health and Management, with differentials of 12% and 8%, respectively, over the others. It is also worth noting the higher proportion of students present in the two occupations that require more qualified training relative to the other three.

Key words: Vocational Education, Quality of Education, Public and Private Schools, Labor Market, Longitudinal Data.

LISTA DE TABELAS

Tabela 1.1 - Estatísticas descritivas das variáveis da amostra – 2014.....	43
Tabela 1.2 - Resultados do algoritmo CEM.....	44
Tabela 1.3 - Efeito médio do tratamento nas áreas de conhecimento do ENEM, por kernel (0,01) .	46
Tabela A.1.1 - Estimação do Propensity Score pós Lasso	55
Tabela A.1.2 - Análise de sensibilidade (Rosenbaum Bounds) para as áreas de conhecimento e a média geral do ENEM.....	55
Tabela 2.1 - Estatísticas Descritivas das variáveis da amostra – 2014.....	67
Tabela 2.2 - Decomposição agregada na distribuição das notas nas áreas do ENEM	75
Tabela 2.3 - Decomposição dos efeitos característico e retorno - MG	78
Tabela B.2.0.1 - Estimativas da Regressão Quantílica Incondicional por Escola – MG	86
Tabela B.2.0.2 - Estimativas da Regressão Quantílica Incondicional por Escola - CH.....	87
Tabela B.2.0.3 - Estimativas da Regressão Quantílica Incondicional por Escola - CN.....	88
Tabela B.2.0.4 - Estimativas da Regressão Quantílica Incondicional por Escola - LC	89
Tabela B.2.0.5 - Estimativas da Regressão Quantílica Incondicional por Escola - MT	90
Tabela B.2.0.6 - Estimativas da Regressão Quantílica Incondicional por Escola - R.....	91
Tabela 3.1 - Estatísticas descritivas das variáveis da amostra por tipos de escolas e por ano.	107
Tabela 3.2 - Estatística descritiva das variáveis de resultados por cursos, 2013.....	107
Tabela 3.3 - Condições de Momentos das covariadas antes do balanceamento por entropia	110
Tabela 3.4 - Condições de Momentos das covariadas após o balanceamento por entropia.....	110
Tabela 3.5 - Impacto das EEEP sobre a Remuneração e Ocupações dos Concludentes no Mercado de Trabalho – 2013	112
Tabela 3.6 - Análise de Sensibilidade (Rosenbaum bounds)	113
Tabela 3.7 - Resultado do Modelo Multinomial Logit – Múltiplos Tratamentos	115
Tabela 3.8 - Impacto dos Múltiplos Tratamentos sobre as Variáveis de resultado – 2013	118
Tabela 3.9 - Análise de Sensibilidade - Remuneração.....	120
Tabela 3.10 - Análise de Sensibilidade - Ocupações	121

LISTAS DE QUADROS

Quadro 2.1 -Descrições das Variáveis.....	66
Quadro 3.1 - Variáveis na Amostra	104

LISTA DE FIGURAS

Figura 1 - Construção dos grupos de tratamento e controle.....	53
--	----

LISTA DE GRÁFICOS

Gráfico A.1.1 - Sobreposição do propensity Score pelo vizinho mais próximo	54
Gráfico A.1.2 - Sobreposição do propensity Score pelo kernel	54
Gráfico A.1.3 - Sobreposição do propensity Score pelo Radius	54
Gráfico 2.1 - Decomposição dos Efeitos – MG.....	72
Gráfico 2.2 - Decomposição dos Efeitos – CH	72
Gráfico 2.3 - Decomposição dos Efeitos – CN	73
Gráfico 2.4 - Decomposição dos Efeitos – MT	73
Gráfico 2.5 - Decomposição dos Efeitos – LC.....	73
Gráfico 2.6 - Decomposição dos Efeitos – R	73
Gráfico A.2.1 -Distribuição das notas em Ciências Naturais	85
Gráfico A.2.2 -Distribuição das notas em Linguagens e Códigos.....	85
Gráfico A.2.3 - Distribuição das notas em Ciências Humanas	85
Gráfico A.2.4 - Distribuição das notas em Matemática	85
Gráfico A.2.5 - Distribuição das notas em Redação	85
Gráfico A.2.6 - Distribuição das notas em Média Geral	85
Gráfico 3.1 - Escore de Propensão antes do balanceamento por entropia.....	111
Gráfico 3.2 - Escore de Propensão após o balanceamento por entropia.....	111
Gráfico 3.3 - Sobreposição do Escore de Propensão - Saúde.....	114
Gráfico 3.4 - Sobreposição do Escore de Propensão - Gestão.....	114
Gráfico 3.5 - Sobreposição do Escore de Propensão - Informação.....	115
Gráfico 3.6 - Sobreposição do Escore de Propensão – Hospitalidade e Lazer.....	115
Gráfico 3.7 - Sobreposição do Escore de Propensão – Outras áreas.....	115

SUMÁRIO

INTRODUÇÃO GERAL	16
CAPÍTULO 1	19
AVALIAÇÃO DO ENSINO PROFISSIONALIZANTE COM DADOS LONGITUDINAIS	19
Introdução	19
2. Ensino Profissionalizante.....	23
2.1 Escolas Estaduais de Ensino Profissionalizante no Ceará	26
3 Metodologia	28
3.1 Métodos de Avaliação de Impacto	28
3.1.1 Pareamento por Escore de Propensão (PEP).....	29
3.2 Estimação do Escore de Propensão via seleção robusta de variáveis de controles. 32	
3.2.1 Método de seleção de variável por meio dos métodos Lasso e Pós-Lasso	33
3.3 Algoritmo CEM.....	36
3.4 Análise de Sensibilidade	37
3.5 Variáveis e Bases de Dados	38
4 Resultados	40
4.1 Análise Descritiva	40
4.2 Resultados do Pré-Pareamento.....	44
4.3 Seleção das Variáveis de Controle: Método do Lasso	44
4.4 Estimação do pareamento com escore de propensão	45
Considerações Finais	46
Referências Bibliográficas	49
Apêndice	53
CAPÍTULO 2	56
DECOMPOSIÇÃO QUANTÍLICA INCONDICIONAL DOS DIFERENCIAIS DE DESEMPENHO ENTRE ENSINO PROFISSIONALIZANTE E PRIVADO	56
Introdução	56
2 Método de Decomposição	58
2.2.1 Variáveis do Modelo	65

2.2.2 Base de Dados	65
3 Resultados	67
3.1 Análise descritiva	67
3.2 Regressão FIR nos Quantis Incondicionais.....	69
3.3 Decomposição da regressão FIR: Efeito Composição e Estrutural.....	71
Conclusões	80
Referências.....	82
Apêndice A	84
Apêndice B	86
CAPÍTULO 3	92
ENSINO PROFISSIONALIZANTE E INSERÇÃO NO MERCADO DE TRABALHO... 92	
Introdução	92
2 Metodologia	96
2.1 Balanceamento por Entropia	96
2.2 PEP com Múltiplos Tratamentos	99
2.3 Variáveis e Bases de Dados	101
3 Resultados	104
3.1 Análise Descritiva	104
3.2 Resultados dos modelos.....	108
3.2.1 Resultados do Modelo Balanceamento por Entropia	108
3.2.2 Resultados do Modelo com múltiplos tratamentos	114
Conclusões	122
Referências.....	125
CONCLUSÃO GERAL	127
REFERÊNCIAS	130

INTRODUÇÃO GERAL

Embora seja consenso que possuir uma boa educação básica seja a forma mais eficaz para se alcançar o desenvolvimento, esta ainda se apresenta de baixa qualidade e com alguns entraves, principalmente nos países em desenvolvimento, por exemplo, o Brasil. Muito tem-se discutido em formas de melhorá-la nas diferentes fases que a compõe, a saber, ensino fundamental I e II, e ensino médio. Esta última etapa é considerada uma das mais problemáticas, apresentando altos índices de evasão, repetência e baixo desempenho em testes padronizados, principalmente comparada ao ensino privado. Além disso, essa fase é marcada pela transição para o ensino superior e/ou para o mercado de trabalho, o que requer bastante atenção por partes dos gestores e políticos para que se possam desenvolver um ensino médio de qualidade e atrativo aos jovens, e que os capacitem para a escolha que desejarem tomar. Porém, muitos jovens não se sentem preparados a entrarem em alguma universidade e nem de adquirem um emprego de qualidade, levando muitas vezes, à condição de nem estudar e nem trabalhar, e isso eleva os índices de jovens ociosos em situação de vulnerabilidade social, contribuindo assim, para o ciclo da pobreza e desigualdade socioeconômica.

Segundo Leme e Escardíbul (2016), existem divergências entre países quando se trata do tipo de ensino que os jovens devem ter durante essa etapa. Portugal, Espanha e Suécia possuem um currículo especializado, onde os jovens escolhem uma especialização e a seguem durante o ensino secundário. Por outro lado, países como Finlândia, Dinamarca e EUA seguem um currículo geral, onde os alunos, apesar de serem capazes de escolher entre as diferentes classes em áreas distintas, não são obrigados a seguir uma única especialização e, portanto, recebem uma educação mais geral. A partir da aplicação de um modelo de diferença em diferença, esses autores procuraram estimar os possíveis efeitos dessas diferenças institucionais que influenciam os resultados dos alunos, usando dados do PISA para 20 países diferentes. Os resultados sugerem que a escolha de um sistema ou outro não

explica as diferenças entre os países, seja no desempenho médio ou na diferença dos resultados dos testes dos alunos.

No Brasil, dentre os tipos de políticas pública que buscam melhorar a qualidade de ensino e reduzir a quantidade de crianças e adolescentes ociosos, encontra-se o ensino integral, o qual aumenta a carga horária dos alunos nas escolas e fornecem outras atividades que sejam atrativas a permanências deles. Além desse tipo de ensino, existe o ensino profissionalizante, que busca preparar os jovens para a inserção no mercado de trabalho, através de cursos técnicos voltados para diversas áreas.

As Escolas de Educação de Ensino Profissional (EEEP), criadas em 2008, no estado do Ceará, levou em consideração esses dois tipos de ensino, integral e profissionalizante, e proporciona ao jovem, durante o ensino médio, cursar as disciplinas básicas¹ dessa etapa, e participarem de cursos técnicos e profissionalizantes voltados para diversas áreas (saúde, agricultura, computação, finanças, etc.). Essa ação possibilitaria ao aluno concluir o ensino médio e se profissionalizar para o mercado de trabalho. Com isso, além de formar o jovem para o mercado de trabalho, o ensino das escolas profissionalizantes também o capacita a ingressar em universidades.

Nesse contexto, essa tese está estruturada em três capítulos que abordam assuntos distintos relacionados ao ensino profissionalizante, focados nas EEEP do Ceará.

No primeiro capítulo, busca-se avaliar o desempenho dos alunos das EEEP nos exames do ENEM. Para a mensuração do efeito dessas escolas, algumas contribuições podem ser consideradas. Primeira, utilização de dados longitudinais, criados a partir do cruzamento das bases, SPAECE 2011, 2012, Censo Escolar 2011 e ENEM 2014. Segunda, utilização do algoritmo CEM (*Coarsened Exact Matching*) para isolar os efeitos dos alunos de melhor desempenho antes de ingressarem nas escolas profissionalizantes. Terceira, condução da avaliação a partir dos métodos Lasso e pós-Lasso, recentemente desenvolvidos por Belloni et al (2015), no processo de seleção de variáveis para aplicação do modelo de score de propensão. Feito o pareamento pelo método de Kernel, será estimado o efeito médio do tratamento.

¹ Língua Portuguesa, Artes, Inglês, Espanhol, Educação Física, História, Geografia, Filosofia, Sociologia, Matemática, Biologia, Física e Química.

No segundo capítulo, busca-se trazer novas evidências sobre os diferenciais de desempenho entre o ensino público e privado, analisando as diferenças nas distribuições de notas no ENEM 2014, para as cinco áreas de conhecimento e na média geral, entre alunos das escolas de ensino profissionalizante e alunos das escolas privadas. Além de propor uma nova construção para o grupo das escolas públicas, este trabalho buscou inovar ao utilizar o método de Regressão Quantílica Incondicional (RQI) proposto por Firpo *et al* (2007), o qual emprega regressões com base na função de influência recentrada (FIR) combinadas com um procedimento de reponderação baseado em DiNardo *et al* (1996) e permite aplicar o método de decomposição de Oaxaca-Blinder.

Por fim, no terceiro capítulo, pretende-se verificar o efeito do ensino profissionalizante sobre os rendimentos e os tipos de ocupações dos jovens no mercado de trabalho. Uma das contribuições deste estudo reside na construção da base de dados para mensurar este efeito, pois através da construção de identificadores foi possível gerar uma amostra com informações longitudinais cruzando diferentes bases, a saber, Censo escolar 2011 e 2012, e Relação Anual de Informações Sociais (RAIS) 2013. A abordagem metodológica seguiu o artigo de Watson e Elliot (2016), os quais combinam o método da entropia, desenvolvido por Hainmueller (2012) com o pareamento por score de propensão - PEP, visando assim, equilibrar com maior robustez as covariadas. Para aplicação deste método, será utilizada uma variável binária para o tratamento, o qual irá permitir encontrar o efeito tratamento médio sobre os tratados das escolas profissionalizantes sobre a remuneração e as ocupações dos concludentes do ensino médio em 2011 e 2012 no mercado de trabalho em 2013. Além disso, visando encontrar o efeito do tipo de curso escolhido sobre a remuneração e os tipos de ocupações, este estudo optou por considerar uma nova variável de tratamento, o qual irá assumir múltiplos valores a partir dos tipos de cursos presentes na amostra. Em decorrência dessa variável ser agora definida como categórica, será utilizado a metodologia do pareamento por score de propensão com múltiplos tratamentos, a qual é abordada por Rosenbaum e Rubin (1983a), Imbens (2000), Lechner (2001).

CAPÍTULO 1

AVALIAÇÃO DO ENSINO PROFISSIONALIZANTE COM DADOS LONGITUDINAIS

Introdução

Por ser considerada uma das principais forças motrizes para impulsionar o crescimento e desenvolvimento econômico, a educação tem sido mantida na agenda de políticas públicas como ação prioritária para melhorar sua qualidade, principalmente em países cujo desempenho relativo a outros, aferido por instituições internacionais, seja considerado baixo, a exemplo do Brasil. Segundo Ganimian e Rocha (2011), Hanushek e Woessmann (2012) e Levy e Schady (2013), a magnitude econômica de um país segue sua grandeza educacional, embora regiões como a América Latina apresentem índices educacionais inferiores a outras pelo que seus níveis de renda poderiam prever. Daí, reconhece-se que a diferença entre países com altas e baixas rendas depende essencialmente de quão rápido eles crescem no longo prazo. Para tanto, há que se enfatizar, em análise comparativa de países, a necessidade de se medir níveis educacionais em termos qualitativos a partir de escores em exames internacionais (BARRO, 2013). Dentre as várias alternativas de estruturas pedagógicas que visam a melhoria da qualidade educacional, o presente artigo avalia a distinção na geração, por meio de desempenho, da qualidade estudantil entre o ensino profissionalizante e o tradicional.

De acordo com diretrizes do Relatório da OECD (2016) referente ao baixo desempenho educacional, os gestores de política deveriam priorizar a educação e traduzir essa prioridade em recursos adicionais, embora reconheça que lidar com esse problema exige uma abordagem multidimensional adaptada às circunstâncias nacionais e locais; ou seja, organizar as escolas e sistemas de ensino para que possam oferecer oportunidades de educação infantil para todos; identificar alunos e escolas de baixa performance e intervir com políticas adequadas. Dessa forma, reduzir o número de alunos de baixo desempenho não é apenas um objetivo em si mesmo, mas uma forma eficaz de melhorar o desempenho geral de um sistema de educação, assim, compreender os fatores que contribuem para esse problema é apenas o

primeiro passo, pois deve ser seguido imediatamente pela implementação de políticas de forma a melhorar a educação.

Embora o Brasil tenha apresentado progresso na educação básica, Relatório Nacional do PISA-2012, e maior avanço absoluto na proficiência em Matemática quando comparados os exames entre 2003 e 2012, destacando-se o ensino médio, a educação brasileira ainda encontra-se em um patamar distante do ambicionado pela sociedade, que destaca a educação como o alicerce mais estável da competitividade econômica e da superação das desigualdades sociais e regionais. Problema torna-se mais agravante quando se compara o desempenho dos alunos pertencentes a rede pública² de ensino com a rede privada, cujos alunos pertencentes a este tipo de ensino apresentam média em matemática superior em aproximadamente 16% média dos alunos das redes federal, estadual e municipal (RELATORIO NACIONAL DO PISA, 2012).

Moraes e Beluzzo (2014) afirmam que o diferencial de desempenho escolar entre escolas públicas e privadas é favorável às escolas particulares, ainda que esse efeito seja diminuído quando controlam por *peer effects*. A análise desse diferencial ao longo da distribuição revela que as escolas privadas tem melhor desempenho em todos os quantis, e a maior diferença ocorre em torno do centro da distribuição, no qual o desempenho das escolas privadas é aproximadamente 1,5 desvios padrão melhor do que o das escolas públicas.

Esse desempenho reflete no ingresso de estudantes nas Instituições de Ensino Superior público no país, pois apesar de ser provido publicamente, não é necessariamente consumido pela população mais pobre ou de condição socioeconômica mais vulnerável, devido ao processo seletivo restritivo a que os estudantes se submetem para conseguir uma vaga em uma instituição pública. Esses processos seletivos para ingresso na universidade pública, como vestibulares e atualmente o Exame Nacional do Ensino Médio (ENEM), dado que são baseados em critérios meritocráticos, estão usualmente associados ao capital humano ou habilidade do indivíduo, assim, são selecionados aqueles alunos que possuem melhor desempenho (BACALHAU; MATOS, 2013).

² Com exceção das escolas militares e federais.

Dentre os tipos de políticas voltados para a melhoria da educação pública no ensino médio encontra-se a concepção do ensino profissionalizante. Embora a criação desse tipo de ensino tenha sido guiada para atender jovens que viviam à margem da sociedade (TAVARES, 2012), após a criação da primeira Lei de Diretrizes e Bases da Educação Nacional, ele passou a ser constituído juntamente com o nível médio, e acrescentou o foco no ensino superior como um de seus propósitos. Dessa forma, estudantes de escolas profissionais não seriam prejudicados, pois, além de estarem preparados para o mercado de trabalho, também poderiam ingressar numa universidade.

No Brasil, Araújo et al (2014) encontraram uma relação positiva entre participar de ensino técnico e profissionalizante e desempenho escolar no ENEM, além de favorecer a entrada no mercado de trabalho, porém, Polidano e Tabasso (2016) observaram que na Austrália esse tipo de combinação de ensino reduz as chances de o estudante ingressar em uma universidade.

A integração das duas redes de ensino, profissional e geral, pode-se romper a dualidade estrutural entre formação para o trabalho e preparação para a universidade (CASTRO; TIEZZI, 2005; MARTINS, 2012; MOEHLECKE, 2012). Diversos autores³ buscam verificar os efeitos desse tipo de ensino sobre a inserção no mercado de trabalho, todavia, ainda, são poucos os estudos que avaliam os impactos do ensino profissionalizante sobre a educação pública brasileira, destaques para Araújo et al (2014), o qual avaliou, com base em informações do ENEM 2009, se os alunos que realizam o currículo específico da Educação Profissional e Tecnológica (EPT) apresentam aumento na proficiência em disciplinas básicas, e para Mariano et al (2015), que elaboraram uma avaliação sobre esse tipo de política e encontraram o efeito do ensino das Escolas Estaduais de Ensino Profissionalizantes sobre o desempenho no ENEM 2013, bem como o impacto desse sobre as chances de os alunos estarem trabalhando. Estes trabalhos utilizam dados em cortes transversais, os quais não permitem o acompanhamento dos alunos durante o processo de maturação desse tipo de ensino, lacuna esta a ser preenchida com o presente estudo de caso, por meio da aplicação em dados longitudinais.

³ NEUMAN; ZIDERMAN 1989; CHEN; WEKO, 2009; SEVERNINI; ORELLANO, 2010; ASSUNÇÃO; GONZAGA, 2010; ARAÚJO et al, 2014; MARIANO et al, 2015

O projeto de criação de Escolas Estaduais de Educação Profissional (EEEP) do Ceará visava, além do ensino com as disciplinas básicas do currículo do ensino médio, os cursos técnicos e profissionalizantes voltados para diversas áreas (saúde, agricultura, computação, finanças, etc.), cuja escolha atenderia a realidade local onde a instituição seria implantada. Essa ação possibilitaria ao aluno concluir o ensino médio e se profissionalizar para o mercado de trabalho, bem como capacitá-lo a ingressar em universidades. Silva (2013) observa o desempenho das escolas estaduais no Exame Nacional do Ensino Médio (ENEM), a partir da comparação entre as médias escolares nos anos 2009 a 2011 e constatou que o estado do Ceará melhorou a performance com a participação dos alunos das EEEP. Além disso, segundo informações disponibilizadas pelo INEP (2015) referentes a nota média por escola no ENEM/2014, 80% das melhores escolas públicas do Ceará são de ensino profissional.

Com isso, este trabalho procura ampliar o debate existente na literatura da Educação sobre o ensino profissionalizante, com ênfase nos efeitos gerados no desempenho dos alunos no ENEM. Para tanto, busca-se trazer novas evidências acerca dessa linha pedagógica nas escolas profissionais públicas.

Uma das contribuições deste estudo reside na base de dados utilizada para mensurar este efeito, pois, por meio da construção de identificadores, foi possível gerar uma amostra com informações longitudinais procedendo-se o cruzamento de diferentes bases – SPAECE 2011, 2012; Censo Escolar 2011 e ENEM 2014 – até então desconhecida em outros trabalhos pelos autores deste⁴.

Dessa forma, pôde-se observar o aluno no 9º ano do ensino fundamental e acompanhá-lo durante o ensino médio, verificando assim, a trajetória destes nas escolas e permitindo a criação dos grupos de tratados, alunos que ingressaram o ensino médio nas escolas profissionalizantes em 2012 e estavam cursando o 3º ano em 2014 nessas escolas, e controles, mesma definição, porém para as escolas regulares. E através das informações desses alunos em 2011, período anterior ao ensino médio, este trabalho poderá observar os fatores que possivelmente determinaram a entrada destes nas escolas profissionalizantes em

⁴ As dificuldades e esforços dispendidos para a geração final da amostra a ser utilizada ditaram a impossibilidade de expandi-la com a inclusão de outros estados.

2012, tais como, características pessoais, *background* educacional, escolaridade dos pais e características de infraestrutura das escolas.

E, objetivando isolar o efeito-aluno⁵, será aplicado o método de pareamento CEM (*Coarsened Exact Matching*) a partir das notas de matemática e português no SPAECE 2011. Além disso, este trabalho também contribui com a literatura para o Brasil na escolha da opção metodológica para a condução dessa avaliação, pois baseou-se em um modelo recentemente desenvolvido por Belloni et al (2015) e, ainda não utilizado em dados brasileiros, o qual aplica a técnica de seleção do lasso e do pós-lasso a um conjunto de variáveis explicativas, selecionando aquelas consideradas importantes na determinação do tratamento. Após esses procedimentos, será encontrado o Efeito Médio de Tratamento sobre os tratados (EMTT).

Em sequência, o artigo está organizado com as seguintes seções: aspectos teóricos da literatura sobre as escolas profissionalizante e as EEEP no estado do Ceará; a abordagem metodológica; resultados e conclusões.

2. Ensino Profissionalizante

Para Crouch et al (1999), existem três justificativas para o desenvolvimento de políticas públicas voltadas para a qualificação mais especializada da mão de obra, em oposição a educação formal. Primeira, melhores condições de um país participar de um mercado internacional competitivo; segunda, estas políticas tendem a contribuir para a redução do desemprego, principalmente entre os jovens; terceira, a qualificação da mão de obra permitiria que as economias de países em desenvolvimento evoluíssem de uma situação equilíbrio gerado pela produção de commodities de valor agregado inferior, resultado da utilização de mão de obra barata pouco qualificada e de baixa produtividade, para situações de equilíbrio com a produção de mercadorias de maior valor agregado, proporcionando assim, melhor distribuição da renda.

Dentre as políticas públicas voltadas para a educação, encontra-se o ensino profissionalizante, porém, apesar da importância, ainda é um assunto pouco explorado na

⁵ Conhecimento e capacidades adquiridos pelos alunos no período anterior a entrada nas escolas profissionalizantes.

literatura. Alguns autores buscam verificar os fatores históricos (KUENZER, 2005; FOLEY, 2007; TAVARES, 2012; MARTINS, 2012), os efeitos sobre a inserção no mercado de trabalho (NEUMAN; ZIDERMAN 1989; CHEN; WEKO, 2009; SEVERNINI; ORELLANO, 2010; ASSUNÇÃO; GONZAGA, 2010; ARAÚJO et al, 2014) e sobre o desempenho em exames para ingressar em universidades (ARAÚJO et al, 2014; MARIANO et al, 2015).

Dentre os que podem ser observados, existem limitações relacionados aos dados, pois utilizam dados em cross-section, dos quais não permitem o acompanhamento dos alunos no processo de adaptação e incorporação desse tipo de ensino.

Segundo Kuenzer (2005), existia uma dualidade estrutural na educação que se caracterizava pela existência de tipos diferentes de escola para classes sociais distintas, onde a trajetória escolar de estudantes situados nas classes de renda média e alta se constituía de uma educação básica voltada para cursos universitários, enquanto que para os demais restava como alternativa uma base educacional associada à formação em cursos técnicos.

A maior participação de jovens com nível socioeconômico mais baixo no ensino profissionalizante foi comprovada por Foley (2007), na Austrália, o qual objetivava estudar o perfil dos participantes do sistema de ensino e formação profissional. Para tanto, o autor utilizou dados sobre os estudantes matriculados na rede pública em 2001 e constatou que os alunos de origens socioeconômicas mais elevadas são maioria na participação dos setores de ensino superior e secundário, o que não ocorre no caso do setor de ensino profissionalizante.

Nesse esforço de preparação para o mercado de trabalho, Severnini e Orellano (2010) questionou se os programas de qualificação profissional realmente contribuem para os treinados obterem uma melhoria de bem-estar e se esse ensino contribui para aumentar a probabilidade de inserção do indivíduo no mercado de trabalho brasileiro. Para tanto, os autores investigaram se este tipo de ensino aumentou a probabilidade de inserção no mercado de trabalho e a elevação de renda dos egressos até meados da década de 1990, comparativamente aos que não cursaram esse tipo de ensino. Através de microdados da Pesquisa sobre Padrões de Vida (PPV) de 1996 do IBGE encontraram que os egressos de cursos profissionalizantes de nível básico tinham renda esperada 37% maior que a de indivíduos que não fizeram esse tipo de curso no ensino fundamental. Por outro lado, para os

egressos do ensino profissional de nível tecnológico, observa-se uma redução de 27% da renda esperada, comparativamente aos que não participaram desse tipo de curso no ensino superior.

Assunção e Gonzaga (2010) apresentam os resultados de uma análise realizada com base nos microdados do suplemento especial sobre educação profissional da PNAD/2007, cujo objetivo reside em analisar a inserção da população brasileira nos cursos de educação profissional e verificar o impacto da educação profissional sobre o rendimento dos trabalhadores brasileiros. Dos resultados, observaram que a inserção da educação profissional em famílias com renda *per capita* inferior a dois salários mínimos é bem menor que nas demais faixas de renda e que a educação profissional aumenta a produtividade dos trabalhadores. Dessa forma, os autores sugerem que essa educação mereça atenção no desenho de políticas públicas.

Além de fazer referência à “dualidade” existente no ensino, Tavares (2012) também diagnosticou que, depois da criação da primeira Lei de Diretrizes e Bases da Educação Nacional, isso viria a mudar, uma vez que esta passou a constituir o ensino profissionalizante juntamente com o ensino médio e estabeleceu que este tivesse como um de seus propósitos o foco no ensino superior. Logo, já se poderia esperar que os jovens participantes do ensino profissionalizante não seriam prejudicados, pois além de estarem preparados para o mercado de trabalho também podem ingressar em uma universidade. Além disso, a escola de ensino profissionalizante também é colocada como um meio de promover as habilidades pessoais do estudante, a vivência com as outras pessoas, a inserção no mercado de trabalho e o melhor desempenho nas disciplinas regulares (ARAÚJO et al, 2014). Segundo Martins (2012), pode-se romper a dualidade estrutural entre formação para o trabalho e preparação para a universidade a partir da integração das duas redes de ensino, profissional e geral.

Buscando analisar o desempenho escolar e inserção no mercado de trabalho dos alunos das escolas técnicas profissionalizantes do Brasil, Araújo et al (2014) usaram os resultados do ENEM-2009 e, através de um modelo de pareamento com escore de propensão para escolas federais e privadas, encontraram uma relação positiva entre participar de ensino técnico e profissionalizante e desempenho escolar, além de favorecer a entrada no mercado de trabalho. Mariano et al (2015) encontraram resultados semelhantes para o estado do Ceará,

observaram que os alunos das escolas profissionalizantes apresentam diferenciais significativos em todas as áreas de conhecimento do ENEM 2013 e possuem maiores chances de ultrapassarem as notas de cortes referentes aos cursos da UFC, porém estes possuem menor probabilidade de estar trabalhando no período que prestaram o exame, em relação aos alunos das escolas regulares.

2.1 Escolas Estaduais de Ensino Profissionalizante no Ceará

No Ceará, a oferta de cursos técnicos era realizada pelo Instituto Centro de Ensino Tecnológico – CENTEC, pelo Instituto Federal do Ceará (IFCE) e pelo Sistema S, onde havia uma oferta de cento e quinze cursos, nos quais estavam matriculados 14.606 jovens e adultos, segundo dados da Secretaria da Ciência e Tecnologia do Ceará - SECITECE, em 2005, representando apenas três e meio por cento dos alunos matriculados no ensino médio, portanto, uma oferta insignificante, considerando o percentual de atendimento da rede de educação no ensino médio (NIBON, 2014). Dessa forma, tinha-se, assim, um cenário estadual que evidenciava a necessidade de desenvolvimento de uma política de educação profissional que potencializasse a articulação entre o ensino e o mercado de trabalho e, ao mesmo tempo, ampliasse as possibilidades de qualificação do processo de aprendizagem (RELATÓRIO DE GESTÃO, 2014).

Em 2007, a SEDUC aderiu ao Programa Brasil Profissionalizado e no segundo semestre de 2008, foram instituídos os Centros Educacionais para a Juventude – CEJOVEM, posteriormente denominados de Escolas Estaduais de Educação Profissional (EEEP), o qual, passou a disponibilizar, além do ensino básico do currículo do ensino médio, cursos técnicos e profissionalizantes voltados para diversas áreas e disciplinas que buscam à formação pessoal e convívio com a sociedade. Essa ação proporciona ao aluno concluir o ensino médio e se profissionalizar em áreas direcionadas para o mercado de trabalho. Para tanto, o estudante adquire, como primeira experiência profissional, o estágio obrigatório na área cursada, o qual se realiza no último ano de participação na escola. Este deve ser monitorado por um acompanhamento técnico-pedagógico por profissionais diplomados de cada área específica, visando avaliar e orientar os estudantes. Esse monitoramento é realizado por meio de um Sistema Informatizado de Captação e Estágios – SICE para cada curso de cada escola,

o qual permite gerenciar a captação de vagas, encaminhar os alunos às concedentes, acompanhar os processos de formalização jurídica do estágio e da frequência mensal do estagiário.

Segundo informações divulgadas pela SEDUC, em 2008, eram 25 escolas em 20 municípios, em 2014 esse número aumentou para 106 e 112, respectivamente. Em 2016, existem 115 Escolas Estaduais de Educação Profissional (EEEP) no estado, perfazendo um atendimento a cerca de 40 mil alunos, e oferecendo 53 cursos técnicos em diversas áreas, tais como: saúde, agricultura, computação, finanças, dentre outras, as quais são selecionadas e implantadas na EEEP conforme a realidade local da cidade onde esta é implantada. Em relação aos recursos gastos na construção, manutenção e funcionamento das escolas já foram gastos mais de um bilhão de reais em parcerias do governo estadual e federal (SEDUC, 2015).

No período de 2008 a 2014, foram investidos R\$1.036.097.010,22 (Um bilhão, trinta e seis milhões, noventa e sete mil, dez reais e vinte e dois centavos), 71% provenientes do Governo do Estado e 29% do Fundo Nacional de Desenvolvimento da Educação do Ministério da Educação (FNDE)/MEC. Segundo o Relatório de Gestão (2014) o modelo cearense foi inspirado no Programa de Desenvolvimento dos Centros de Ensino Experimental – PROCENTRO - implementado no Estado de Pernambuco. Os Centros de Ensino Experimental (CEE), também conhecidos como ginásios experimentais, eram escolas modelo cuja diretriz pedagógica é prover o ensino integral a estudantes de ensino médio, e atualmente implantada em São Paulo, Goiás e Rio de Janeiro. Diferentemente, as EEEP optaram por incluir no ensino médio a formação profissionalizante.

O processo seletivo dos alunos se dá a partir dos critérios definidos na Portaria de matrícula, o qual é publicada a cada ano, dentre esses, se pode citar: idade mínima de 14 anos, completados até a data do fechamento do censo (exceção dos cursos do eixo saúde, que exigem idade mínima de 14,5 anos); ter concluído o ensino fundamental; média de todas as disciplinas cursadas nas séries finais do ensino fundamental. Os alunos são classificados considerando-se o número de vagas disponíveis, não existindo assim, nota de corte. Vale ressaltar que 80% das vagas são destinadas a alunos oriundos da rede pública e 20% são oriundos da rede particular.

A contratação de professores ocorre a partir de uma seleção que compreende análise de currículo, participação em seminário sobre a rotina de trabalho das EEEP, e entrevista. Já os professores das disciplinas técnicas são contratados via CENTEC, com regime CLT. Estes passam por seleção organizada pelo próprio CENTEC, que compreende provas de: Português, Didática, Educação Profissional e prova específica.

Segundo um Informe do IPECE (2013), a melhora no desempenho do estado do Ceará na participação no ENEM dos anos 2009-2011 pode estar relacionada a maior participação dos alunos das EEEP do estado. Para isso, o trabalho utilizou as médias no exame nesse período e observou que 29 escolas profissionais tiveram médias melhores ou iguais à nacional. Além disso, no ranking das melhores escolas do Ceará, as profissionalizantes ocuparam 33 posições das 50. Além disso, o Relatório de Gestão (2014) constatou que houve uma ampliação da aprovação de egressos das EEEP na universidade e uma redução da inserção no mercado de trabalho e segundo informações do INEP (2015) referentes ao exame ENEM/2014, 8 (oito) das 10 (dez) melhores escolas são de ensino profissional, destaque para Adriano Nobre EEEP, na cidade de Itapajé, no interior do Ceará, o qual obteve média geral de 588.16.

3 Metodologia

3.1 Métodos de Avaliação de Impacto

Avaliações de impacto têm por objetivo inferir sobre os resultados advindos da ocorrência de uma determinada ação a um grupo de agentes ou entidades relativo a outro de controle. No campo social, busca-se mensurar o grau de sucesso ocorrido aos indivíduos sob essa ação ao compará-los a um contrafactual que capta o diferencial resultante na ausência de tal intervenção.

O procedimento é feito através da observação de uma amostra de agentes, os quais são particionados em um grupo composto pelos denominados tratados e outro grupo contendo os não-tratados ou controle, onde a variável binária $d_i = 1$ indica que o indivíduo i recebeu o tratamento; Y_{i1} e Y_{i0} são os resultados após a adesão ao programa e caso não fosse beneficiada pelo programa, respectivamente. Busca-se encontrar o resultado de $E[Y_{1i} - Y_{0i}]$, o qual pode ser reescrito como $E[Y_i|d_i = 1] - E[Y_i|d_i = 0] = E[Y_{1i}|d_i = 1] - E[Y_{0i}|d_i =$

0]. Somando e subtraindo $E[Y_{0i}|d_i = 1]$, isto é, o resultado esperado para o indivíduo i que recebeu o tratamento caso não o tivesse, obtem-se:

$$\begin{aligned} \underbrace{E[Y_i|d_i = 1] - E[Y_i|d_i = 0]}_{\text{Diferença Observada}} &= \underbrace{E[Y_{1i}|d_i = 1] - E[Y_{0i}|d_i = 1]}_{\text{Efeito médio do Tratamento nos tratados (EMTT)}} + \\ &\underbrace{E[Y_{0i}|d_i = 1] - E[Y_{0i}|d_i = 0]}_{\text{Viés de Seleção}} \end{aligned} \quad (1)$$

Se a definição dos grupos não ocorrer de forma aleatória, então devem existir outras características dos agentes envolvidos que determinem a designação para cada grupo, ou seja, o pressuposto da hipótese de identificação requer que existam indivíduos de ambos os grupos, tratamento e controle, para cada característica X_i para o qual se deseja comparar. Para esse tipo de modelo, chamados de não-experimentais, haverá o problema de auto-seleção (BECKER; ICHINO, 2002; ANGRIST; PISCHKE, 2009). Dessa forma, o valor da variável dependente (Y) passa a ser independente da condição (tratado ou controle), uma vez que os indivíduos com características observáveis idênticas possuem a mesma chance de receber o tratamento, logo, $\{Y_{0i}, Y_{1i}\} \perp d_i | X_i$. Dessa forma, segundo Angrist e Pischke (2009), condicionando para o vetor X_i , o viés de seleção some. Formalmente,

Condicionando às variáveis observáveis, X_i , e assumindo a hipótese da independência condicional (HIC), tem-se que:

$$\begin{aligned} \underbrace{E[Y_i|X_i, d_i = 1] - E[Y_i|X_i, d_i = 0]}_{\text{Diferença Observada}} &= \underbrace{E[Y_{1i}|X_i] - E[Y_{0i}|X_i]}_{\text{Efeito médio do Tratamento nos tratados (EMTT)}} + \underbrace{E[Y_{0i}|X_i] - E[Y_{0i}|X_i]}_{\text{Viés de Seleção}} = \\ &= E[Y_{1i} - Y_{0i} | X_i] \end{aligned} \quad (2)$$

3.1.1 Pareamento por Escore de Propensão (PEP)

O Pareamento por Escore de Propensão (PEP), desenvolvido por Rosenbaum e Rubin (1983), lida com o problema da dimensionalidade, aproximando-se a características da estimação da relação causal de um experimento aleatório. Para tanto, a hipótese da independência condicional, descrita anteriormente, deve ser satisfeita a fim de se construir um grupo de controle similar ao grupo de tratamento tomando com base na distribuição de variáveis observadas. De forma geral, o PEP corresponde à probabilidade condicional de uma

unidade receber o tratamento em virtude de um conjunto das características observáveis X , o qual pode ser calculado conforme a equação:

$$P(X) = P(D = 1 | X) \quad (3)$$

Esse método apresenta uma solução prática para o problema da multidimensionalidade no pareamento e o efeito de tratamento pode ser determinado de acordo com a equação (4):

$$E(Y_1 - Y_0 | D = 1, P(X)) = E(Y_1 | D = 1, P(X)) - E(Y_0 | D = 0, P(X)) \quad (4)$$

O viés associado às diferenças do nível das variáveis observadas é eliminado ajustando as diferenças entre as unidades de tratamento e controle, gerando um estimador que permite que a distribuição das características observadas (X) do grupo de controle e do grupo de tratamento seja idêntica. Então, o contrafactual pode ser construído a partir dos resultados do grupo de tratamento, de acordo com:

$$E(Y_0 | P(X), D = 1) = E(Y_0 | P(X), D = 0) = E(Y_0 | P(X)) \quad (5)$$

O processo de seleção ocorre através de características observáveis, onde as unidades que as possuam de forma idêntica tenham a mesma probabilidade de serem alocadas como tratamento ou controle. Rosenbaum e Rubin (1983) propõe o Teorema do Escore de Propensão, o qual torna prático o pareamento, reduzindo o número de variáveis do vetor X_i a um único escalar, a probabilidade de recebimento do tratamento, dado as características observadas.

Angrist e Pischke (2009) mostra que esse teorema pode ser enunciado da seguinte forma: Supondo que a Hipótese de Independência Condicional (HIC) seja satisfeita tal que $\{Y_{0i}, Y_{1i}\} \perp d_i | X_i$, então $\{Y_{0i}, Y_{1i}\} \perp d_i | P(X_i)$, ou seja, se os resultados potenciais são independentes da variável de tratamento condicionada a um vetor multivariado X_i , então os resultados potenciais são independentes da variável de tratamento condicionada a uma função escalar desse mesmo vetor, que é o escore de propensão, definido como $P(X_i) \equiv E[d_i | X_i] = P[d_i = 1 | X_i]$.

Heckman, LaLonde e Smith (1999) afirmam que a hipótese de suporte comum garante que as observações de tratamento têm observações de comparação "nas proximidades" na distribuição de escore de propensão, ou seja, para algum $c > 0$, $c < p(x) < 1 - c$. Segundo Lee (2006), outra hipótese necessária para a estimação dos efeitos de tratamento utilizando métodos de pareamento é o *critério de balanceamento*, o qual é satisfeito quando, para cada

valor do escore de propensão, X tem a distribuição similar para os grupos de tratamento e controle, ou seja, $D \perp X | P(X)$.

A estimação por escore de propensão ocorre da seguinte forma: estima-se $P(X_i)$ com algum modelo paramétrico, como probit ou logit e em seguida, a estimação do efeito do tratamento⁶ pode ser procedida pelo pareamento do valor encontrado no primeiro passo ou através de algum esquema de pesos (CALIENDO; KOPEINIG, 2005). Os principais métodos de pesos utilizados para a estimação do EMTT no procedimento PEP são definidos e denominados em Caliendo e Kopeinig (2005) como *nearest-neighbor*, *radius*, *stratification*, *kernel*, os quais serão testados no pareamento, e conduzido o teste de balanceamento com cada um deles para se selecionar o mais adequado na inferência do EMTT.

Para que esses estimadores baseados no escore de propensão possam ser considerados sem viés é necessário que as condições de receber o tratamento sejam exógenas. Além disso, outro ponto importante a ser analisado é quanto à inclusão de variáveis no modelo como controle. Alguns autores sugerem que elas sejam escolhidas tanto pela relevância na designação e participação no tratamento, levando em consideração os fatores econômicos, sociais e políticos do programa, quanto nas variáveis de resultado, cujos impactos se desejam mensurar. Assim, o método PEP depende do grau em que as características observadas explicam a participação no programa.

Geralmente, a intuição econômica é utilizada para escolher as covariáveis, porém, esta pode não ser precisa em determinar de maneira exata quais variáveis de controle devem ser utilizadas, além de a forma funcional pode estar sujeita a equívocos. Dessa forma, se a escolha das covariadas não inclui variáveis determinantes do recebimento do tratamento, a presença destas características não observáveis fará com que o método utilizado não seja capaz de reproduzir os resultados de um experimento aleatório. Assim, no caso da escolha incorreta dessas variáveis e da forma funcional, o modelo poderá gerar estimativas viesadas, caso a variável de tratamento não se comporte de forma exógena quando condicionada ao conjunto de controles.

⁶ Esse procedimento se dá dentro da região de suporte comum.

3.2 Estimação do Escore de Propensão via seleção robusta de variáveis de controles

A estimação do escore de propensão está sujeita a duas escolhas realizadas a priori pelo pesquisador, seja em base intuitiva ou teórica. Primeira, a seleção de variáveis que serão utilizadas como controle e, segunda, a forma funcional que tais variáveis serão utilizadas. Assim, considerando que a dimensão de X é p , seja s o número de variáveis escolhidas pelo pesquisador e w o número apropriado de variáveis a serem selecionadas. No caso da escolha inapropriada de variáveis, $s \neq w$, pode ocorrer omissão de variável ou excesso de variáveis selecionadas. No primeiro caso, as estimativas do escore de propensão tornar-se-iam viesadas, visto que provavelmente o conjunto de variáveis relevantes é correlacionado. No segundo caso, as estimativas do escore de propensão não seriam viesadas, porém, haveria perda de eficiência na inferência ocasionada pela inclusão de variáveis desnecessárias. Além disso, com os recentes avanços da computação, a existências de grandes bases de dados tem proporcionado um desafio adicional. Tendo em vista o vasto potencial de variáveis de controle, métodos como MQO ou máxima verossimilhança são inaplicáveis se o número de variáveis exceder o número de observações. Portanto, a própria existência das estimativas depende da eliminação de certas variáveis de controle potenciais. Ou seja, métodos tradicionais de estimação do escore de propensão podem ser viesados, ineficientes e requerem a eliminação de variáveis potencialmente relevantes.

Técnicas de regularização (ou *shrinkage*) tem sido aplicadas em problemas envolvendo efeitos de tratamento. Belloni et al (2014), por exemplo, aplicam a técnica do Lasso para estimar o efeito de tratamento por meio da seleção robusta de variáveis instrumentais fracas. Farrell (2013), por sua vez, aplica a técnica do lasso grupado para obter estimativas robustas a heterogeneidade dos efeitos de tratamento médio. Adicionalmente, demonstra que o seu estimador atinge o limite semiparamétrico de eficiência. Resultados de simulação e empíricos mostram um bom comportamento em amostras finitas.

Belloni et al (2015) desenvolveram um método para estimação robusta do escore de propensão utilizando dupla seleção tanto para modelos com alta dimensão ($p \gg n$) ou baixa dimensão ($p \ll n$). Além disso, permitem estimação para efeitos heterogêneos como efeito médio sobre o tratamento local (LATE) e o efeito médio sobre o tratamento quantílico (LQTE). Tal técnica pode ainda ser aplicada a casos endógenos ou exógenos para a variável

de tratamento. Técnicas Bayesianas para modelos de escore de propensão também têm sido aplicadas, especialmente na área de saúde. Estes trabalhos incluem entre outros, Hahn, Carvalho e Puelz (2016), Wang, Pagamini e Dominici (2012), entre outros.

Dessa forma, o presente estudo propõe-se em aplicar a técnica de seleção do lasso e do pós-lasso na estimação do escore de propensão, similarmente a Belloni et al (2015), para avaliar o EMTT das escolas EEEP. Após a estimação robusta do escore de propensão serão aplicados procedimentos de *matching* para realizar comparações entre os grupos tratados e não tratados.

3.2.1 Método de seleção de variável por meio dos métodos Lasso e Pós-Lasso

O método Lasso (*Least Absolute Shrinkage and Selection Operator*), originalmente proposto por Tibsharani (1996), consiste na estimação de parâmetros através da imposição de restrições sobre o valor absoluto dos coeficientes estimados. Belloni et al (2015) desenvolveram métodos para avaliar a inferência de modelos cuja estimação seja realizada por métodos de seleção como o Lasso. Tais resultados se aplicam a modelos de EMTT, com ou sem variáveis endógenas para a variável de tratamento.

Supondo que o escore de propensão seja calculado utilizando a função *link* logit, então:

$$E(Y_i|X) = \frac{\exp(g(X))}{1 + \exp(g(X))} + \varepsilon_i \quad (6)$$

As técnicas tradicionais de estimação de (6) utilizam a intuição econômica para determinar quais variáveis e qual a forma funcional de $g(X)$. Seja $\beta = (\beta_1, \dots, \beta_p)$ o vetor de parâmetros de $X \in X$, em que X é o suporte de X . Assuma que X seja formado por variáveis de controle e por transformações dessas variáveis⁷ e que $p \leq \dim(X)$. O método Lasso selecionará s ($s \ll p$) parâmetros após a aplicação do seguinte problema de minimização:

$$\beta_L = \arg \min \hat{Q}(\beta) + \frac{\lambda}{n} \|\beta\|_1 \quad (7)$$

Onde, $\hat{Q}(\beta) = E[(g_i - X\beta)^2]$, $g_i = \{0,1\}$, sendo 1 se o indivíduo i participou do programa e 0 se não participou, $\|\beta\|_1 = \sum_{j=1}^p |\beta_j|$. Isto é, o método minimiza a soma dos quadrados dos resíduos condicionada a restrições nos parâmetros.

⁷ Estas podem incluir variáveis não transformadas e variáveis que sofrem alguma transformação, tais como, interação entre as variáveis, variáveis polinomiais, *B-Splines* dentre outras.

O termo λ , chamado de *turning point*, corresponde ao peso que será dado à restrição, indicando que quanto maior seu valor, maior será o efeito da restrição sobre os coeficientes estimados. Observe que a equação (7) pode ser rescrita da seguinte forma:

$$\beta_L = \arg \min \sum_{i=1}^n (g_i - X\beta)^2 \quad (8)$$

$$\text{sujeito a } \sum_{j=1}^p |\beta_j| \leq t \quad (9)$$

Esta última forma deixa mais explícita o modo pelo qual a otimização restrita é aplicada pelo método Lasso. De fato, a estimação dos parâmetros será feita deste que a soma não ultrapasse o *turning point*. A escolha do λ é elemento crucial no processo de seleção das variáveis. Tibshirani (1996) propôs a escolha de λ via correlação cruzada, ou seja, o método é aplicado considerando um intervalo para o valor do λ , de onde seleciona-se aquele que minimiza o erro quadrático médio. Tal estratégia, entretanto, não é diretamente aplicável a problemas de estimação de modelos estruturais, por meio de variáveis instrumentais ou no caso de avaliação de políticas, pois, o interesse não é necessariamente realizar previsão.

Hanh e Carvalho (2015) utilizam métodos bayesianos para determinar o valor do *turning point* através da análise gráfica. Eles aplicam tal metodologia para estimação do problema do investidor que maximiza a relação entre retorno e risco por meio de pesos ótimos. Bickel et al (2009) propuseram o seguinte valor: $\lambda = 2 \cdot c\sigma\sqrt{2n\log(2p/\zeta)}$. Em que, $c > 1$, $1 - \zeta$ refere-se ao intervalo de confiança para que a probabilidade de rejeição ficar próximo de 1, σ consiste no desvio padrão do erro. Tal derivação é motivada buscando obter *near-optimal* taxas de convergência dos estimadores após a seleção.

Todavia, tal escolha possui duas limitações. Primeira, foi derivada supondo a homocedasticidade. Segunda, não é factível, tendo em vista que σ não é observável. Ao contrário, Belloni et al (2010) propuseram um λ factível e aplicável na presença de heteroscedasticidade, cujo procedimento será seguido neste trabalho para estimação da equação (7).

Portanto, o objetivo deste trabalho consiste em aplicar o método de seleção de variáveis Lasso sobre o escore de propensão. Isso permitirá obter um escore através da estimação de $g(X)$ composta de variáveis selecionadas de forma robusta. Apesar de o método não requerer a escolha de variáveis por meio da teoria econômica, em certas situações, pode haver o

interesse de algumas variáveis não serem submetidas à restrições, pois pode ocorrer de algumas delas possuírem relevância teórica suficiente, de forma que sua manutenção no modelo seja requerida. Belloni et al (2010), por exemplo, estimam equações de rendimento para o EUA utilizando o método Lasso para selecionar variáveis instrumentais relevantes, porém excluem variáveis como sexo, experiência e experiência ao quadrado de sujeição à restrições.

O Lasso é um caso particular de um conjunto de métodos de seleção de variáveis, conhecidos como *shrinkage methods*, e possui duas vantagens frente às outras técnicas de seleção de variáveis. Primeiro, seleciona variáveis ao forçar certos parâmetros irrelevantes a obter o valor zero, no sentido de não gerar redução do erro quadrático médio com a inclusão da variável. Segundo, a forma funcional da equação (7) é convexa. Portanto, o problema da otimização é computacionalmente mais simples e os ótimos obtidos serão globais. Estas duas propriedades fazem com que o método do lasso seja bastante apropriado para selecionar variáveis de controle relevantes para explicar o comportamento da variável binária ou da variável de tratamento, como em Belloni *et al* (2014).

Entretanto, o método do lasso possui também limitações, notadamente, caso haja forte de multicolinearidade (ZOU; HASTIE, 2005). Isso ocorre justamente pela forma através da qual a seleção de variáveis é realizada, em que as variáveis que possuem elevada correlação têm seus coeficientes estimados forçados a zero. Mais grave, porém, é o fato de o Lasso produzir estimativas viesadas. O método foi desenvolvido para selecionar variáveis de forma a aumentar o poder preditivo dos modelos, ou seja, minimizar a função perda quadrática. Todavia, à medida que a variância do erro quadrático médio se reduz, o viés se eleva⁸. Para reduzir o peso do viés, métodos alternativos têm sido propostos. Por exemplo, Zou e Hastie (2002) propuseram o *elastic net*, que busca suavizar a perda com o viés do Lasso sem, no entanto, perder as qualidades de seleção de variáveis. Belloni *et al* (2012, 2013) por sua vez, adotam o pós-Lasso que consiste em reestimar a equação (8) por MQO após a seleção de variáveis. Isto é, inicialmente realiza-se a seleção de variáveis, em que s dentre as p possíveis

⁸ De fato, sendo y_0 a variável dependente e $\hat{f}(x_0)$ os previsores estimados então, o erro quadrático médio pode ser decomposto em $E(y_0 - \hat{f}(x_0))^2 = Var(\hat{f}(x_0)) + [Viés(\hat{f}(x_0))]^2 + Var(\epsilon)$, em que: ϵ consiste no erro de aproximar y_0 por $\hat{f}(x_0)$. Para maiores detalhes ver, Hastie et al (2009).

variáveis são escolhidas ($s \ll p$). Em seguida, a equação (7) é reestimada considerando apenas as variáveis selecionadas. Por fim, utiliza-se a estimativa do primeiro estágio para estimar a equação estrutural de segundo estágio. Esta será a técnica utilizada neste trabalho.

Para compor o conjunto de variáveis a serem selecionadas pelo método Lasso, o qual irá determinar a probabilidade do aluno pertencer ao grupo de tratados, ou seja, iniciar e permanecer o ensino médio nas EEEP de 2012 a 2014, foram inseridas variáveis de características pessoais e educacionais dos alunos, status educacional dos pais e as condições de infraestrutura das escolas no período de 2011, período anterior a entrada desses alunos nas escolas profissionalizantes. Essas variáveis foram transformadas, uma vez que algumas transformações também podem ter poder de explicação sobre os tratados.

Vale notar que apesar da limitação da factibilidade imposta pelos métodos tradicionais de estimação, o método do lasso não se restringe a modelos com dimensões baixas ($p \ll n$). Ou seja, mesmo que os modelos tenham dimensões elevadas ($p \gg n$), ainda assim, a método do lasso pode ser aplicado. Esta possibilidade permite a obtenção de um escore de propensão muito mais explicativo e com maior probabilidade de refletir o grupo dos tratados.

Por não possuir acesso a tais informações, Mariano et al (2015) não trabalhou com dados longitudinais, assim, não considerou variáveis de *background* educacional dos alunos, tais como, a nota de proficiência em Matemática e Português no SPAECE 2011, os quais acredita-se serem bastante relevantes para determinação do aluno ingressar nas EEEP e para o procedimento de pareamento de forma a captar somente o efeito escola.

3.3 Algoritmo CEM

Dado que existe um processo seletivo para a entrada dos alunos nas EEEP, pode haver argumentos que relacionem o bom desempenho dessas escolas às capacidades dos alunos selecionados. Com isso, visando isolar os efeitos de alunos que já eram bons antes de ingressarem nas escolas profissionalizantes, este trabalho optou por aplicar o algoritmo CEM e fazer o pareamento ex-ante pela nota dos alunos no 9º ano do ensino fundamental nas proficiências de matemática e português obtidas pelo SPAECE 2011, de forma que, a amostra resultante não tenha diferença/desequilíbrio no desempenho prévio entre tratados e grupo de controle.

Segundo Blackwell et. al. (2009), o CEM, desenvolvido por Iacus, King e Porro (2008), é um método de pareamento que reduz o desequilíbrio na distribuição empírica entre tratados e controles, permitindo que o equilíbrio seja pré-escolhido pelo próprio pesquisador. Esse algoritmo não exige nenhuma hipótese sobre o processo de geração de dados, com exceção da ignorabilidade, e assegura que os desequilíbrios entre os grupos após o pareamento não sejam maiores que um limite previamente selecionado. Os autores mostraram que o CEM permite maior robustez de análise de outros métodos, como o Pareamento por Escore de Propensão PEP, razão pela qual será usado neste trabalho⁹.

Iacus, King e Porro (2008) desenvolve uma medida de desequilíbrio global, dada pela estatística \mathcal{L}_1 , onde esta é baseada na diferença entre histogramas multidimensionais de todas as covariadas de pre-tratamento nos grupos de tratados e controles e pode ser representada por: $\mathcal{L}_1(f, g) = \frac{1}{2} \sum_{l_1 \dots l_k} |f_{l_1 \dots l_k} - g_{l_1 \dots l_k}|$. O equilíbrio perfeito global é encontrado por $\mathcal{L}_1 = 0$, e valores elevados para \mathcal{L}_1 indicam grandes desequilíbrios entre os grupos, com valor máximo de $\mathcal{L}_1 = 1$, no qual indica completa separação. Denote as frequências relativas dos dados pareados por f^m e g^m , então uma boa solução para o pareamento produzirá uma redução na estatística \mathcal{L}_1 , ou seja, $\mathcal{L}_1(f^m, g^m) \leq \mathcal{L}_1(f, g)$.

3.4 Análise de Sensibilidade

Se variáveis não observadas, que afetam o processo de seleção, também afetam os resultados, então os estimadores baseados no escore de propensão não são estimadores consistentes do efeito do tratamento. Dessa forma, a omissão dessas variáveis pode gerar resultados do efeito médio do tratamento sobre o tratado estimado viesados. O método conhecido como *Rosenbaum bounds* (ROSENBAUM, 2002; DIPRETE; GANGL, 2004), permite determinar quão "forte" deve ser a influência de uma variável omitida sobre a seleção na participação. Considerando-se que a probabilidade de participação de um indivíduo i seja dada por:

$$\pi_i = \Pr(D_i = 1|x_i) = F(\beta x_i + \gamma x_i) \quad (10)$$

⁹ Alguns autores têm destacado ter tido sucesso em aplicações recentes com o CEM: Datta (2015), Aroca et al (2014), Schurer et al (2015).

Como já explicitado D_i é igual a 1 se o indivíduo recebe o tratamento e 0 caso não receba; x_i são as características observadas do indivíduo i ; u_i corresponde à variável não observada e γ representa o efeito de u_i sobre a decisão de participação no programa. Se não existir viés de seleção, então γ será igual a zero e a probabilidade de participação será exclusivamente determinada pelas características observáveis. Entretanto, na presença de viés de seleção, dois indivíduos com as mesmas co-variáveis observadas x terão diferentes chances de receber tratamento, $\frac{\theta_i}{1-\theta_i}$ e $\frac{\theta_j}{1-\theta_j}$ e a *odds ratio* é dada por:

$$\frac{\frac{\theta_i}{1-\theta_i}}{\frac{\theta_j}{1-\theta_j}} = \frac{\theta_i(1-\theta_j)}{\theta_j(1-\theta_i)} = \exp[\gamma(u_i - u_j)] \quad (11)$$

Se os indivíduos possuírem as mesmas características observáveis, então o vetor x se cancela. Deste modo, se não houver diferenças nas variáveis não observadas ($u_i = u_j$) e se estas variáveis não influenciarem a probabilidade de participação ($\gamma = 0$), a *odds ratio* será igual a 1, implicando a não-existência de viés de seleção. Segue-se então, que se suas *odds* de participação diferirem, isto é, se a *odds ratio* for diferente de 1, só pode ser devido à presença de não-observáveis. A análise de sensibilidade avalia o quanto do efeito do programa é alterado pela mudança nos valores de γ e de $u_i - u_j$. Isso significa examinar os limites da *odds ratio* de participação. Rosenbaum (2002) mostra que (11) implica os seguintes limites para a *odds ratio*):

$$\frac{1}{e^\gamma} \leq \frac{\alpha_i(1-\theta_j)}{\theta_i(1-\theta_i)} \leq e^\gamma \quad (12)$$

Os indivíduos pareados possuem a mesma probabilidade de participação apenas se $e^\gamma=1$. Entretanto, se $e^\gamma = 2$, então indivíduos aparentemente similares em termos de x irão diferir nas probabilidades de receberem tratamento por um fator de até 2.

3.5 Variáveis e Bases de Dados

Para avaliar o impacto da política de profissionalização no ensino médio das EEEP do Ceará sobre o desempenho dos alunos no ENEM foram utilizados como indicadores de impacto as notas dos alunos nas cinco áreas e na média geral do ENEM, quais sejam: Ciências da Natureza (CN), Ciências Humanas (CH), Linguagem e Códigos (LC), Matemática (MT), Redação (R), Média Geral (MG).

As informações referentes ao desempenho nas cinco áreas de conhecimento do ENEM 2014¹⁰ e dotações de infraestrutura das escolas em 2011 foram retiradas do INEP, através dos Microdados do ENEM e do censo escolar. As características pessoais e educacionais, status educacional dos pais dos alunos das escolas da rede estadual de educação do Ceará e a identificação das escolas profissionalizantes foram obtidos da Secretaria de Educação do Ceará – SEDUC através do SPAECE 2011. Dessa forma, para construção da amostra a ser utilizada por este trabalho, fez-se necessário a junção de três diferentes bases de dados, SPAECE, ENEM e censo escolar, o qual só foi possível através da disponibilização de identificadores criados em pela SEDUC para acompanhar o aluno no 9º ano do ensino fundamental ao 3º do ensino médio.

Para atender o objetivo proposto por este trabalho, precisa-se construir um grupo que pertence ao programa (grupo de tratados) e outro grupo que não foi beneficiado (grupo de controle). O grupo de tratamento será composto pelos alunos que ingressaram o ensino médio nas escolas profissionalizantes em 2012 e estava cursando o 3º ano em 2014 nessas escolas, mesma definição para o grupo de controle, porém, refere-se as escolas regulares. Para construir o grupo de tratados é necessário acompanhar esses alunos durante o ensino médio, ou seja, identificar se estes permaneceram nas escolas profissionalizantes e se repetiram o ano durante o período de 2012 a 2014. Assim, utilizou-se a base de dados do SPAECE 2012 e ENEM 2014 para essa identificação.

Vale ressaltar que não foi possível observar esses alunos em 2013, pois o SPAECE 2013 foi apenas amostral. Porém, acredita-se que isso não comprometa os resultados a serem encontrados por este trabalho, pois o acompanhamento desses alunos nos anos de 2012, 2013 e 2014 serve apenas para identificar se estes permaneceram nas escolas profissionalizantes durante o ensino médio. Dado que foi possível observar os alunos que fizeram o 1º ano nas EEEP em 2012 e o 3º ano em 2014 na mesma escola, espera-se que, este aluno também cursou o 2º ano na referida escola, pois dificilmente esse aluno sairia da profissionalizante em um ano para voltar no ano seguinte.

¹⁰ Último ano com informações disponíveis sobre o desempenho individual no ENEM.

Inicialmente, para compor a amostra, os alunos devem ter feito o SPAECE e ter cursado o 9º ano do ensino fundamental em 2011 para que se possa extrair as variáveis que irão determinar a entrada nas escolas profissionalizantes em 2012, e ter participado do ENEM em 2014 para obtenção das variáveis de resultado, perfazendo um total de 40.435 alunos. Destes, 35.680 estudantes puderam ser observados também no SPAECE 2012, onde 35.312 iniciaram o ensino médio nesse ano¹¹, onde, 5.491 alunos ingressaram nas escolas profissionalizantes. Em 2014, 5.092 estudantes permaneceram nas EEEP, sendo que, 24 repetiram o 2º ano e 5.068 estavam cursando o 3º ano, e destes, 4.977 estavam presentes nos dias das provas do ENEM 2014, porém 4.635 possuíam informações referentes às características pessoais (Grupo de tratados). Nas escolas regulares, 29.658 permaneceram nas referidas escolas em 2014, onde 6 alunos estavam no 1º ano, 1.937 no 2º ano e 27.715 estavam no 3º ano, sendo que, destes, apenas 22.201 foram fazer o exame neste ano e 19.811 responderam ao questionário pessoal (grupo de controle)¹².

4 Resultados

4.1 Análise Descritiva

Estatísticas descritivas das variáveis dependentes e de algumas variáveis de controles¹³ relacionadas aos alunos de EEEP e escolas regulares do Ceará, extraídas das bases de dados do SPAECE 2011, Censo escolar 2011 e ENEM 2014, estão explicitadas na Tabela 1.1.

Observa-se que as médias das notas no ENEM 2014 dos alunos das EEEP são superiores às de alunos de escolas regulares e o coeficiente de variação é menor ou igual em todas as provas. Em Ciências da Natureza (CN), o diferencial de médias é de 40,12 pontos a favor das escolas profissionalizantes e detém a mesma variação relativa. As áreas Ciências Humanas (CH) e Linguagem e Códigos (LC) apresentam diferença, 52,32 e 49,39 pontos, respectivamente, com dispersão relativa à média de 16% e 36% superior à EEEP, respectivamente.

Apresentando diferenciais mais elevados e menos homogêneo, estão as provas de Matemática (MT) e Redação (R), com diferença de 57,51 e 143,77 pontos, e variação relativa

¹¹ 368 alunos repetiram o 9º ano em 2012

¹² Para mais detalhes ver Apêndice

¹³ Total de 95 variáveis de controle.

de 4,7% e 56,6%, superior à EEEP, respectivamente. Na Média Geral, os alunos destas escolas apresentaram desempenho superior em, aproximadamente, 68 pontos, 15,4 % a mais que os alunos das escolas regulares, porém, observa-se que, estas apresentam variação relativa à média superior à EEEP de, aproximadamente 21%.

A partir dessa análise descritiva das variáveis dependentes, espera-se que o efeito da política do ensino profissionalizante seja significativo sobre a performance dos alunos no ENEM, porém não seja suficiente somente através da comparação entre esses grupos, pois apenas por não ter passado pela intervenção não significa que o grupo de não tratados representa bem o que ocorreria com o grupo de tratamento caso este não tivesse sido tratado, ou seja, a ausência do tratamento para alguns não gera automaticamente o contrafactual de não tratamento para outros. Dessa forma, são necessárias metodologias apropriadas que busquem isolar o efeito dos programas dos efeitos causados por outros fatores que afetam as variáveis de resultado de interesse.

Sobre as variáveis relacionadas às características dos alunos observadas em 2011, período anterior a entrada destes no ensino médio, presentes na Tabela 1.1, pode-se afirmar que os alunos das escolas profissionalizantes apresentaram pontuação nas provas de Matemática e Português no exame do SPAECE superior aos alunos das escolas regulares em, aproximadamente, 10%. Estas variáveis foram inseridas visando representar um *background* educacional e habilidade dos alunos no ensino fundamental e podem ter sido importantes no processo de seleção desses alunos no ingresso das escolas profissionalizantes.

Ambas as escolas, EEEP e regulares, apresentam características semelhantes relacionadas aos alunos em 2011, idade média de 15,7 anos, mais de 50% são mulheres e aproximadamente, 20% se consideram brancos, porém, observa-se que os alunos que ingressaram nas escolas profissionalizantes possuem pais com maiores níveis de escolaridade, 21% e 31% dos alunos apresentam pai e mãe, respectivamente, com no mínimo o ensino médio incompleto, enquanto a proporção dos alunos das escolas regulares foram 14 e 20%. Vale ressaltar que a proporção dos alunos que não souberam responder sobre a escolaridade do pai e da mãe, pertencentes às EEEP e regulares, foram 29%, 17%, 32% e 22%, respectivamente.

Dentre as variáveis apresentadas na tabela 1.1 que representam as condições socioeconômicas dos alunos em 2011, é possível afirmar que os alunos que ingressaram nas profissionalizantes apresentam melhores condições que os alunos que foram para as escolas regulares, uma vez que, a proporção de alunos cuja família recebe o benefício BF é menor, a proporção que possuem carro, empregada e máquina de lavar é superior em, aproximadamente, 8%, 28%, 65% e 29%, respectivamente.

As características relacionadas à infraestrutura das escolas foram utilizadas por este trabalho, pois espera-se que estas tenham influenciado a tomada de decisão do aluno na escolha da escola após o término do ensino fundamental. Destas expostas na tabela 1, as escolas profissionalizantes apresentam algumas características melhores que as regulares, tais como: 99% apresentam sala de professores, 88% possuem biblioteca e apresentam número de computadores superior às escolas regulares em 59%.

Tabela 1.1 - Estatísticas descritivas das variáveis da amostra – 2014

Variáveis		EEEP			Regular		
		Média	Desvio-padrão	C.V	Média	Desvio-padrão	C.V
Var. Resultados	Nota_CN	489,82	65,29	0,13	449,70	58,92	0,13
	Nota_CH	550,34	66,15	0,12	498,02	69,57	0,14
	Nota_MT	479,07	102,10	0,21	421,56	90,85	0,22
	Nota_LC	512,98	71,55	0,14	463,59	88,06	0,19
	Nota_Red	533,46	158,54	0,30	389,69	184,61	0,47
	Nota_MG	512,98	71,55	0,14	444,51	74,90	0,17
Caract. de Alunos	Spaece_MT_2011	268,67	42,35	0,16	244,49	41,70	0,17
	Spaece_PT_2011	263,37	40,76	0,15	237,74	42,92	0,18
	Idade	15,7	7,82	0,50	15,7	7,27	0,46
	Sexo	0,42	0,49	1,16	0,43	0,49	1,14
Escolaridade Pai	Raça	0,21	0,41	1,95	0,19	0,39	2,05
	E_pai01	0,07	0,25	3,57	0,11	0,31	2,82
	E_pai02	0,43	0,49	1,14	0,43	0,49	1,14
	E_pai03	0,17	0,38	2,24	0,11	0,31	2,82
	E_pai04	0,04	0,21	5,25	0,03	0,18	6,00
Escolaridade Mãe	E_mae01	0,03	0,17	5,67	0,05	0,23	4,60
	E_mae02	0,49	0,50	1,02	0,53	0,49	0,92
	E_mae03	0,23	0,42	1,83	0,14	0,35	2,50
	E_mae04	0,08	0,28	3,50	0,06	0,24	4,00
Cond. Sócioecon.	Bolsa_familia	0,72	0,44	0,61	0,80	0,39	0,49
	Carro	0,51	0,49	0,96	0,47	0,49	1,04
	Empregada	0,09	0,28	3,11	0,07	0,27	3,86
	Computador	0,33	0,47	1,42	0,20	0,40	2,00
	Maquina_lavar	0,31	0,46	1,48	0,24	0,42	1,75
Caract. Escolas	Lixo_coleta	0,99	0,06	0,06	0,99	0,08	0,08
	Sala_diretoria	0,98	0,12	0,12	0,98	0,12	0,12
	Sala_professores	0,99	0,04	0,04	0,96	0,18	0,19
	Laboratório_inform	0,99	0,12	0,12	0,99	0,06	0,06
	Biblioteca	0,88	0,32	0,36	0,83	0,36	0,43
	N° Computadores	68,31	22,13	0,32	42,9	19,35	0,45

Fonte: Elaboração Própria com base nos Microdados do SPAECE 2011, Censo Escolar de 2011 e ENEM 2014

Nota: Escolaridade (01) não estudou, (02) ensino fundamental, (03) nível médio, (04) superior.

4.2 Resultados do Pré-Pareamento

O efeito-aluno pode viesar os resultados da estimação quando se deseja encontrar o efeito da política das escolas profissionalizantes sobre o desempenho escolar do aluno no ENEM. Com isso, como estratégia para isolar os efeitos dos alunos considerados bons antes de ingressarem nas escolas profissionalizantes, aplicou-se o algoritmo CEM (*Coarsened Exact Matching*) a partir das variáveis de proficiências em matemática e português no SPAECE 2011 dos alunos no 9º ano do ensino fundamental.

A tabela 1.2 apresenta as medidas de desequilíbrios antes e após o pareamento. Observa-se que o \mathcal{L}_1 multivariado e de cada variável reduziu para valores mais próximos a zero. Além disso, a diferença entre as médias dos grupos tratados e controle reduziram nas duas proficiências, 24,17 para 0,05, em matemática, e 25,62 para 0,0833, em português. Esse algoritmo permitiu identificar os alunos nos dois grupos, controle e tratados, que podem ser comparados. Observa-se que 19.660 e 4.634 estudantes pertencentes a escolas regulares e profissionalizantes, respectivamente, podem ser pareados a partir das variáveis estabelecidas, de forma a isolar o efeito-aluno. Dessa forma, a aplicação do PEP se restringirá somente aos alunos contidos no pareamento e identificados pelo algoritmo CEM.

Tabela 1.2 - Resultados do algoritmo CEM

Variáveis	Antes Pareamento		Pós pareamento	
	\mathcal{L}_1	Diferença de médias	\mathcal{L}_1	Diferença de médias
MT_SPAECE_2011	0.2424	24.17	0.0522	0.2836
PT_SPAECE_2011	0.2526	25.62	0.0479	0.0833
\mathcal{L}_1 Multivariado	0.3458		0.2253	
Nº Tratados	4635		4634	
Nº Controles	19811		19660	

Fonte: Elaboração Própria

4.3 Seleção das Variáveis de Controle: Método do Lasso

A seleção das variáveis inseridas no modelo PEP foi regida pelo método Lasso. Com base nas 95 variáveis de controle contidas na amostra, foram criadas 449 variáveis a partir das interações entre elas e potência das variáveis, e aplicado o processo de seleção. Destas, o método proposto por Belloni et al (2015) selecionou 39 variáveis consideradas importantes

na determinação do tratamento, das quais, 5 são em valores originais, 7 na forma de potências e 27 são resultados das interações.

4.4 Estimação do pareamento com escore de propensão

Seguindo os procedimentos de Becker e Ichino (2002), inicialmente estima-se um modelo logit binário de o aluno pertencer ou não a uma escola profissionalizante utilizando, como características observáveis, as variáveis selecionadas pelo método Lasso para descrever os grupos de tratamento e controle. Posteriormente, descartam-se da amostra aqueles alunos com escore de propensão fora do suporte.

Os resultados na tabela A.1.1 no apêndice mostram que trinta das trinta e nove variáveis foram significativas na determinação do tratamento quando considerado apenas os alunos que foram pareáveis pelo algoritmo CEM. Verifica-se que a região de suporte comum foi o intervalo, cujo escore de propensão varia de 0.001769, 0.9999984. Essa região assegura que as observações do grupo de tratamento tenham observações comparáveis do grupo de controle quanto às variáveis selecionadas pelo método Lasso.

Dessa forma, o *matching* para se obter o Efeito Médio de Tratamento sobre os Tratados (EMTT) se dará dentro deste intervalo. Porém, outros métodos fazem-se necessário para se estimar o EMTT, além de se utilizar somente o escore de propensão, tais como: *nearest-neighbor*, *Radius* e *Kernel*. Um procedimento importante com relação à construção do escore de propensão e de implementação do pareamento é a comprovação das condições de balanceamento. Os gráficos A.1.1 a A.1.3 no apêndice apresentam os resultados do teste de sobreposição da densidade entre os grupos antes e após o pareamento por esses métodos. Observa-se sobreposições ajustadas e semelhantes entre os métodos de kernel e do *Radius*. Dessa forma, optou-se por escolher o pareamento pelo *kernel* (0,01), no qual foi utilizado para encontrar o EMTT, cujos resultados encontram-se na tabela 1.3.

Tabela 1.3 - Efeito médio do tratamento nas áreas de conhecimento do ENEM, por kernel (0,01)

Variáveis	EEEP	Regular	EMTT	Estatística t
Notas_CN	490,92	445,11	45,11	13,57
Notas_CH	553,05	498,19	54,86	14,30
Notas_MT	484,05	417,55	66,50	13,03
Notas_LC	517,04	458,85	58,19	12,23
Notas_Red	540,86	386,54	154,32	15,36
Notas_MG	517,18	441,39	75,79	18,41

Fonte: Elaboração própria

Os diferenciais de notas entre os dois tipos de escola estimados por pareamento com primeiro *kernel* estão dispostos na Tabela 1.3, a qual revela serem tais diferenciais significativos a favor das EEEP em todas as áreas de conhecimento. Em média, um aluno pertencente a escola profissional apresenta desempenho superior em Linguagens e Códigos (LC) e em Ciências Humanas (CH), aproximadamente, 12,7% e 11%, respectivamente, a mais do que um aluno da escola regular pertencente ao grupo de controle. Em Matemática (MT) o diferencial é 66,5 pontos, 15,92% em favor dos alunos de ensino profissional. Menor e Maior efeito foi observado na área de Ciências da Natureza (CN) e em Redação (Red), de 45,11 pontos (10,13%) e 154,32 pontos (39,9%), respectivamente. Na Média Geral, o impacto das EEEP foi de 75,79 pontos (aproximadamente, 17,17%). Em suma, o ensino profissional de nível médio no Ceará demonstra ser mais competente em potencializar habilidades do conhecimento cognitivo comparativamente ao ensino regular. Os resultados da Tabela A.1.2 no apêndice mostram que o efeito do tratamento sobre os tratados apresenta-se robusto a uma possível presença de viés de seleção em todas as áreas de conhecimento e na média geral.

Considerações Finais

Este trabalho procura ampliar o debate existente na literatura da Economia da Educação sobre o ensino profissionalizante, com ênfase nos efeitos gerados no desempenho escolar. Para tanto, busca-se trazer novas evidências sobre o papel das escolas profissionais públicas do Ceará em nível educacional, ao se comparar alunos dessas escolas com outros

do ensino regular, isolando o efeito-aluno e testando-se a existência de diferenciais significativos entre eles.

Uma das contribuições deste estudo reside na base de dados utilizada para mensurar este efeito, pois a partir da construção de identificadores em parceria com a SEDUC foi possível montar uma amostra com informações longitudinais através do cruzamento de diferentes bases, a saber, SPAECE 2011, 2012; Censo Escolar 2011 e ENEM 2014, até então, não observado pelos autores em outro trabalho. Dessa forma, pôde-se observar o aluno no 9º ano do ensino fundamental e acompanhá-lo durante o ensino médio, verificando assim, a trajetória destes nas escolas.

As informações referentes ao desempenho nas cinco áreas de conhecimento do ENEM 2014 e dotações de infraestrutura das escolas em 2011 foram retiradas do INEP, por meio dos Microdados do ENEM e do censo escolar, respectivamente. As características pessoais e educacionais, status educacional dos pais dos alunos das escolas da rede estadual de educação do Ceará e a identificação das escolas profissionalizantes foram obtidos da SEDUC através do SPAECE 2011, 2012.

Para atender o objetivo proposto por este trabalho, construiu-se um grupo pertencente ao programa e outro grupo que não foi beneficiado, no qual, o grupo de tratamento foi composto pelos alunos que ingressaram o ensino médio nas escolas profissionalizantes em 2012 e estavam cursando o 3º ano em 2014 nessas escolas, mesma definição para o grupo de controle, porém, refere-se as escolas regulares, e com base nas informações desses alunos em 2011 foi possível observar os fatores que possivelmente determinaram a entrada destes nas escolas profissionalizantes em 2012.

Como estratégia para isolar os efeitos dos alunos considerados bons antes de ingressarem nas escolas profissionalizantes, este trabalho aplicou o algoritmo CEM (*Coarsened Exact Matching*) a partir das variáveis de proficiências em matemática e português no SPAECE 2011 para os alunos no 9º ano do ensino fundamental, uma vez que esse efeito pode viesar os resultados da estimação quando se deseja encontrar o efeito da política das escolas profissionalizantes sobre o desempenho escolar do aluno no ENEM. Este método estabeleceu que 19.660 e 4.634 estudantes pertencentes a escolas regulares e

profissionalizantes, respectivamente, podem ser pareados a partir das variáveis estabelecidas, de forma a isolar o efeito-aluno.

Este trabalho também contribui com a literatura para o Brasil na escolha da opção metodológica para a condução dessa avaliação, pois baseou-se em um modelo recentemente desenvolvido por Belloni et al (2015), o qual aplica a técnica de seleção do lasso e do pós-lasso na estimação do escore de propensão. Esse método selecionou 39 das 449 variáveis criadas a partir das 95 variáveis originais.

Para a mensuração do Efeito Médio do Tratamento sobre os Tratados (EMTT) realizou-se o pareamento através dos métodos *nearest-neighbor*, *radius* e *kernel*, que apesar de apresentarem resultados semelhantes, optou-se por *kernel* (0,01) devido ser o mais balanceado para determinação do efeito. Verificou-se que os alunos das escolas profissionalizantes apresentam diferenciais significativos em todas as áreas de conhecimento, entre os que fizeram EEEP e aqueles que cursaram escola regular apenas, cujo efeito mais expressivo foi em Redação, 39,9%, a mais para um aluno da escola profissional. Em média, um aluno pertencente a escola profissional apresenta desempenho superior em Linguagens e Códigos (LC) e em Ciências Humanas (CH), aproximadamente, 12,7% e 11%, respectivamente, a mais do que um aluno da escola regular e em Matemática (MT) o diferencial é 66,5 pontos, 15,9% em favor dos alunos de ensino profissional. Menor efeito foi observado na área de Ciências da Natureza (CN) de 45,11 pontos (10,1%). Na Média Geral, o impacto das EEEP foi de 75,79 pontos 17,17%. Além disso, através do teste de sensibilidade verificou-se que os resultados se apresentam robustos à presença de variáveis omitidas.

Diante do exposto, as EEEP do estado do Ceará demonstram ser mais competentes em potencializar habilidades do conhecimento cognitivo comparativamente ao ensino regular, pois ao observar um background educacional do aluno em 2011 e considerá-lo para a obtenção do pareamento, este trabalho minimizou o efeito-aluno e procurou isolar o efeito-escola sobre o desempenho escolar em 2014. Vale ressaltar que estes resultados dependem das formas funcionais e das interações a priori estabelecidas na aplicação do método do Lasso, assim, sugere-se testar outras formas e observar se o EMTT permanecesse significativa.

Referências Bibliográficas

- ANGRIST, J., PISCHKE, J.S. **Most Harmless Econometrics: an empiricist's companion**. Princeton, New Jersey: Princeton University Press. 2009
- ARAÚJO, A. J. N; CHEIN, F.; PINTO, C. **Ensino Profissionalizante, Desempenho Escolar e Inserção Produtiva: Uma Análise com dados do ENEM**. 2014 Disponível: <http://www.anpec.org.br/encontro/2014/submissao/files_I/i12-ac3a8f487db438fc6278e117ab468a01.pdf>
- AROCA, P., BRIDA, J.G, VOLO, S. Tourism statistics: correcting data inadequacy using coarsened exact matching. Working Paper. **School of Economics and Management at the Free University of Bozen**, 2014.
- ASSUNÇÃO J; GONZAGA G. Educação Profissional no Brasil: Inserção e retorno. **Série Cenários**, n. 3, Brasília: SENALDN, 2010.
- BACALHAU, P.; MATTOS, E. **A Provisão Pública de Ensino Superior como Mecanismo de Seleção por Habilidade: Evidências para o Brasil**. 2013 Disponível em: http://www.anpec.org.br/encontro/2013/files_I/i12-e15541f20ab44aacc725c19288802a8d.pdf Acesso em: 10 Jan. 2016 .
- BARRO, Robert J. Education and Growth. **Annals of Economics and Finance**, v.14, n.2, p.301-328, 2013.
- BECKER, S., ICHINO, A. Estimation of Average Treatment Effects based on Propensity Scores. **The Stata Journal**, v.2 n.4, p.358-377. 2002
- BELLONI et al. High Dimensional Methods and Inference on Structural and Treatment Effects. **Journal of Economic Perspectives**, v. 28, n2, p. 29-50, 2014.
- BELLONI, A; HANSEN, C.; CHERNOZHUKOV, V. Inference Methods for High Dimensional Sparse Econometric Models. **Advances in Economics and Econometrics**, 10th World Congress of the Econometric Society, 2013.
- BELLONI, A; HANSEN, C.; CHEN, D.; CHERNOZHUKOV, V. **LASSO Methods for Gaussian Instrumental Variable Models**. Working paper, Duke University, 2010.
- BELLONI, A; HANSEN, C.; CHEN, D.; CHERNOZHUKOV, V. Sparse Models and Methods for Optimal Instruments with Application to Eminent Domain. **Econometrica**, v.80, n.6, p. 2369-2429, 2012.
- BELLONI, A; HANSEN, C.; FERNANDEZ-VAL, I; CHERNOZHUKOV, V. **Program Evaluation with High-Dimensional data**. CeMMAP working papers. Centre for Microdata Methods and Practice, Institute for Fiscal Studies, 2015.
- BICKEL, P. J.; RITOV, Y.; TSYBAKOV, A. B. Simultaneous analysis of Lasso and Dantzig selector, **Annals of Statistics**, v.37, n.4, p. 1705-1732, 2009.

- BLACKWELL, M. IACUS, S. KING, G. PORRO, G. CEM: Coarsened exact matching in Stata. **The Stata Journal**. v. 9, n. 4, p. 524-546, 2009.
- CALIENDO, M.; KOPEINIG, S. Some Practical Guidance for the Implementation of Propensity Score Matching. **Journal of Economic Surveys**, v.22 n.1, p. 31-72. 2008
- CASTRO, M.H.G.; TIEZZI, S. A reforma do ensino médio e a implantação do ENEM no Brasil. In: BROCK, C.; SCHWARTZMAN, S. (Org.). **Os desafios da educação no Brasil**. Rio de Janeiro: Nova Fronteira, 2005. p. 119-154. Disponível em: <<http://www.schwartzman.org.br/simon/desafios/4ensinomedio.pdf>> Acesso em 12 junho 2015
- CHEN, X; WEKO, T. US Department of Education NCES 2009. **Aurora**, v. 202, p. 502-7334, 2009.
- CROUCH, C., D.; SAKO, M. **Are skills the answer?** The political economy of skill creation in advanced industrial countries. New York: Oxford University Press. 1999
- DATTA, N. Evaluating Impacts of Watershed Development Program on Agricultural Productivity, Income, and Livelihood in Bhalki Watershed of Bardhaman District, West Bengal. **World Development**. v. 66, p. 443-456, 2015.
- DIPRETE, T.; GANGL, M. 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, Abr. 2004. Disponível em: <http://www.wjh.harvard.edu/~winship/cfa_papers/HBprop_021204.pdf>. Acesso em: 12/11/2015.
- FARREL, M. Robust Inference on Average Treatment Effects with Possibly More Covariates than Observations. **Journal of Econometrics**, v.189 n.1 p. 1–23, November 2013.
- FOLEY, P. **The Socio-Economic Status of Vocational Education and Training Students in Australia**. National Centre for Vocational Education Research Ltd. PO Box 8288, Stational Arcade, Adelaide, SA 5000, Australia, 2007.
- GANIMIAN, A. J.; ROCHA, A. S. **Measuring Up?. How Did Latin America and the Caribbean Perform on the 2009 Programme for International Student Assessment (PISA)?**. Partnership for Educational Revitalization in the Americas (PREAL), 60 p., 2011
- HAHN, R.; CARVALHO, C.; PUELZ, D. Bayesian Regularized Regressions For Treatment effects estimation from observational data. Working Paper, **Duke University**, 2015.
- HANUSHEK, E. A.; WOESSMANN, L. Schooling, Educational Achievement, and the Latin American Growth Puzzle. **Journal of Development Economics**, v.99, n.2, p.497-512, 2012.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learnings: Data Mining, Inference and Prediction**. Springer Series in Statistics, Second Edition, 2012.
- HECKMAN, J.; LALONDE, R.; SMITH, J. The economics and econometrics of active labor market programs. In: ASHENFELTER, O., CARD, D. (Eds.) **The Handbook of Labor Economics**. Amsterdam: North Holland. v.3A, part.6, cap.31. 1999.
- IACUS, S. M.; KING, G; PORRO, G. Matching for causal inference without balance checking. <http://gking.harvard.edu/files/cem.pdf>. 2008.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS – INEP. Disponível em: <http://www.inep.gov.br/> Acesso em: 10 julho 2015.

KUENZER, A. Exclusão incluyente e inclusão excludente: a nova forma de dualidade estrutural que objetiva as novas relações entre educação e trabalho. In: SAVIANI, D.; SANFELICE, J.L.; LOMBARDI, J.C. (Org.). **Capitalismo, trabalho e educação**. 3. ed. Campinas: Autores Associados, 2005. p. 77-96.

LEE, W. Propensity Score Matching and Variations on the Balancing Test. **Mimeo**. Melbourne Institute of Applied Economics and Social Research. 2006

LEVY, S., SCHADY, N. Latin America's Social Policy Challenge: Education, Social Insurance, Redistribution. **The Journal of Economic Perspectives**, v.27, n.2, p.193-218, 2013.

MARIANO, F. Z.; ARRAES, R. A; SOUZA, N. O. **Desempenho Escolar e Inserção no Mercado de Trabalho**: Uma Avaliação das Escolas Estadual de Ensino Profissionalizantes (EEEP) do Ceará. In. XI Encontro Economia do Ceará em Debate. IPECE 2015. Disponível em:

<http://www2.ipece.ce.gov.br/encontro/2015/trabalhos/Desempenho%20escolar%20e%20inser%20no%20mercado%20de%20trabalho.pdf> Acesso em: 10 Jan. 2016

MARTINS, A. P. Pressupostos de Gramsci na educação profissional e tecnológica de nível médio. **Revista de Educação, Ciência e Tecnologia**, v. 1, n. 2, 2012.

MOEHLECKE, Sabrina. O Ensino Médio e as novas diretrizes curriculares nacionais: entre recorrências e novas inquietações. **Revista Brasileira de Educação**, v. 17, n. 49, Rio de Janeiro. Jan- abril, 2012. Disponível em: http://www.scielo.br/scielo.php?pid=S1413-24782012000100003&script=sci_arttext Acesso em: 10 julho 2015

MORAES, A. G. E.; BELLUZZO, W. O diferencial de desempenho escolar entre escolas públicas e privadas no Brasil. **Nova Economia**. Belo Horizonte. V.24 n.2 p.409-430 maio-agosto, 2014

NEUMAN, S; ZIDERMAN, A. Vocational Secondary Schools Can Be More Cost Effective than Academic Schools: The Case of Israel. **Comparative Education**, v. 25, n. 2, p. 151-163, 1989.

POLIDANO, C.; TABASSO, D. **Fully Integrating Upper-Secondary Vocational and Academic Courses: A Flexible New Way?** Discussion Paper n. 9694, Jan. 2016

RELATÓRIO DA OECD. “Who are the low-performing students?”, *PISA in Focus*, n. 60, OECD Publishing, Paris. 2016 Disponível em: <http://dx.doi.org/10.1787/5jm3xh670q7g-en>. Acesso em: 10 julho 2016

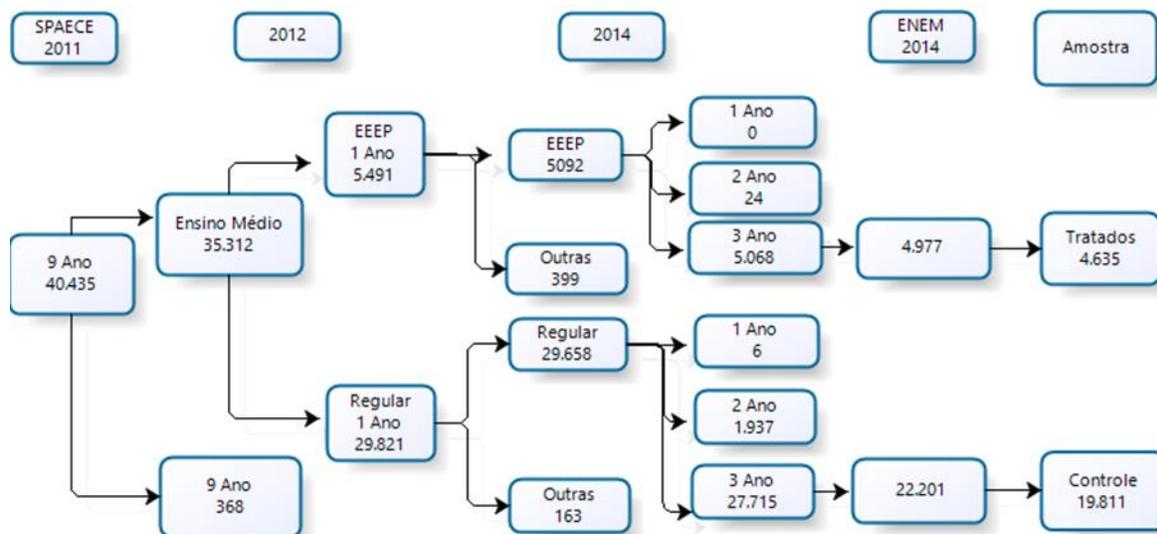
RELATÓRIO DE GESTÃO “O PENSAR E O FAZER DA EDUCAÇÃO PROFISSIONAL NO CEARÁ – 2008 a 2014”. Secretaria da Educação. Coordenadoria de Educação Profissional. – 1. ed. – Fortaleza: Secretaria da Educação, 2014

RELATÓRIO NACIONAL PISA 2012. Resultados Brasileiros, OCDE 2012. Disponível em: http://download.inep.gov.br/acoes_internacionais/pisa/resultados/2014/relatorio_nacional_pisa_2012_resultados_brasileiros.pdf Acesso em: 22 Fev 2016

- ROSENBAUM, P. **Observational Studies**, Springer, New York. 2002.
- ROSENBAUM, P.; RUBIN, R. The Central Role of the Propensity Score in Observational Studies for Causal Effects. **Biometrika**, v.70 n.1, p. 41-55. 1983
- SCHURER, S. ALSPACH, M. MaCRAE, J. MARTIN, G. L., The Medical Care Costs of Mood Disorders: A Coarsened Exact Matching Approach, **IZA Discussion Papers**, 2015.
- SECRETARIA DE EDUCAÇÃO DO CEARÁ - SEDUC. Disponível em: <http://www.seduc.ce.gov.br/index.php/educacao-profissional> Acesso em: 15 de Jun. 2015
- SEVERNINI, E. R.; ORELLANO, V. I. F. O efeito do ensino profissionalizante sobre a probabilidade de inserção no mercado de trabalho e sobre a renda no período pré-Planfor. **Revista Economia**, 2010.
- SILVA, V. H. O. Análise da Participação das Escolas Públicas Estaduais Cearenses no Exame Nacional do Ensino Médio (ENEM): 2009-2011. **IPECE Informe**, n. 54. Fortaleza, 2013. Disponível em: < [http://www.ipece.ce.gov.br/publicacoes/ipece-informe/Ipece Informe 54 A26 fevereiro 2013.pdf](http://www.ipece.ce.gov.br/publicacoes/ipece-informe/Ipece%20Informe%2054%20A26%20fevereiro%202013.pdf)> Acesso em: 12 junho 2015
- TAVARES, P. A. **Três Ensaios em Economia da Educação**. São Paulo, 2012. Disponível em: <http://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/11542/Tr%C3%AAs%20Ensaios%20em%20Economia%20da%20Educa%C3%A7%C3%A3o.pdf?sequence=1&isAllowed=y>. Acesso em: 12 junho 2015
- TIBSHARANI, R. Regression Shrinkage and Selection via the Lasso. **Journal of the Royal Statistical Society, Series B-Methodological**, v.58, n.1, p.267-288, 1996.
- WANG, C., PARMIGIANI, G. DOMINICI, F. Bayesian effect estimation accounting for adjustment uncertainty. **Biometrics** 68 p. 661–671, 2012.
- ZOU, H., HASTIE, T. Regularization and variable selection via elastic net. **Journal of Royal Statistics Society B**, v. 67 n.2, p. 301-320, 2005.

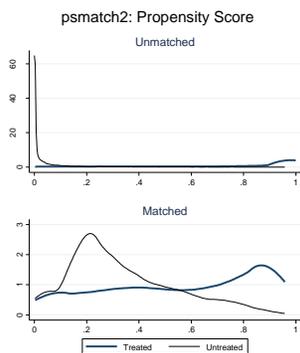
Apêndice

Figura 1 - Construção dos grupos de tratamento e controle



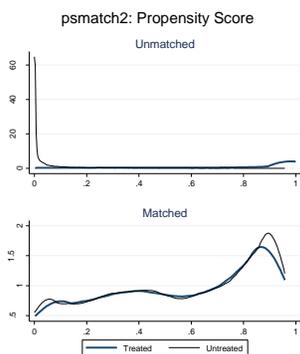
Fonte: Elaboração Própria com base nos Microdados do SPAECE 2011, 2012; Censo Escolar 2011 e ENEM 2014

Gráfico A.1.1 - Sobreposição do propensity Score pelo vizinho mais próximo



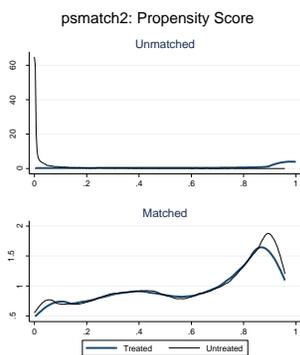
Fonte: Elaboração própria

Gráfico A.1.2 - Sobreposição do propensity Score pelo kernel



Fonte: Elaboração própria

Gráfico A.1.3 - Sobreposição do propensity Score pelo Radius



Fonte: Elaboração própria

Tabela A.1.1 - Estimação do Propensity Score pós Lasso

Variáveis	Coefficientes	p-valor	Variáveis	Coefficientes	p-valor
V01	-1.7223	0.000	V21	-0.3718	0.000
V02	-0.7620	0.000	V22	-1.4898	0.000
V03	-0.5714	0.000	V23	-0.9858	0.000
V04	-1.2251	0.000	V24	1.6798	0.000
V05	-0.2555	0.000	V25	-1.5219	0.000
V06	0.0040	0.000	V26	-1.4398	0.000
V07	-0.0001	0.000	V27	3.4337	0.401
V08	-5.53e-06	0.000	V28	-0.2325	0.138
V09	-0.0010	0.000	V29	0.5187	0.000
V10	0.8466	0.000	V30	0.0001	0.100
V11	-0.8677	0.000	V31	-0.0003	0.000
V12	-4.4204	0.000	V32	0.0019	0.203
V13	0.5045	0.000	V33	-0.0017	0.290
V14	0.0655	0.385	V34	-0.7624	0.000
V15	0.1268	0.263	V35	-0.3802	0.000
V16	0.1752	0.173	V36	-0.3058	0.037
V17	-0.0581	0.544	V37	0.5767	0.000
V18	0.2035	0.019	V38	-0.2440	0.001
V19	2.0868	0.000	V39	0.0012	0.000
V20	2.0451	0.000	Constant	-3.7773	0.000

Fonte: Elaboração Própria

Tabela A.1.2 - Análise de sensibilidade (Rosenbaum Bounds) para as áreas de conhecimento e a média geral do ENEM

Γ	CH	CN	MT	LC	R	MG
	p^+	p^+	p^+	p^+	p^+	p^+
1	0.000	0.000	0.000	0.000	0.000	0.000
1.05	0.000	0.000	0.000	0.000	0.000	0.000
1.1	0.000	0.000	0.000	0.000	0.000	0.000
1.15	0.000	0.000	0.000	0.000	0.000	0.000
1.2	0.000	0.000	0.000	0.000	0.000	0.000
1.25	0.000	0.000	0.000	0.000	0.000	0.000
1.3	0.000	0.000	0.000	0.000	0.000	0.000
1.35	0.000	0.000	0.000	0.000	0.000	0.000
1.4	0.000	0.000	0.000	0.000	0.000	0.000
1.45	0.000	0.000	0.000	0.000	0.000	0.000
1.5	0.000	0.000	0.000	0.000	0.000	0.000
1.55	0.000	0.000	0.000	0.000	0.000	0.000
1.6	0.000	0.000	0.000	0.000	0.000	0.000
1.65	0.000	0.000	0.000	0.000	0.000	0.000
1.7	0.000	0.000	0.000	0.000	0.000	0.000
1.75	0.000	0.000	0.000	0.000	0.000	0.000
1.8	0.000	0.000	0.000	0.000	0.000	0.000
1.85	0.000	0.000	0.000	0.000	0.000	0.000
1.9	0.000	0.000	0.000	0.000	0.000	0.000
1.95	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000

Fonte: Elaboração Própria

CAPÍTULO 2

DECOMPOSIÇÃO QUANTÍLICA INCONDICIONAL DOS DIFERENCIAIS DE DESEMPENHO ENTRE ENSINO PROFISSIONALIZANTE E PRIVADO

Introdução

De acordo com informações do Ministério da Educação (MEC) 2016, o Brasil registrou melhoras no Índice de Desenvolvimento da Educação Básica (IDEB) para o ensino fundamental, porém, para o ensino médio, este permaneceu estagnado, considerando todas as redes de ensino no ano de 2015. No entanto, apesar dessa evolução, apenas o ensino fundamental I conseguiu atingir a meta, superando-a em 0,3 pontos.

No Ceará, considerando escolas públicas e privadas, apenas os anos iniciais e finais do ensino fundamental atingiram a meta proposta para 2015, registrando desempenho de 5,9 e 4,8 pontos, respectivamente. O mesmo resultado pode ser observado para o ensino público separadamente, porém, com IDEB de 5,7 para o fundamental I e 4,5 pontos, para o fundamental II, respectivamente, superando a meta em 1,5 e 0,5, respectivamente. Mesmo estando abaixo do esperado para 2015, o ensino privado registra um IDEB superior ao do ensino público, porém, este último atingiu melhor evolução educacional no ensino fundamental em relação a 2013, reduzindo o diferencial de desempenho ainda existente entre as duas redes de ensino.

Estudos que tratam do diferencial de desempenho entre escolas públicas e privadas evidenciam uma superioridade do ensino particular, mesmo após controlar um conjunto de características pessoais, socioeconômica e escolaridade dos pais de alunos (FRANÇA; GONÇALVES, 2010; MORAES; BELUZZO, 2014; COSTA; ARRAES 2014).

É comum na literatura trabalhos que comparam os diferenciais de notas de testes padronizados entre alunos provenientes das escolas públicas com os alunos advindos de escolas privadas. Buscando mensurar as diferenças de desempenho entre esses tipos de escolas na fase inicial do ensino fundamental, França e Gonçalves (2010) estimam o efeito do ensino privado sobre os resultados de alunos da quarta série, a partir da aplicação do método dos Mínimos Quadrados Ponderados por *Propensity Score* aos dados do SAEB/2003. De acordo com os resultados encontrados pelos autores, há uma grande diferença de notas entre

os dois setores de ensino, sendo que, o desempenho acadêmico de alunos da rede privada é superior em 0,9 desvios padrão, aproximadamente, mesmo após controlar um conjunto de características socioeconômicas, escolares e regionais.

Costa e Arraes (2014) questionam os resultados de estudos que consideram as suposições de identificação da metodologia de pareamento no escore de propensão e regressões lineares na investigação da diferença de resultados entre as escolas públicas e privadas. Para tanto, utilizam o método de identificação parcial de Manski (1989), o qual permite estimar limites com suposições menos restritivas e verificar se as estimações pontuais das metodologias usuais não ultrapassam os limites do intervalo. A partir de informações do SAEB 2005 para estudantes do 5º e 9º anos do ensino fundamental, os autores comprovaram que essas metodologias usuais sobreestimam o efeito das escolas privadas, porém esse efeito permanece positivo e significativo.

Moraes e Belluzzo (2014) utilizaram a decomposição por quantis de Melly (2005) para a distribuição condicional de notas, a partir da base de dados do SAEB 2005 para os alunos do 9º ano do ensino fundamental. Os autores encontraram que o diferencial é favorável às escolas privadas, em todos os quantis da distribuição, chegando a 1,5 erro padrão em torno do 600 percentil. Ademais, a análise do contrafactual indica que o desempenho das escolas particulares é relativamente menor na cauda inferior da distribuição.

Depreende-se dos trabalhos supracitados que os alunos são influenciados por diversos fatores não controlados, tornando-os bastante heterogêneos, razão da forte explicação de fatores não observáveis. Ao contrário, este estudo busca reduzir a influência desses fatores ao se construir grupos de alunos mais homogêneos para se aferir o diferencial de rendimento escolar líquido, notadamente, entre os alunos provenientes de escolas públicas profissionalizantes e privadas no Ceará, a partir dos resultados destes no ENEM 2014. Vale ressaltar que, os autores não encontraram trabalhos na literatura que analisem o diferencial de notas entre alunos do ensino médio das escolas públicas e privadas, impedindo que seja feita uma comparação.

A pertinência para tal escolha prende-se ao fato de as escolas de ensino profissional apresentarem desempenho superior às demais escolas públicas, uma vez que os estudantes ingressantes passam por um processo seletivo mais rigoroso, conseqüentemente, acredita-se que os melhores alunos provenientes do ensino fundamental são selecionados, além de

possuírem estruturas mais adequadas e especializadas para proporcionar um ensino de melhor qualidade. Com isso, pretende-se responder aos seguintes questionamentos: Esse novo modelo de escola pública provê uma estrutura de ensino equiparável ao da escola privada? Como os alunos dos dois tipos de escola se comparam ao longo da distribuição de rendimento escolar? Quais fatores observáveis mais influenciam os desempenhos dos alunos das duas escolas? Como os fatores não observáveis se comparam quanto ao diferencial de rendimentos entre os alunos?

Além de propor uma nova construção para o grupo das escolas públicas, este trabalho buscou inovar ao utilizar o método de Regressão Quantílica Incondicional (RQI) proposto por Firpo *et al* (2007), o qual é considerada uma generalização do método de Oaxaca-Blinder, porém, emprega regressões com base na função de influência recentrada (FIR) combinadas com um procedimento de reponderação inspirado em DiNardo et al (1996). Esta foi escolhida em detrimento do método tradicional de Regressão Quantílica (RQ) condicional proposto por Koenker e Bassett (1978), pois este não permite interpretar as estimativas como o impacto marginal da variável explicativa sobre a variável dependente (FIRPO et al, 2007; FIRPO et al, 2009).

Em sequência, o capítulo está organizado com as seguintes seções: método de decomposição de Firpo et al. (2007), resultados do modelo e conclusões.

2 Método de Decomposição

Quando se pretende realizar uma decomposição do tipo Oaxaca-Blinder em diferentes quantis ao longo da distribuição, não se pode simplesmente executar RQ nos quantis de interesse e usar os coeficientes estimados no método OB, pois, diferentemente dos obtidos via Mínimos Quadrados (MQ), que possuem interpretação condicional e incondicional, os coeficientes estimados por regressões quantílicas têm apenas a interpretação condicional.

Sabe-se que o método MQ estima os coeficientes (β) de uma média condicional, $E[Y|X] = X\beta$, da qual interpreta-se β como $\partial E[Y|X]/\partial X$. Aplicando a lei das expectativas iteradas, $E_X[E[Y|X]] = E[Y] = E[X]\beta$, β também pode ser explicado como o efeito das alterações na média de X na média incondicional de Y, porém essa lei não é válida para quantis, ou seja, se aplicar a esperança condicional em quantil τ específico, $Q_\tau(X)$, não se

encontra o quantil incondicional, $E_X[Q_\tau(X)] \neq Q_\tau$. Consequentemente, os coeficientes da regressão linear quantílica $Q_\tau(X) = X\beta$, tem somente interpretação condicional, de modo que não se pode interpretá-lo como o efeito das alterações na média de X no quantil τ th marginal (incondicional).

Dessa forma, RQ fornece uma abordagem dos impactos diferenciados de covariáveis ao longo da distribuição de um resultado, assim, esta é utilizada para avaliar o efeito de uma variável em um quantil do resultado condicional a valores específicos de outras variáveis. Porém, segundo Borah e Basu (2013), na maioria dos casos, RQ pode gerar resultados que muitas vezes não são generalizáveis ou interpretável.

Na média, é suficiente estimar somente o valor médio condicional nos quantis de interesse, embora todo o contrafactual da distribuição acumulada incondicional, $F_{Y_B^A}(Y)$, deve ser estimada, e um contrafactual quantil incondicional de interesse pode ser recuperado por inversão, $Q_{\tau,B}^A = F_{Y_B^A}^{-1}(\tau)$, no qual os subscritos A e B, representam os dois grupos em comparação.

Vários procedimentos para a obtenção de distribuição contrafactual $F_{Y_B^A}(Y)$ tem sido sugeridos na literatura. Juhn, Murphy e Pierce (1993), Machado e Mata (2005) e Melly (2005) visam substituir cada resultado em B com um contrafactual Y_B^A . Juhn, Murphy e Pierce (1993) utilizam uma abordagem de imputação residual, enquanto Machado e Mata (2005) se baseia em simulações e regressões Quantil condicional. DiNardo et al. (1996) propuseram uma abordagem semiparamétrica reponderada, que usa um fator de reponderação estimado de "impor" as características dos indivíduos do grupo B sobre os indivíduos do grupo A.

Existem também métodos que, primeiro estimam a distribuição condicional $F_{Y_A|X_A}(Y|X)$ e depois integrá-lo sobre a distribuição dos indivíduos do grupo B, $F_{X_B}(X)$ para obter a distribuição contrafactual $F_{Y_B^A}(Y)$. Estes incluem os métodos paramétricos de Donald, Green e Paarsch (2000) e Fortin e Lemieux (1998), bem como uma abordagem menos restritiva de Chernozhukov, Fernandez-Val e Melly (2013).

Se o objetivo for apenas realizar uma decomposição agregada do quantil, ou seja, estimar os efeitos explicáveis e não explicáveis da variável de resultado sem o interesse de se fazer uma decomposição mais detalhada, qualquer um desses métodos citados anteriormente poderia ser utilizado, caso contrário seriam limitados e inapropriados. Apesar de alguns

permitirem, os procedimentos variam em grau de complexidade ou a decomposição deve ser feita sequencialmente, assim, os resultados irão depender da ordem que as características foram introduzidas. Firpo et al. (2007) propuseram um procedimento simples que permite fazer uma decomposição detalhada no quantil de interesse, cujos resultados independem da sequência. O método baseia-se na função de regressão de influência (FI) ou RQI de Firpo et al. (2009).

A RQI tem sido aplicado em diversos trabalhos na literatura, tais como: Fortin (2008), Edoka (2012), Le and Booth (2013), Borah e Basu (2013), Rubil (2013), Fisher and Marchand (2014), Hirschand Winters (2014), Kassenboehmer e Sinning (2014), Morin (2015) e Nguyen (2015). Segundo Firpo et al (2007), existem duas formas de se obter o efeito de uma covariável no quantil incondicional. Primeira, usar as estimativas dos coeficientes da RQ para recuperar o impacto incondicional, porém, apesar de intuitivamente simples, muitas vezes torna-se intratável. Especificamente, eles mostram que o efeito parcial de uma covariável em um quantil incondicional de Y pode ser escrito como uma média ponderada (através da distribuição de X) do efeito parcial sobre um quantil condicional específico de Y , que corresponde ao quantil incondicional de interesse. Portanto, se conseguir mapear todos os quantis incondicionais de Y para os quantis condicionais correspondentes sob diferentes argumentos condicionado, então tal abordagem ponderada pode ser facilmente implementada. No entanto, é evidente que isso é uma tarefa árdua, requer técnicas não paramétricas, e é muitas vezes intratável.

Uma alternativa para a abordagem acima mencionada foi proposta por Machado e Mata (2005), que assumiu uma mudança na distribuição incondicional ao longo do tempo e a decompôs em componentes que são atribuíveis a mudanças na distribuição marginal de diferentes X s. No entanto, esse método captura o efeito total de uma alteração na distribuição marginal de X sobre todos os quantis incondicionais de Y , mas não apenas em um quantil específico.

A segunda abordagem foi proposta por Firpo et al. (2007), o qual soluciona o problema de intratabilidade e supera a limitação do modelo condicional. Eles sugerem um modelo RQI baseadas nos conceitos de Função de Influência (FI) e na Função de Influência Recentrada (FIR), tal como utilizado na robusta literatura estatística (HAMPEL et al., 1986). Uma FI é uma ferramenta analítica que pode ser usada para avaliar o efeito (influência) de uma

remoção/adição de uma observação sobre o valor de uma estatística, $v(F)$, sem ter que recalculá-la a estatística, e é definida por:

$$FI(y, v(F)) = \lim_{\varepsilon \rightarrow 0} \left\{ \frac{v[(1 - \varepsilon) \cdot F + \varepsilon \cdot \delta_y] - v(F)}{\varepsilon} \right\}, 0 \leq \varepsilon \leq 1 \quad (1)$$

Onde F representa a função de distribuição acumulada de Y e δ_y é uma distribuição que agregada no valor y .

Uma FIR é obtida adicionando a estatística $v(F)$ a FI:

$$FIR(y, v) = v(F) + FI(y, v) \quad (2)$$

Uma característica dessa abordagem é que a expectativa é igual a $v(F)$, ou seja, se a estatística de interesse for a média, então, a FI será o residual avaliado em y e a FIR será o próprio valor de y :

$$FI(y, \mu) = \lim_{\varepsilon \rightarrow 0} \left\{ \frac{(1 - \varepsilon) \cdot \mu + \varepsilon \cdot y - \mu}{\varepsilon} \right\} = y - \mu, e FIR(y, \mu) = \mu + y - \mu = y \quad (3)$$

Por conseguinte, a regressão da FIR para a média, em X , produz os mesmos coeficientes da regressão por MQ.

Quando a estatística de interesse é um quantil τ específico da distribuição:

$$FI(y, q_\tau) = \frac{\tau - I\{Y \leq q_\tau\}}{f_Y(q_\tau)} \quad (4)$$

Onde q_τ se refere ao quantil τ th da distribuição incondicional de Y , $f_Y(q_\tau)$ é a função densidade de probabilidade de Y avaliado no τ –ésimo quantil (q_τ) e $I\{Y \leq q_\tau\}$ é uma variável indicadora que delimita a ocorrência de Y até o limite do quantil. Por definição,

$$FIR(y, q_\tau) = q_\tau + FI(y, q_\tau) \quad (5)$$

Quando a esperança condicional da $FIR(y, q_\tau)$ é modelada como uma função das variáveis explicativas, $X, E[FIR(y, q_\tau)|X = x] = m_\tau(x)$, uma regressão FIR pode ser vista como uma RQI (FIRPO et al., 2009). Por definição da FIR, $E_X E[FIR(y, q_\tau)|X] = q_\tau$, então $E_X(dm_\tau(x)/dX)$ pode ser interpretado como efeito marginal de uma pequena variação na distribuição das covariadas no τ quantil incondicional de Y , mantendo tudo o mais constante.

Firpo et al. (2007) afirmam que os procedimentos para estimação da RQI são simples e semelhantes à regressão MQ e se dá da seguinte forma: Primeiro, para um quantil τ específico,

estima-se a RIF do quantil τ de Y seguindo (3.2.4) e (3.2.5). q_τ é estimado usando a estimativa da amostra do quantil τ incondicional, assim como, a densidade $f_y(q_\tau)$ no ponto q_τ é estimada usando método de kernel. Segundo, aplica-se MQ na $FIR(y, q_\tau)$ sobre as covariadas observadas, X . Os autores também destacam as medidas para calcular o efeito quantílico incondicional, o qual mede o impacto de uma covariável sobre Y num quantil específico, onde, na regressão FIR-MQO, pressupõe-se que, os efeitos parciais quantílicos incondicionais são os coeficientes estimados¹⁴.

Dessa forma, enquanto a regressão quantílica condicional provê estimativas do retorno de características individuais, *coeteris paribus*, onde, este retorno é variável entre os indivíduos de acordo com o quantil incondicional ao qual ele pertence, a regressão quantílica incondicional estima, *coeteris paribus*, o efeito de pequenas mudanças em uma característica dos indivíduos em cada quantil da distribuição, permitindo assim, avaliar o efeito sobre diversas estatísticas da distribuição (FOURNIER; KOSKE, 2012).

Uma característica do método RQI é que os resultados podem ser diretamente aplicados ao modelo de decomposição de Oaxaca-Blinder¹⁵, o qual será utilizado por este trabalho para verificar os fatores que contribuem para os diferenciais de notas no ENEM entre os alunos das escolas públicas profissionalizante e os estudantes das escolas privadas do Ceará ao longo de toda a distribuição, definidos genericamente como A e B, aplica-se um método de decomposição proposto por Firpo et al. (2007). Generalizando a decomposição de Oaxaca-Blinder, pode-se representar a diferença nas distribuições de notas entre os dois tipos de escolas por meio de alguma estatística dessas distribuições. Assim, o diferencial total entre as distribuições é escrito como:

$$\Delta^v = v(F_{yA}) - v(F_{yB}) \quad (6)$$

Onde $v(F_{yr})$ representa uma estatística da distribuição de notas para a escola $r = A, B$. A decomposição irá dividir Δ^v em um componente associado a diferenças de características observáveis dos indivíduos e um componente associado a diferenças na estrutura de notas entre os dois tipos de escolas. Para tanto, faz-se necessário a construção de um contrafactual (F_{yC}) para que se possa simular uma distribuição de notas compondo a estrutura de notas das

¹⁴Podem ser implementados pelo software STATA a partir do comando *rifreg*. Disponível no site: <http://faculty.arts.ubc.ca/nfortin/datahead.html>

¹⁵ A regressão FIR quantílica pode ser usada na decomposição de Oaxaca através do comando *oaxaca8* no software STATA.

escolas A e a distribuição de características (observáveis e não observáveis) dos alunos da escola B. Construindo esse contrafactual e algumas manipulações algébricas pode-se obter:

$$\Delta^v = [v(F_{yB}) - v(F_{yC}) + [v(F_{yC}) - v(F_{yA})]] \quad (7)$$

$$\Delta^v = \Delta_S^v + \Delta_X^v \quad (8)$$

O primeiro termo constitui a *parte não-explicada* da decomposição – efeito coeficiente (retorno) ou discriminação quando se refere a mercado de trabalho, definido como uma medida da diferença entre os coeficientes estimados das covariadas para o grupo B em relação ao A. O último termo deste diferencial total é o efeito quantidade, considerado a *parte explicada* da decomposição – efeito dotação ou característico, o qual refere-se à diferença nas dotações dos dois grupos, onde a diferença de nota entre os alunos das escolas públicas profissionalizantes frente aos das privadas pode ser parcialmente atribuída ao fato de possuírem características mais favoráveis para atingir notas mais elevadas. Aplicando o valor esperado das funções de influencia recentrada e assumindo uma especificação linear, tem-se as equações (9) e (10) para as distribuições observadas e para a distribuição contrafactual, respectivamente.

$$v(F_{yB}) = E[FIR(y_r; v_r)|X, R = r] = X_r \beta_r \quad \text{para } r = A, B \quad (9)$$

$$v(F_{yC}) = E[FIR(y_A; v_C)|X, R = B] = X_C \beta_C \quad (10)$$

Ao aplicar a decomposição de OB sob a suposição de linearidade, a expectativa da estatística contrafactual em que se adota a distribuição de características da escola B e a estrutura de notas da escola A é estimada de forma que $\bar{X}_C \beta_C = \bar{X}_B \beta_A$. Porém, quando a linearidade não se mantém, essa igualdade passa a não ser válida.

A partir de Barsky et al. (2002), Firpo et al. (2007) propõe uma abordagem que combina o método de regressões FIR com reponderações inspirada na técnica de DiNardo et al. (1996). Esse método estabelece o uso de um fator de reponderação, o qual pode ser definido por:

$$\varphi(X) = \frac{\Pr[R = B|X]}{\Pr[R = A|X]} \cdot \frac{\Pr[R = A]}{\Pr[R = B]} \quad (11)$$

Onde $\Pr[R = r|X]$ é a probabilidade de um aluno pertencer a escola r dados as suas características X e $\Pr[R = r]$ denota a proporção de alunos na escola r. vale ressaltar que esse fator de ponderação pode ser estimado através de um modelo de probabilidade para $\Pr[R = B|X]$, e usando as probabilidades previstas para calcular o valor $\widehat{\varphi(X)}$ para cada observação.

$$\widehat{\beta}_r = (\sum_{i \in r} \widehat{w}_r \cdot X_i \cdot X_i')^{-1} \cdot \sum_{i \in r} \widehat{w}_r \cdot \widehat{FIR}(y_{ri}; v_r) \cdot X_i, \text{ para } r = A, B \quad (12)$$

Onde \widehat{w}_r é um fator de ponderação baseado na própria composição da amostra total. Para o contrafactual, estima-se

$$\widehat{\beta}_C = (\sum_{i \in A} \widehat{\varphi}(X_i) \cdot X_i \cdot X_i')^{-1} \cdot \sum_{i \in A} \widehat{\varphi}(X_i) \cdot \widehat{FIR}(y_{Ai}; v_C) \cdot X_i \quad (13)$$

$$\overline{X}_C = \sum_{i \in A} \widehat{\varphi}(X_i) \cdot X_i \quad (14)$$

A decomposição é dada por:

$$\widehat{\Delta}^v = [\overline{X}_B \widehat{\beta}_B - \overline{X}_C \widehat{\beta}_C] + [\overline{X}_C \widehat{\beta}_C - \overline{X}_A \widehat{\beta}_A] \quad (15)$$

$$\widehat{\Delta}^v = \widehat{\Delta}_S^v + \widehat{\Delta}_X^v \quad (16)$$

A estimativa do efeito composição, $\widehat{\Delta}_S^v$, pode ser dividida em dois componentes em que um termo padrão $(\overline{X}_C - \overline{X}_A) \widehat{\beta}_A$ é somado a um erro de especificação $\overline{X}_C (\widehat{\beta}_C - \widehat{\beta}_A)$, de forma que:

$$\widehat{\Delta}_X^v = (\overline{X}_C - \overline{X}_A) \widehat{\beta}_A + \overline{X}_C (\widehat{\beta}_C - \widehat{\beta}_A) \quad (17)$$

O erro de especificação tende a ser nulo, se a especificação linear do modelo for correta, uma vez que $plim(\widehat{\beta}_C) = plim(\widehat{\beta}_A)$. Na decomposição detalhada, em que é possível estimar a contribuição de cada variável explicativa, o efeito composição pode ser escrito como:

$$\widehat{\Delta}_S^v = \sum_{k=1}^K (\overline{X}_{Ck} - \overline{X}_{Ak}) \widehat{\beta}_A \quad (18)$$

O efeito estrutural das notas passa a ser definido por:

$$\widehat{\Delta}_B^v = \overline{X}_B (\widehat{\beta}_B - \widehat{\beta}_C) + (\overline{X}_B - \overline{X}_C) \widehat{\beta}_C \quad (19)$$

Esse efeito se reduz ao primeiro termo, uma vez que, em grandes amostras e com a estimação correta do fator de ponderação, o erro de reponderação tende a ser nulo. Vale ressaltar que, o método de regressões quantílicas incondicionais fornece uma estimativa de equilíbrio parcial da variável de interesse, assumindo que mudanças marginais em X não exercem impacto sobre a distribuição conjunta de X e y. Além disso, o método também assume a forte independência entre heterogeneidade não observada e as características observadas. Segundo Fournier e Kosque (2012), embora estas suposições não se sustentem na prática, mas argumentam que uma comparação entre estimativas de diferentes quantis permanece válida, desde que o viés potencial seja o mesmo em toda a população da amostra.

2.2.1 Base de Dados

As informações referentes ao desempenho nas cinco áreas de conhecimento e na média geral do ENEM 2014¹⁶ às condições socioeconômicas dos alunos das EEEP da rede estadual de educação e das escolas privadas do Ceará foram retiradas do INEP. A identificação das escolas profissionalizantes e suas dotações de infraestrutura de todas as escolas foram obtidas da SEDUC e Censo Escolar/2014, respectivamente.

A amostra utilizada por este trabalho é constituída apenas pelos alunos com as seguintes características: estão cursando o ensino médio das escolas de ensino profissionalizante e privado do estado¹⁷; concluíram em 2014¹⁸; estiveram presentes nos dias de provas do ENEM do referente ano; suas escolas possuem informações disponibilizadas pelo censo de 2014. Dessa forma, obteve-se um total de 20.947 alunos, sendo 8.955 da rede estadual de ensino profissional e 11.992 das escolas privadas.

2.2.2 Variáveis do Modelo

As variáveis, descritas no quadro 2.1, foram escolhidas com o intuito de captar os diferenciais nas notas das cinco áreas e na média geral do ENEM, compostas por vetores com as características pessoais dos alunos, status educacional dos pais e as condições de infraestrutura das escolas.

A variável que reflete a renda familiar está disponível em intervalos de salários mínimos, optou-se por fazer um corte de dois salários, tendo em vista que são pais de alunos de escolas públicas, pressupondo-se assim, não possuírem renda familiar muito elevada. Para facilitar a interpretação da medida dos índices CSE e Infra, as variáveis foram normalizadas, de forma a assumir valores entre 0 a 1; quanto maior o índice, mais favorável na disposição de equipamentos.

¹⁶Último ano com informações disponíveis sobre o desempenho individual no ENEM.

¹⁷ Foram excluídos os alunos das escolas estaduais de ensino regular, municipais e federais.

¹⁸ Foram excluídos os alunos que já concluíram o ensino médio ou que não irão concluir em 2014.

Quadro 2.1 -Descrições das Variáveis

Variáveis	Descrição
Dependentes: Notas logaritimizadas do aluno em cada uma das áreas: Ciências da Natureza (CN), Ciências Humanas (CH), Linguagem e Códigos (LC), Matemática (MT), Redação (R), Média Geral (MG).	
Explicativas	
Alunos	
Priv	1 se o aluno pertence a escola privada; 0 se o aluno pertence a EEEP
Idade	Anos de idade
Sexo	1 se masculino; 0 caso contrário (c.c)
Raça	1 se branca; 0 c.c
Estado civil	1 se solteiro; 0 c.c
Anos_Fund	1 se concluiu o ensino fundamental em 8 anos; 0 c.c
Anos_Méd	1 se concluiu o ensino médio em 3 anos; 0 c.c
Escolaridade_Pai	
E_Pai1	1 se o pai não estudou; 0 c.c
E_Pai2	1 se o pai possui de escolaridade o ensino fundamental incompleto ou completo; 0 c.c
E_Pai3	1 se o pai possui de escolaridade o ensino médio incompleto ou completo; 0 c.c
E_Pai4	1 se o pai possui de escolaridade o ensino superior incompleto ou completo; 0 c.c
Escolaridade_Mãe	
E_mae1	1 se a mãe não estudou; 0 c.c
E_mae2	1 se a mãe possui de escolaridade o ensino fundamental incompleto ou completo; 0 c.c
E_mae3	1 se a mãe possui de escolaridade o ensino médio incompleto ou completo; 0 c.c
E_mae4	1 se a mãe possui de escolaridade o ensino superior incompleto ou completo; 0 c.c
Condições Socioeconômicas	
Renda00	1 se a família não possui renda; 0 c.c
Renda01	1 se a família possui até 1 salário mínimo; 0 c.c
Renda02	1 se a família possui de 1 até 2 salário mínimo; 0 c.c
Renda03	1 se a família possui de 2 até 5 salário mínimo; 0 c.c
Renda04	1 se a família possui de 5 até 10 salário mínimo; 0 c.c
Renda05	1 se a família possui de 10 até 15 salário mínimo; 0 c.c
CSE*	Índice de condição socioeconômica
Escolas	
Infra**	Índice de infraestrutura da escola

Fonte: Elaboração própria

Notas: (*) Índice criado a partir da análise de componente principal que engloba a existência das seguintes características dos alunos obtidas pelo ENEM 2013: Se possui TV, DVD, rádio, microcomputador, automóvel, máquina de lavar, geladeira, freezer, telefone fixo, celular, acesso à internet, TV por assinatura, aspirador de pó, empregada, banheiro.

(**)Índice gerado pela análise de componentes principais que engloba a existência das seguintes características das escolas obtidas no Censo Escolar 2013: abastecimento de água por rede pública, de esgoto sanitário por rede pública, de coleta periódica de lixo, de sala de diretoria, de sala de professores, de laboratório de informática e de ciências, de quadra de esportes, de cozinha, de biblioteca, de parque infantil, de sanitário adequado a aluno portador de necessidades especiais (PNEs), de dependências e vias adequadas a PNEs, de televisão, de videocassete, de DVD, de parabólica, de copiadora, de retroprojeto, de impressora, de computadores, de acesso à internet, e de alimentação escolar para os alunos.

3 Resultados

3.1 Análise descritiva

Estatísticas descritivas das notas nas cinco áreas e na média geral do ENEM, das variáveis de controles relacionadas aos alunos de EEEP e privadas do Ceará em 2014 estão explicitadas na Tabela abaixo.

Tabela 2.1 - Estatísticas Descritivas das variáveis da amostra – 2014

Variáveis	EEEP			Privadas		
	Media	Desvio-padrão	Coefficiente de variação	Media	Desvio-padrão	Coefficiente de variação
NOTA_CN	493.58	63.7490	0.1291	545.82	85.8326	0.1572
NOTA_CH	556.17	63.8829	0.1148	593.56	70.9712	0.1195
NOTA_LC	519.56	58.9450	0.1134	552.23	64.4928	0.1167
NOTA_MT	486.98	95.5583	0.1962	558.90	135.454	0.2423
NOTA_R	549.30	139.2176	0.2534	633.27	158.1479	0.2497
NOTA_MG	521.12	64.5028	0.1237	576.76	85.6248	0.1484
Alunos						
Idade	17.467	0.8507	20.532	17.643	2.6064	6.7691
Sexo	0.4240	0.4942	1.1655	0.4654	0.4988	1.0716
Raça	0.1934	0.3949	2.0422	0.4237	0.4941	1.1661
Estado civil	0.9946	0.0730	0.0734	0.9908	0.0953	0.0962
Anos_Fund	0.3794	0.4852	1.2788	0.4590	0.4983	1.0855
Anos_Médio	0.3570	0.4791	1.3421	0.5270	0.4992	0.9473
Escolaridade_Pai*						
E_Pai1	0.0692	0.2538	3.6667	0.0149	0.1212	8.1240
E_Pai2	0.5786	0.4937	0.8533	0.2317	0.4219	1.8208
E_Pai3	0.2944	0.4558	1.5479	0.3940	0.4886	1.2399
E_Pai4	0.0501	0.2182	4.3527	0.2655	0.4416	1.6629
Escolaridade_Mãe**						
E_mae1	0.0279	0.1647	5.9011	0.0068	0.0824	12.0522
E_mae2	0.4943	0.4999	1.0113	0.1571	0.3639	2.3156
E_mae3	0.3563	0.4789	1.3440	0.3922	0.4882	1.2447
E_mae4	0.0904	0.2868	3.1712	0.2952	0.4561	1.5449
Condições Socioeconômicas***						
Renda00	0.0155	0.1236	7.9643	0.0049	0.0699	14.2223
Renda01	0.6391	0.4802	0.7513	0.1323	0.3388	2.5606
Renda02	0.2797	0.4488	1.6047	0.2635	0.4405	1.6718
Renda03	0.0599	0.2374	3.9595	0.2990	0.4578	1.5311
Renda04	0.0052	0.0722	13.7677	0.1752	0.3802	2.1692
Renda05	0.0002	0.0149	66.9220	0.0617	0.2407	3.8967
CSE	0.4806	0.1461	0.3040	0.6740	0.1847	0.2740
Infra_escola	0.3547	0.0903	0.2546	0.4144	0.1490	0.3597

Fonte: Elaboração Própria

Nota:(*) A proporção de alunos com pai que possui pós-graduação (E_pai5) foi 0.007 para as profissionalizantes e 0.093 para as privadas, perfazendo 100% quando somada as demais. (**) A proporção com mãe que possui pós-graduação (E_mae5) foi 0.030 para as profissionalizantes e 0.148 para as privadas, perfazendo 100% quando somada as demais. (***) A proporção de alunos com renda familiar maior que 15 salários mínimos (renda_6) foi 0.0001 para as profissionalizantes e 0.0631 para as privadas, perfazendo 100% quando somada as demais.

Observa-se que, embora as médias das notas dos alunos das escolas de ensino privado serem superiores às dos alunos de escolas profissionalizantes em todas as provas, há que se considerar também a variabilidade dessas notas. Em Ciências da Natureza (CN), o diferencial de médias é de 52.24 pontos a favor das escolas privadas, embora estas apresentem uma variação relativa à média 21,7% ($0,1572/0,1291$) superior à EEEP. As áreas Ciências Humanas (CH) e Linguagem e Códigos (LC) apresentam diferença semelhante, 37,39 e 32,67 pontos, respectivamente, com dispersão relativa à média de 4% e 3% superior à EEEP, respectivamente.

Apresentando diferenciais mais elevados e menos homogêneo, estão as provas de Matemática (MT) e Redação (R), com diferença de 71,92 e 83,97 pontos, respectivamente, e variação relativa semelhante nos dois tipos de escolas, e 23,5%, em MT, superior à EEEP. Na Média Geral, os alunos das escolas particulares apresentaram desempenho superior em, aproximadamente, 56 pontos, 11% a mais que os alunos das escolares de ensino profissionalizante, porém, observa-se que, as escolas particulares apresentam variação relativa à média superior à EEEP de, aproximadamente 20%.

Os gráficos no Apêndice A mostram a distribuição das proficiências dos alunos nas cinco áreas e na média geral, para as escolas particulares e profissionalizantes, representados pelas linhas contínuas e pontilhadas, respectivamente. Através destes, pode-se observar que o gráfico para os alunos do ensino privado encontra-se mais concentrado para a direita, indicando que a frequência de alunos com notas maiores é superior aos alunos pertencentes ao ensino profissional.

Referente as demais variáveis presentes na Tabela 2.1, observa-se que apesar de apresentarem idade média semelhante, 17 anos, o grupo das escolas privadas é menos homogêneo ao das escolas profissionalizantes, pois apresenta variação absoluta de aproximadamente 2 anos, enquanto que no outro grupo, o desvio-padrão é de meses.

Além disso, a proporção de homens e alunos solteiros são semelhantes, nas EEEP e privadas, 42% e 46% são do sexo masculino, respectivamente, e em ambos os grupos, 99%

são solteiros. Quando se refere a raça, a diferença entre os percentuais aumenta, aproximadamente, 19% são considerados brancos nas profissionalizantes e 42% nas escolas particulares.

As variáveis Anos_Fund e Anos_Médio apresentam a proporção de alunos que terminaram o ensino fundamental em 8 anos e o Médio em 3 anos, os quais, para as profissionalizantes, foram de 38% e 36%, e para as particulares, 46% e 53%, respectivamente. Isso mostra indícios que os alunos que fizeram o 3 ano nas escolas profissionalizantes apresentam problemas de repetência, maiores que as particulares.

Ao analisar as escolaridades do pai e da mãe, observa-se que, os alunos das escolas privadas apresentam pais com níveis mais elevados de estudo. Para as EEEP, a maior proporção reside em pai e mãe com fundamental incompleto ou completo, 57,86% e 49,43%, respectivamente. Enquanto que, para as escolas particulares, pais com ensino médio incompleto ou completo são maioria, aproximadamente 39%. Ademais, a proporção de pais com ensino superior nas escolas de ensino privado é maior do que no grupo das profissionalizantes, aproximadamente 20 pontos percentuais a mais.

Através das variáveis de condições socioeconômicas, observa-se que a renda familiar dos alunos das escolas profissionalizantes concentra-se até dois salários mínimos, diferentemente dos alunos das escolas privadas, cuja concentração se dá a partir de três salários. Além disso, estes alunos apresentam índice mais elevados que os alunos das EEEP, 0.6740 e 0.4806, respectivamente. Apesar de as EEEP apresentarem boa infraestrutura (RELATÓRIO DE GESTÃO, 2014), as escolas particulares detém uma melhor oferta de equipamentos aos alunos, conforme atestam os respectivos índices com valores 0.3547 e 0.4144.

3.2 Regressão FIR nos Quantis Incondicionais

Para decompor os diferenciais de notas entre os alunos das escolas profissionalizantes e das privadas por meio da metodologia de Firpo (2007), faz necessário estimar as regressões RIF, os quais tem o objetivo de estimar efeitos das covariáveis sobre os quantis incondicionais do log das notas das provas do ENEM, cujos resultados para os quantis 0.1, 0.5 e 0.9 encontram-se nas tabelas do apêndice B. Pode-se observar que, os efeitos não são constantes ao longo da distribuição, justificando assim, a análise por regressões quantílicas, além de diferirem entre os dois tipos de escolas.

Os coeficientes estimados para a variável idade foram negativos e significativos em todos os quantis analisados, para as cinco áreas de conhecimento e na média geral, e nos dois tipos de ensino, profissionalizante e privado, indicando que, quanto maior a idade do aluno, menor será o desempenho dele no ENEM, independente da parte da distribuição no qual o aluno se encontra, porém, esse efeito decresce a medida que se caminha para parte superior da distribuição.

Os resultados mostram que o efeito da variável sexo difere entre os tipos de escolas, ao longo da distribuição e entre as áreas de conhecimento. Na média geral, Ciências Humanas, Ciências da Natureza e Matemática, considerando a mediana e o quantil 0.9, pode-se observar que, o efeito é positivo e significativo em ambas escolas, porém com efeito maior nas profissionalizantes. Enquanto que, no quantil 0.1, o efeito difere entre as áreas e entre as escolas. Na média geral, este foi insignificante para as EEEP e negativo significativo para as escolas particulares. Em CH, o efeito foi positivo e significativo para as profissionalizantes e insignificante para as privadas. Já em CN e MT, ambos foram positivos e significativos. Apresentando resultados negativos e significativos em todos os quantis analisados e nos dois tipos de escolas, encontra-se a área de Redação, porém em Linguagens e Códigos, observou-se esse tipo de efeito apenas para as escolas particulares, na mediana e no quantil 0.9.

Referente a raça, verificou-se que, no primeiro quantil, o efeito é predominantemente insignificante, exceto em redação, o qual foi positivo significativo para ambos os tipos de escolas. Na mediana, se pode observar, na MG, LC, MT e R, efeitos insignificantes para as EEEP e positivos significativos para as privadas, enquanto que, em CH e CN, os efeitos foram positivos e significantes nas EEEP, e nas escolas particulares foram insignificantes e positivo significativo, respectivamente. No quantil 0.9, esse efeito mostrou-se positivo e significativo em todas as áreas para as escolas de ensino privado e nas áreas de CH e LC para as EEEP.

Além desses resultados, observa-se que alunos cujo estado civil é solteiro não apresentam efeito significativos sobre as notas do ENEM no primeiro quantil da distribuição, na mediana, exceto em LC para as EEEP e MT para as escolas particulares, e no quantil 0.9, nas áreas CH e LC para ensino privado, e em todas as áreas nas EEEP, exceto em CH. Diferentemente da variável que indica se o aluno concluiu o ensino fundamental em oito anos, cujo efeito foi positivo e significativo em todos os quantis analisados, áreas de conhecimentos e escolas, exceto no primeiro quantil de Matemática para o ensino profissionalizante, cujo

efeito foi insignificante. Já para a variável *Anos_medio*, o efeito foi insignificante para as EEEP, em todas as áreas e quantis, exceto no quantil 0.1 em Redação.

Referente as *dummies* de escolaridade dos pais dos alunos de ensino profissionalizante, observa-se que todos os efeitos foram insignificantes, nas partes inferiores e superiores da distribuição de notas, em todas as áreas, e na mediana, exceto na variável que indica se o pai tem ensino médio incompleto ou completo para as áreas de conhecimento MG e CN, e nas variáveis *E_pai2* e *E_pai4* em CN, cujos efeitos foram positivos. Porém, nas escolas privadas observa-se, predominantemente, efeitos negativos e significativos, exceto para as variáveis *E_pai1*, nos quantis 0.9 da MG, 0.1 em CH, CN e MT, *E_pai_3* na mediana em LC e no quantil 0.1 de LC, porém para *E_pai4*, o efeito mostrou-se, na maioria dos casos, insignificante.

Resultado semelhante pôde ser observado para as escolaridades das mães dos alunos das escolas profissionalizantes, com exceção da média geral para *E_mae2* e *E_mae3*, no quantil 0.1, e na mediana, para *E_mae3* na MG e *E_mae2* em CN. Já, no ensino privado, os efeitos variaram entre insignificantes e negativos significativos, tanto por quantil, quanto por área de conhecimento.

Quanto ao efeito das faixas de renda sobre as variações nas notas do ENEM, verifica-se que, a partir da média geral, para as escolas profissionalizantes, todos os efeitos foram positivos e significativos, exceto na mediana para renda de 10 a 15 salários mínimos e no quantil 0.9 para famílias que não possuem renda e para aquelas que possuem de 5 a 15 S.M. Já, nas escolas particulares, o efeito foi predominantemente negativo e significativo, com exceção de algumas faixas de renda no primeiro quantil e na mediana.

Além disso, efeitos insignificantes puderam ser observados ao longo da distribuição para o índice de condição socioeconômica e de infraestrutura das escolas, considerando a média geral para os alunos das escolas profissionalizantes, enquanto que, nas escolas privadas, esse efeito foi positivo e significativo em todos os quantis analisados.

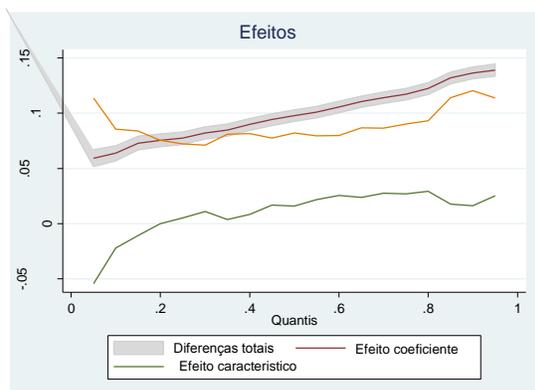
3.3 Decomposição da regressão FIR: Efeito Composição e Estrutural

A decomposição quantílica irá identificar quais os fatores, efeitos composição (característicos) ou estrutural (coeficientes), determinam os diferenciais de notas por tipo de escola (privada e EEEP) nos quantis da distribuição utilizando as estimativas do FIR. Para

tanto, este procedimento foi estimado para os 18 quantis estabelecidos, 0.05 a 0.95, cujos diferenciais de notas profissionalizantes-privadas, juntamente com os efeitos característicos e coeficientes são ilustradas suas trajetórias através dos gráficos 2.1 a 2.6, porém somente alguns quantis foram selecionados para serem expostos na tabela 2.2. Em todas as notas, as estimativas dos diferenciais mostraram-se significativas a 1%, assim, é possível verificar vantagens dos alunos das escolares particulares em relação aos alunos do ensino profissional.

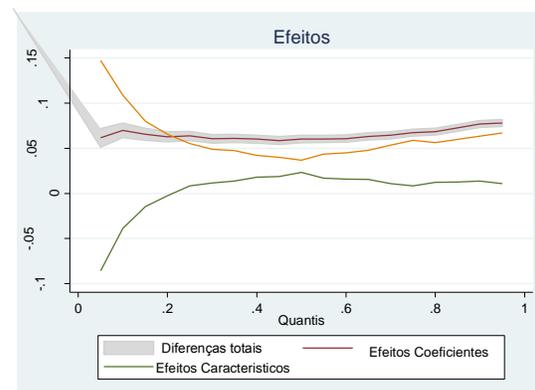
Em Matemática, Ciências da Natureza e na Média Geral, o aumento da diferença de notas na parte superior da distribuição indica o efeito *glassceiling*. Observa-se, que o diferencial total entre as notas dos alunos das escolas profissionalizantes e privadas na média geral é monotonamente crescente nos quantis. No quantil 0.1, a diferença é de 0.0637, atingindo 0.0978 na mediana e 0.1365 no quantil 0.9, ou seja, 6,57%, 10,27% e 14,62% a mais para os alunos das escolas privadas. Dessa forma, pode-se dizer que, em média, a diferença está aumentando à medida que se move a quantis mais elevados da distribuição, significando que, o diferencial de notas, a favor dos alunos das escolas particulares, é maior para aqueles que apresentam desempenho superior.

Gráfico 2.1 - Decomposição dos Efeitos – MG



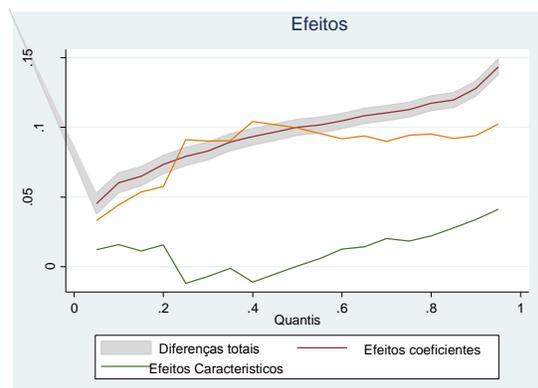
Fonte: Elaboração Própria

Gráfico 2.2 - Decomposição dos Efeitos – CH



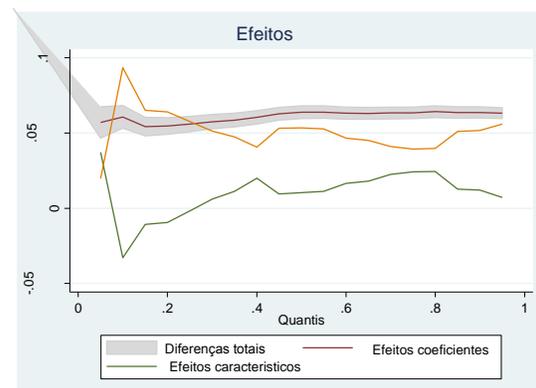
Fonte: Elaboração Própria

Gráfico 2.3 - Decomposição dos Efeitos – CN



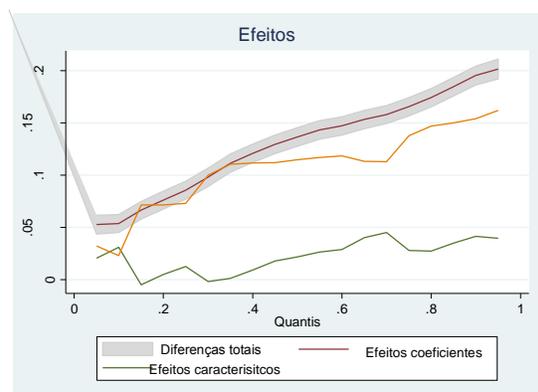
Fonte: Elaboração Própria

Gráfico 2.4 - Decomposição dos Efeitos – LC



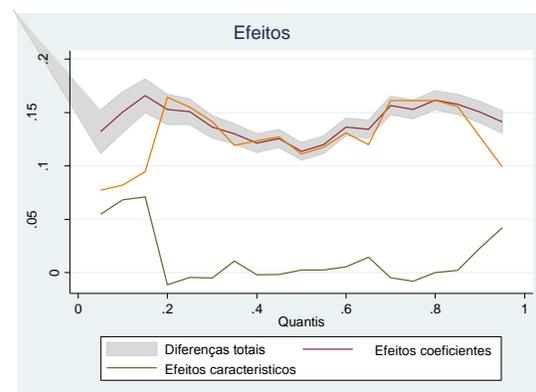
Fonte: Elaboração Própria

Gráfico 2.5 - Decomposição dos Efeitos – MT



Fonte: Elaboração Própria

Gráfico 2.6 - Decomposição dos Efeitos – R



Fonte: Elaboração Própria

Interpretação semelhante pode ser utilizada para as áreas de Ciências da Natureza e Matemática, cujo diferencial total também é positivo e crescente, porém, observa-se que, em MT o hiato é ainda mais expressivo à medida que se caminha para quantis mais elevados da distribuição. Nesta área, o quantil 0.10 apresenta uma diferença de 5,51% e atinge 21,55% no quantil 0.9, enquanto que, em CN, essas diferenças são de 6,2% e 13,63%, respectivamente.

Em Ciências Humanas e Linguagens e Códigos, apesar de apresentarem diferenciais positivos, estes praticamente permanecem constantes ao longo da distribuição, com leves oscilações. No 10°, 50° e 90° quantil, as diferenças são 0.0698, 0.0602 e 0.0767, respectivamente, para CH, e 0.0606, 0.0637 e 0.0636, respectivamente, para LC.

Apresentando diferenciais mais elevados ao longo de toda a distribuição, encontra-se a área de Redação, o qual oscila de 0.1504, equivalente a 16,22%, no quantil 0.1 para 0.1135 (12,01%) na mediana e 0.1503 (16,21%) no quantil 0.9.

Embora os alunos das escolas profissionalizantes apresentem desempenho superior quando comparados aos alunos das escolas estaduais do Ceará (MARIANO et al, 2015), ainda existem diferenciais positivos de notas em todas as áreas do ENEM e ao longo de toda a distribuição a favor dos alunos da rede privada, principalmente em Redação. A decomposição quantílica de Firpo (2007) permite identificar os fatores que contribuem para esse diferencial de notas, os quais podem ser observados através dos efeitos composição (caraterísticos) e estruturais (coeficientes).

Os resultados dos efeitos característicos para MG e CH apresentam sinais negativos e significantes nos primeiros quantis (0.05, 0.10 e 0.15) da distribuição, nos quantis 0.10 e 0.15 em LC e no percentil 0.25 em CN, indicando que estes agem no sentido de diminuir as diferenças entre as notas. Isso implica que as diferenças de notas entre as escolas em características observáveis preveem vantagens aos estudantes das EEEP nessas áreas e na parte inferior da distribuição, com destaque para o quantil 0.1 em CH e em LC, que agem reduzindo o diferencial em 56% e 54%, respectivamente, como pode ser observado na tabela 2.2 e nos gráficos acima.

Dessa forma, os alunos das escolas profissionalizantes apresentam características produtivas melhores que os alunos das escolas particulares, e o efeito positivo para o efeito estrutural sugere que os retornos das características foram maiores para os alunos das escolas particulares. Com isso, nesses quantis, o que explica os diferenciais de notas são os fatores não observáveis, pois os alunos das EEEP apresentaram características observáveis melhor do que os alunos das escolas privadas. Portanto, o sinal positivo do efeito dos efeitos estruturais foi suficiente para anular as características – efeito composição.

Os efeitos composição nos quantis 0.2, 0.25, 0.35, 0.5 para a MG, 0.2 e 0.25 para CH, 0.2, 0.25, 0.3, 0.45, 0.5 e 0.55 para LC, 0.15 a 0.6 em MT, 0.15, 0.3 a 0.65 e 0.75 em CN, e 0.2 a 0.85 em R, não foram estatisticamente significantes. Para os demais, os efeitos composição foram positivos e significantes, atuando de forma a contribuir com o diferencial de notas, ou seja, os alunos das escolas privadas apresentam melhores características que contribuem para um bom desempenho, principalmente nos quantis mais elevados da

distribuição, pois em todas as áreas de conhecimento e na média geral, esses efeitos foram positivos e significativos nos percentis 0.9 e 0.95, porém nos quantis iniciais em Matemática e Redação, observaram-se os maiores efeitos, 57% e 45%, respectivamente, com destaque para o primeiro, o qual superou o efeito estrutural em contribuir para o aumento do diferencial de notas.

Tabela 2.2 - Decomposição agregada na distribuição das notas nas áreas do ENEM

Quantile	0.1	%	0.25	%	0.5	%	0.75	%	0.9	%
MG										
Diferencial	0.0637*	100	0.0774*	100	0.0978*	100	0.1173*	100	0.1365*	100
Características	-0.022*	-35	0.0053	7	0.0158	16	0.0270**	23	0.0161**	12
Coeficientes	0.0857*	135	0.0720*	93	0.0820*	84	0.0903*	77	0.1204*	88
CH										
Diferencial	0.0698*	100	0.0635*	100	0.0602*	100	0.0672*	100	0.0767*	100
Características	0.0389**	-56	0.0084	-	0.0232*	39	0.0083**	12	0.0135*	18
Coeficientes	0.1088*	156	0.0551*	87	0.0369*	61	0.0588*	88	0.0632*	82
CN										
Diferencial	0.0602*	100	0.0791*	100	0.1000*	100	0.1127*	100	0.1278*	100
Características	0.0158*	26	-0.0120*	-15	0.0004	-	0.0184	-	0.0337***	26
Coeficientes	0.0443*	74	0.0912*	115	0.0996*	100	0.0943*	84	0.0940*	74
MT										
Diferencial	0.0537*	100	0.0855*	100	0.1365*	100	0.1655*	100	0.1952*	100
Características	0.0308**	57	0.0126	-	0.0216	-	0.0278*	17	0.0414*	21
Coeficientes	0.0228*	43	0.0728*	85	0.1148*	84	0.1377*	83	0.1538*	79
LC										
Diferencial	0.0606*	100	0.0558*	100	0.0637*	100	0.0633*	100	0.0636*	100
Características	-0.0328*	-54	-0.0018	-	0.0104	-	0.0240**	38	0.0120*	19
Coeficientes	0.0934*	154	0.0577*	103	0.0533*	84	0.0392*	62	0.0516*	81
R										
Diferencial	0.1504*	100	0.1506*	100	0.1135*	100	0.1528*	100	0.1503*	100
Características	0.0683*	45	-0.0044	-	0.0024	-	-0.0079	-	0.0230*	15
Coeficientes	0.0821*	55	0.1550*	103	0.1111*	98	0.1608*	105	0.1273*	85

Fonte: Elaboração Própria

Nota (*) significante a 1; (**) significante a 5; *** significante a 10.

Contudo, os efeitos não observáveis (coeficientes) são todos positivos e significativos em todos os quantis da distribuição e em todas as áreas do ENEM e são responsáveis por explicar a maior parte do diferencial das notas. Em MT, esse efeito é crescente ao longo da curva, diferente das demais áreas, cujos efeitos oscilam entre os percentis. Esses resultados

indicam que os alunos das escolas privadas são mais capazes de converter insumos educacionais em maiores pontuações em testes padronizados e que existem outros fatores não observáveis pertencentes a estes alunos que contribuem para melhores desempenhos.

Os dois efeitos, característico e estrutural, podem ser decompostos de forma detalhada, como mostra a tabela 2.3 para a MG. Na parte inferior da distribuição, no quantil 0.1, o hiato de notas foi reduzido devido a características referentes a idade, escolaridade da mãe, e faixas de rendas, para os que não possuem renda e os que possuem até dois salários mínimos, com destaque para a variável que indica se o aluno possui renda mensal domiciliar até um salário mínimo, cujo efeito foi responsável por tornar o efeito composição negativo. Resultado semelhante pode ser observado para os demais quantis, porém a característica do aluno ter mãe com o ensino fundamental completo ou incompleto não age mais no sentido de reduzir os diferenciais de notas e na parte superior da distribuição, quantil 0.9, a variável que indica que a família não possui renda também não age reduzindo o hiato das notas.

Por outro lado, a variável gênero age de forma a contribuir para o aumento do diferencial, positiva e significativa, exceto no quantil 0.1, cuja maior contribuição encontra-se no quantil 0.25, aproximadamente, 13. Em partes da distribuição, a raça também possui esse efeito, aumentando os diferenciais em 39,62 e 10, nos quantis 0.25 e 0.75, assim como as escolaridades da mãe e do pai, cujo efeito sobre hiato no quantil 0.5 foi 4,43 e 26,58, respectivamente. Ao longo de toda distribuição a variável que indica se o aluno terminou o ensino fundamental em 8 anos apresentou sinal positivo, contribuindo para aumento das disparidades de notas entre os alunos das profissionalizantes e privadas, cujo efeito foi maior no quantil 0.25, aproximadamente 40. Dessa forma, ter uma boa trajetória no ensino fundamental afeta o desempenho no ENEM, favorecendo os alunos das escolas particulares. Além disso, possuir renda a partir de dois salários, para alguns percentis, também age de forma a aumentar as diferenças de notas, cujo maior efeito pode ser observado no quantil 0.25, nas três faixas de renda. Com isso, dentre as variáveis que contribuem para elevar o hiato de notas na média geral do ENEM, pertencer a elevadas faixas de rendas é o principal fator.

Como se pôde observar na tabela 2.2, a parte do diferencial atribuído a fatores não observáveis foi bastante elevada em todos os percentis, 135 no quantil 0.1, 93 no quantil 0.25,

84 na mediana, 77 e 88 nos quantis 0.75 e 0.9, respectivamente, e isso pode ser comprovado através da decomposição detalhada desse efeito, apresentados na tabela 3, o qual mostra a influência significativa de fatores não observáveis na maioria das variáveis em boa parte da distribuição, principalmente em percentis mais elevados.

Na parte inferior da distribuição, o efeito não observável seria menor e contribuiria para a redução do hiato entre as notas na média geral, caso os alunos das escolas profissionalizantes tivessem um aumento no desempenho para os alunos do sexo masculino, com mães que possuem ensino fundamental completo ou incompleto, e um maior retorno nos alunos das EEEP em todas as faixas de rendas, principalmente para aqueles com renda familiar entre dois e cinco salários mínimos, o qual age de forma a reduzir o efeito estrutural em aproximadamente 240.

Considerando a mediana da distribuição, observa-se que o efeito estrutural de notas poderia ser reduzido, e assim, contribuir para a queda da desigualdade de notas entre as escolas, se os alunos das profissionalizantes melhorasse o desempenho para os alunos do sexo masculino, para os alunos com pais em todos os níveis de escolaridades, exceto os que possuem superior completo ou incompleto, com destaque para os que possuem ensino médio completo ou incompleto, cuja redução sobre os efeitos não observáveis seria 35,36. Além destes, o efeito estrutural reduziria com o aumento do efeito dos alunos com mães que possuem ensino fundamental e médio completo ou incompleto, em aproximadamente, 7 e 21, respectivamente; com um maior retorno em notas na média geral dos alunos em todas as faixas de rendas, exceto na renda05, cujo efeito mais expressivo foi para aqueles com renda familiar entre dois e cinco salários mínimos, reduzindo o efeito não observado em aproximadamente 85.

Resultado semelhante pode ser observado na parte superior da distribuição, porém, o efeito estrutural das notas seria reduzido com o aumento no desempenho para os alunos das EEEP do sexo masculino, solteiros; com pais que possuem ensino fundamental e médio completo ou incompleto com um maior retorno em notas na média geral dos alunos em todas as faixas de rendas, exceto para os que possuem renda familiar entre dez e quinze salários mínimos, cujo efeito mais expressivo foi para aqueles com renda familiar entre dois e cinco salários mínimos, reduzindo o efeito não observado em aproximadamente 41.

Tabela 2.3 - Decomposição dos efeitos característico e retorno - MG

Variáveis	Característicos					Estrutural				
	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9
Total	-0.0222*	0.0053	0.0158	0.0270**	0.0161**	0.0857*	0.0720*	0.0820*	0.0903*	0.1204*
Idade	-0.0061*	-0.0053*	-0.0052*	-0.0037*	-0.0022*	0.2099*	0.2913*	0.3843*	0.3092	0.1951*
Sexo	0.0002	0.0007*	0.0012*	0.0012*	0.0011*	-0.0077*	-0.0093*	-0.0094*	-0.0075*	-0.007*
Raça	0.0012	0.0021**	0.0012	0.0027*	0.0003	-0.0002	-0.0018	0.0014	0.0002*	0.0052**
Estado civil	-0.001	-0.0002	-0.0001	-0.0000	-0.0000	-0.0855	-0.0452	-0.0325	-0.0473**	-0.0641*
Anos_Fund	0.0014*	0.0018*	0.0003*	0.0027*	0.0011*	0.0043	0.0048**	0.0055*	0.0088*	0.0042**
Anos_Médio	0.0007	-0.0001	0.0001	0.0015	-0.0009	0.0026	0.0080*	0.0029	0.0012	0.0107*
E_Pai1	0.0012	0.0001	-0.0006	0.0005	-0.0002	-0.0006	-0.0007	-0.0011*	-0.0008**	-0.0003
E_Pai2	0.0026	-0.0032	-0.0078	-0.0022	-0.0038	-0.0084	-0.0135**	-0.0163*	-0.0150*	-0.0115***
E_Pai3	0.0006	0.0028	0.0042**	0.0016	0.0016	-0.0138	-0.0209**	-0.0290*	-0.0256*	-0.0202**
E_Pai4	0.0009	0.0040	0.0075	0.0034	0.0029	-0.0042	-0.0073	-0.0118**	-0.0088	-0.0045
E_mae1	-0.0002	0.0002	-0.0001	0.0002	0.0004	-0.0004	-0.0002	-0.0000	-0.0000	-0.0000
E_mae2	-0.0078***	-0.0001	-0.0019	0.0066	0.0071	-0.0097*	-0.0056*	-0.0061*	-0.0021	-0.0013
E_mae3	0.0009***	0.0001	0.0007***	0.0000	-0.0000	-0.0130**	-0.0065	-0.0172*	-0.0109**	-0.0098
E_mae4	0.0037	0.0005	0.0031	-0.0009	-0.0002	-0.0068	0.0002	-0.0057	0.0002	-0.0002
Renda00	-0.0066*	-0.0030*	-0.0011*	-0.0005*	-0.0001	-0.0035*	-0.0017*	-0.0010	-0.0008*	-0.0005*
Renda01	-0.3294*	-0.01503*	-0.0666*	-0.0382*	-0.0176*	-0.0965*	-0.0516*	-0.0320*	-0.0263*	-0.0175*
Renda02	-0.0110*	-0.0054*	-0.0027*	-0.0017*	-0.0011*	-0.1909*	-0.1041*	-0.0685*	-0.0609*	-0.0447*
Renda03	0.1634*	0.0810*	0.0435*	0.0318*	0.0205*	-0.2054*	-0.1092*	-0.0693*	-0.0655*	-0.0499*
Renda04	0.1166*	0.0523*	0.0296*	0.0127*	0.0086	-0.1217*	-0.0554*	-0.0334*	-0.0216*	-0.0154*
Renda05	0.0475*	0.0277*	0.0105	0.0123	-0.0003	-0.0479*	-0.0280*	-0.0110	-0.0144	-0.0015
CSE	-0.0037	-0.0022	-0.0010	-0.0008	-0.0008	0.0382**	0.0326**	0.0546*	0.0555*	0.0563*
Infra	0.0016	0.0014	-0.0006	0.0005	-0.0000	0.0419*	0.0567*	0.0802*	0.0445*	0.0137
Constante	-	-	-	-	-	0.6056*	0.1398***	-0.1019***	-0.0212	0.0842

Fonte: Elaboração Própria

Nota (*) significante a 1; (**) significante a 5; (***) significante a 10.

Assim, é importante verificar a distribuição de notas por quantis para que se possa identificar o efeito heterogêneo das características dos alunos e das escolas e do retorno dessas características no hiato de notas entre as escolas particulares e profissionalizantes do Ceará. As contribuições das covariáveis de forma detalhada na redução do hiato, podem servir como forma de identificar os fatores que as políticas públicas poderiam agir quando se pretende equiparar o ensino público ao privado. Porém, não se pôde analisar características relacionadas a habilidade inata do aluno e ao esforço exercido pelos alunos de ambas as redes de ensino, uma vez que estes podem ser fatores relevantes para a determinação da nota do aluno e do diferencial, mas não são observáveis.

Embora este estudo tenha buscado comparar grupos teoricamente mais homogêneos, alunos profissionalizantes com alunos da rede privada, o diferencial de notas no desempenho do ENEM ainda é significativo, ao longo de toda distribuição. Além disso, prevalece a influência de fatores não observáveis que contribuem para o aumento desse diferencial.

Conclusões

Este trabalho procura ampliar o debate existente na literatura da Economia da Educação sobre o ensino público, com ênfase no ensino profissionalizante desenvolvido no Ceará. Para tanto, busca-se trazer novas evidências sobre os diferenciais de desempenho entre o ensino público e privado, ao analisar as diferenças nas distribuições de notas no ENEM, para as cinco áreas de conhecimento e na média geral, entre alunos das EEEP e alunos das escolas privadas.

Essa análise foi feita visando reduzir a influência de fatores não observados ao se construir grupos de alunos mais homogêneos para se aferir o diferencial de rendimento escolar líquido. A pertinência para tal escolha prende-se ao fato de as escolas de ensino profissional apresentarem desempenho superior às demais escolas públicas, como mostrado no capítulo anterior desta tese. Dessa forma, busca-se verificar se esse novo modelo de escola pública provê uma estrutura de ensino equiparável ao da escola privada, ao analisar como os alunos dos dois tipos de escola se comparam ao longo da distribuição de rendimento escolar e quais fatores observáveis mais influenciam os desempenhos dos alunos das duas escolas.

Além de propor uma nova construção para o grupo das escolas públicas, este trabalho buscou inovar ao utilizar o método de Regressão Quantílica Incondicional (RQI) proposto por Firpo *et al* (2007), o qual emprega regressões com base na função de influência recentrada (FIR) combinadas com um procedimento de reponderação inspirado em DiNardo *et al* (1996) e permite aplicar o método de decomposição de Oaxaca-Blinder.

Como não se possui informações identificadas dos alunos do ensino privado para uma análise longitudinal, utilizaram-se apenas as informações referentes ao desempenho dos alunos no ENEM para o ano de 2014, as quais foram extraídas do INEP, enquanto que, a identificação das escolas profissionalizantes e suas dotações de infraestrutura de todas as escolas foram obtidas da SEDUC e Censo Escolar/2014, respectivamente.

Os resultados encontrados mostram que, em todas as áreas, as estimativas dos diferenciais foram significantes, apresentando vantagens aos alunos das escolares particulares em relação aos alunos do ensino profissional. Em Matemática, ciências da Natureza e na média geral, o aumento da diferença de notas na parte superior da distribuição indica o efeito *glassceiling*, pois se observa que, o diferencial total entre as notas dos alunos das escolas profissionalizantes e privadas é crescente nos quantis. Dessa forma, o diferencial

de notas, a favor dos alunos das escolas particulares, é maior para aqueles que apresentam desempenho superior. Em Ciências Humanas e Linguagens e Códigos, apesar de apresentarem diferenciais positivos, estes praticamente permanecem constantes ao longo da distribuição, com leves oscilações, enquanto que em Redação, os diferenciais foram os mais elevados ao longo de toda a distribuição.

Com isso, embora os alunos das escolas profissionalizantes apresentem desempenho superior quando comparados aos alunos das escolas estaduais do Ceará (MARIANO et al, 2015), ainda existem diferenciais positivos de notas em todas as áreas do ENEM e ao longo de toda a distribuição a favor dos alunos da rede privada, principalmente em Redação.

A decomposição quantílica de Firpo (2007) permite identificar os fatores que contribuem para esse diferencial de notas, os quais podem ser observados através dos efeitos composição (característicos) e estruturais (coeficientes). Dos resultados, observou-se que, os efeitos não observáveis (coeficientes) são todos positivos e significativos em todos os quantis da distribuição e em todas as áreas do ENEM e são responsáveis por explicar a maior parte do diferencial das notas, com destaque para matemática, cujo efeito é crescente ao longo da curva. Esses resultados indicam que os alunos das escolas particulares são mais capazes de converter insumos educacionais em maiores pontuações em testes padronizados e que existem outros fatores não observáveis pertencentes a estes alunos que contribuem para melhores desempenhos.

Além disso, verificou-se que o efeito característico age no sentido de diminuir as diferenças entre as notas apenas em alguns quantis da parte inferior da distribuição das notas, principalmente para o quantil 0.1 em CH e em LC, cujo efeito reduz o diferencial em 56 e 54, respectivamente. Porém, na maioria dos casos, esse efeito foi positivo e significativo, atuando de forma a contribuir com o diferencial de notas, ou seja, os alunos das escolas privadas apresentam melhores características que contribuem para um bom desempenho, principalmente nos quantis mais elevados da distribuição, pois em todas as áreas de conhecimento e na média geral, esses efeitos foram positivos e significativos nos percentis 0.9 e 0.95.

Embora este estudo tenha buscado inovar ao comparar grupos teoricamente mais homogêneos, alunos das escolas profissionalizantes com alunos das escolas particulares, o diferencial de notas no desempenho do ENEM ainda é significativo, ao longo de toda

distribuição. Além disso, prevalece a influência de fatores não observáveis que contribuem para o aumento desse diferencial.

Referências

- BARSKY, R.; BOUND, J.; CHARLES, K. K.; LUPTON, J. P. Accounting for the black–white wealth gap: a nonparametric approach. **Journal of the American Statistical Association**, v. 97, n. 459, p. 663-673, 2002.
- BORAH, B. J.; BASU, A. Highlighting Differences Between Conditional and Unconditional Quantile Regression Approaches Through An Application To Assess Medication Adherence. **Health Econ**. v. 22, n.9, p. 1052–1070, 2013.
- CHERNOZHUKOV, V.; FERNANDEZ-VAL, I; MELLY, B. Inference on Counterfactual Distributions. **Econometrica**. v. 81, n.6, p. 2205-2268, 2013.
- COSTA, L.; ARRAES, R. Identificação parcial do efeito das escolas privadas brasileiras. **Pesquisa e Planejamento Econômico (PPE)**. v.44, n.2, agosto de 2014.
- DINARDO, J; FORTIN, N. M.; LEMIEUX, T. Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. **Econometrica**, v. 64: p. 1001-1044, 1996.
- DONALD, S. G.; GREEN, D. A.; PAARSCH, H. P. Differences in Wage Distributions between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates Source. **Review of Economic Studies**, v. 67, p. 609-633, 2000.
- EDOKA; I. P. **Decomposing Differences in Cotinine Distribution between Children and Adolescents from Different Socioeconomic Backgrounds**. HEDG. Working Paper n. 12/29. 2012.
- FRANÇA, M. T. A.; GONÇALVES, F.O. Provisão pública e privada de educação fundamental: diferenças de qualidade medidas por meio de propensity score. **Economia Aplicada**. v.14, n.4, p.373-390, 2010.
- FISHER, J.; MARCHAND, J. Does the retirement consumption puzzle differ across the distribution? **The Journal of Economic Inequality**, v.12, p. 279-296. 2014.
- FIRPO, S.; FORTIN, N.; LEMIEUX, T.. Decomposing wage distributions using recentered influence function regressions. **Mimeo**. University of British Columbia, 2007.
- FIRPO, S.; FORTIN, N.; LEMIEUX, T.. Unconditional Quantile Regressions, **Econometrica**, v. 77, n.3, p. 953-973, 2009.
- FORTIN, N M. The Gender Wage Gap among Young Adults in the United States: The Importance of Money vs. People. **Journal of Human Resources**, v.43, p. 886-920, 2008.
- FORTIN, N. M; LEMIEUX, T. Rank Regressions, Wage Distributions, and the Gender Gap. **Journal of Human Resources** v. 33, p. 610.643, 1998.

FOURNIER, J.; KOSKE, I. Less income inequality and more growth—Are they compatible? Part 7. The drivers of labour earnings inequality—An analysis based on conditional and unconditional quantile regressions. **OECD Economics Department Working Papers**, n. 930, OECD Publishing, 2012.

HAMPEL, F. R.; RONCHETTI, E. M.; ROUSSEEUW, P. J.; STAHEL, W. A. **Robust Statistics: The Approach Based on Influence Functions**. Wiley; New York: 1986.

HIRSCH, B.T.; WINTERS, J.V. An Anatomy of Racial and Ethnic Trends in Male Earnings in the U.S. **Review of Income and Wealth** v. 60, p. 930-947, 2014.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS – INEP. Disponível em: <<http://www.inep.gov.br/>>. Acesso em: 24 de set. de 2016.

JUHN, C.; MURPHY, K. M.; PIERCE, B. Wage Inequality and the Rise in Returns to Skill. **Journal of Political Economy** v. 101, p. 410-442, 1993.

KASSENBOEHMER, S.C.; SINNING, M.G. Distributional changes in the gender wage gap. **Industrial & Labor Relations Review** v. 67, p. 335-361, 2014.

KOENKER, R.; BASSETT, J.R. Regression quantiles. **Econometrica: journal of the Econometric Society**, p. 33-50, 1978.

LE, H.T.; BOOTH, A.L. Inequality in Vietnamese Urban–Rural Living Standards, 1993–2006. **Review of Income and Wealth** v. 60, p. 862–886. 2013.

MACHADO, J. F.; MATA, J. Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression. **Journal of Applied Econometrics**, v.20, p. 445-465, 2005.

MANSKI, C. Anatomy of the selection problem. **The journal of human resource**, v. 24, p. 343-360, 1989.

MARIANO, F. Z.; ARRAES, R. A; SOUZA, N. O. **Desempenho Escolar e Inserção no Mercado de Trabalho: Uma Avaliação das Escolas Estadual de Ensino Profissionalizantes (EEEP) do Ceará**. In. XI Encontro Economia do Ceará em Debate. IPECE 2015. Disponível em: <http://www2.ipece.ce.gov.br/encontro/2015/trabalhos/Desempenho20escolar20e20inserC3A7C3A3o20no20mercado20de20trabalho.pdf> Acesso em: 10 Jan. 2016

MELLY, B. Decomposition of differences in distribution using quantile regression. **Labour economics** v.12, p. 577-590, 2005.

MORAES, A. G. E.; BELLUZZO, W. O diferencial de desempenho escolar entre escolas públicas e privadas no Brasil. **Nova Economia**. Belo Horizonte. v.24 n.2 p.409-430 maio-agosto, 2014.

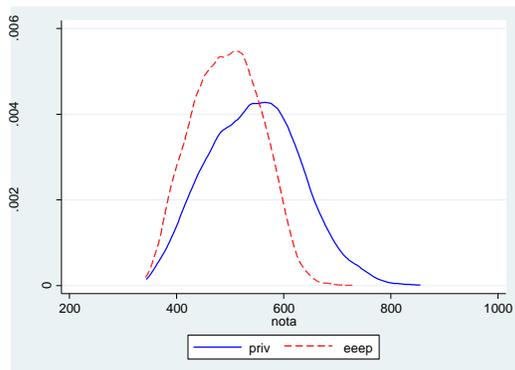
MORIN, L.P. Do Men and Women Respond Differently to Competition? Evidence from a Major Education Reform. **Journal of Labor Economics** v. 33, p. 443-491, 2015

NGUYEN; H. T. **The Evolution of The Gender Test Score Gap Through Seventh Grade: New Insights From Australia Using Quantile Regression and Decompositon**. Bankwest Curtin Economics Centre. Working Paper. n. 15/7, 2015.

RELATÓRIO DE GESTÃO “O PENSAR E O FAZER DA EDUCAÇÃO PROFISSIONAL NO CEARÁ – 2008 a 2014”. Secretaria da Educação. Coordenadoria de Educação Profissional. – 1. ed. – Fortaleza: Secretaria da Educação, 2014

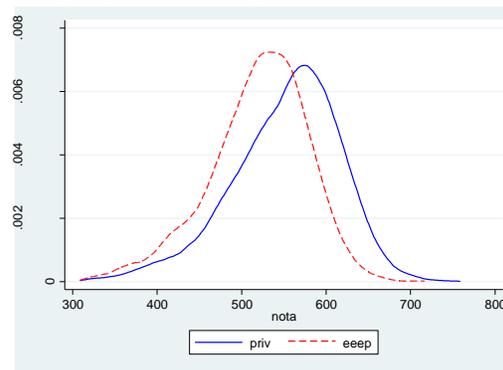
RUBIL; I. **The Great Recession and the Public-Private Wage Gap:** Distributional Decomposition Evidence from Croatia 2008-2011. MPRA Paper n. 46798. 2013.

Gráfico A.2.1 -Distribuição das notas em Ciências Naturais



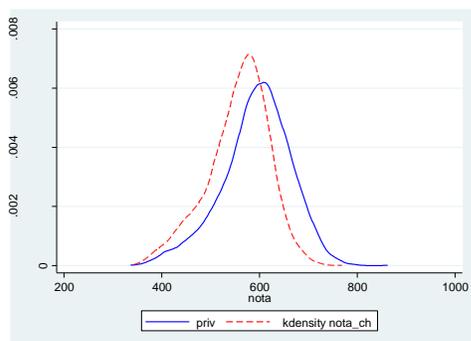
Fonte: Elaboração Própria

Gráfico A.2.2 -Distribuição das notas em Linguagens e Códigos



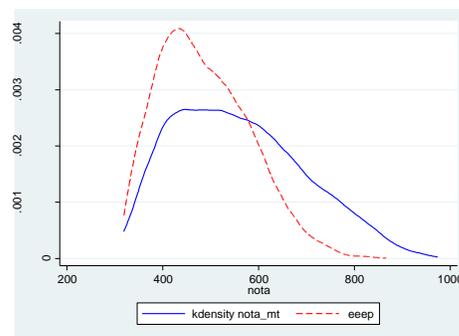
Fonte: Elaboração Própria

Gráfico A.2.3 - Distribuição das notas em Ciências Humanas



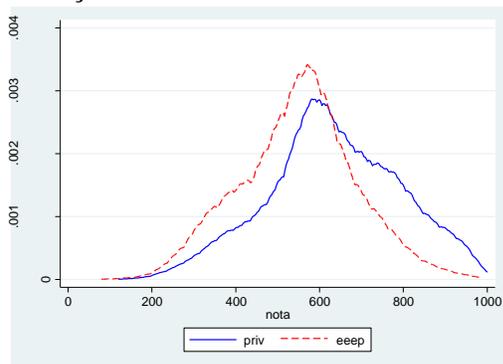
Fonte: Elaboração Própria

Gráfico A.2.4 - Distribuição das notas em Matemática



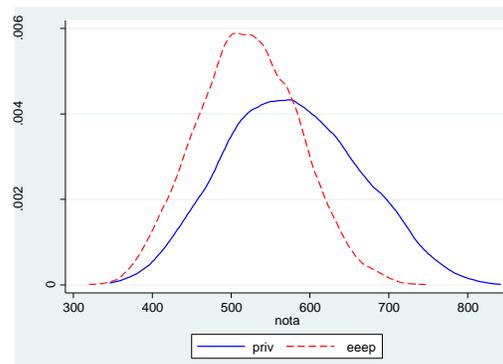
Fonte: Elaboração Própria

Gráfico A.2 - Distribuição das notas em Redação



Fonte: Elaboração Própria

Gráfico 3 - Distribuição das notas em Média Geral



Fonte: Elaboração Própria

Apêndice B

Tabela B.2.1 - Estimativas da Regressão Quantílica Incondicional por Escola – MG

Variáveis	0.1		0.5		0.9	
	EEEP	Particulares	EEEP	Particulares	EEEP	Particulares
Idade	-0.0348*	-0.0229*	-0.0296*	-0.0078*	-0.0125*	-0.0015*
Sexo	0.0064	-0.0102**	0.0291*	0.0088**	0.0276*	0.0123*
Raça	0.0054	0.0049	0.0054	0.0088**	0.0016	0.0140*
Estado civil	0.0426	-0.0436	0.0271	-0.0057	0.0251**	-0.0396**
Anos_Fund	0.0187*	0.0282*	0.0253*	0.0374*	0.0149*	0.0242*
Anos_Médio	0.0043	0.0093**	0.0011	0.0067***	-0.0055	0.0148*
E_Pai1	-0.0237	-0.0644**	0.0115	-0.0665*	0.0045	-0.0180
E_Pai2	-0.0077	-0.0441*	0.0227	-0.0479*	0.0111	-0.0386*
E_Pai3	0.0060	-0.0291*	0.0425**	-0.0311*	0.0170	-0.0343*
E_Pai4	0.0041	-0.0118**	0.0351	-0.0096	0.0136	-0.0033
E_mae1	0.0117	-0.0581	0.0093	0.0001	-0.0197	-0.0243***
E_mae2	0.0232***	-0.0387*	0.0057	-0.0336*	-0.0212	-0.0298*
E_mae3	0.0274**	-0.0058	0.0198***	-0.0241*	-0.0025	-0.0275*
E_mae4	0.0183	-0.0047	0.0153	-0.0042	-0.0010	-0.0020
Renda00	0.6233*	-0.0928**	0.1094*	-0.0983*	0.0168	-0.1034*
Renda01	0.6499*	-0.0796*	0.1314*	-0.1107*	0.0348*	-0.0978*
Renda02	0.6789*	-0.0454*	0.1720*	-0.0881*	0.0706*	-0.0993*
Renda03	0.6838*	-0.0032	0.1821*	-0.0499*	0.0858*	-0.0810*
Renda04	0.6861*	-0.0086	0.1742*	-0.0165**	0.0507	-0.0375*
Renda05	0.7715*	-0.0048	0.1707	-0.0072	-0.0056	-0.0307***
CSE	-0.0196	0.0371**	-0.0052	0.0758*	-0.0042	0.0792*
Infra	0.0269	0.1282*	-0.0107	0.1828*	-0.0014	0.0317**
Constante	5.9553*	6.5609*	6.5421*	6.4402*	6.5380*	6.6223*

Fonte: Elaboração Própria

Nota (*) significante a 1; (**) significante a 5; (***) significante a 10.

Tabela B.2.2 - Estimativas da Regressão Quantílica Incondicional por Escola - CH

Variáveis	0.1		0.5		0.9	
	EEEEP	Particulares	EEEEP	Particulares	EEEEP	Particulares
Idade	-0.0219*	-0.0087*	-0.0198*	-0.0047*	-0.0074*	-0.0010*
Sexo	0.0209*	-0.0022	0.0276*	0.0175*	0.0306*	0.0184*
Raça	0.0045	0.0078	0.0081**	0.0028	0.0103*	0.0050***
Estado civil	0.0885	0.0028	-0.0004	-0.0045	0.0166***	-0.0042
Anos_Fund	0.0197*	0.0289*	0.0178*	0.0245*	0.0101*	0.0151*
Anos_Médio	-0.0036	-0.0001	0.0049	0.0050***	0.0011	0.0074*
E_Pai1	0.0087	-0.0479	0.0022	-0.0405*	-0.0045	-0.0295*
E_Pai2	-0.0002	-0.0298*	0.0053	-0.0332*	-0.0042	-0.0378*
E_Pai3	0.0249	-0.0218**	0.0210	-0.0229*	0.0056	-0.0314*
E_Pai4	0.0122	-0.0112	0.0170	-0.0050	-0.0075	-0.0125***
E_mae1	0.0191	0.0183	-0.0053	-0.0148	-0.0033	0.0161
E_mae2	0.0242	-0.0251**	-0.0058	-0.0248*	0.0084	-0.0036
E_mae3	0.0284	-0.0058	0.0070	-0.0171*	0.0027	-0.0044
E_mae4	0.0108	-0.0002	0.0016	-4.65e-06	0.0040	0.0048
Renda00	0.8756*	-0.1121**	0.1132*	-0.0619*	0.0029	-0.0598*
Renda01	0.8716*	-0.0878*	0.1181*	-0.0704*	0.0294*	-0.0694*
Renda02	0.9167*	-0.0468*	0.1529*	-0.0527*	0.0451*	-0.0648*
Renda03	0.9392*	-0.0031	0.1604*	-0.0303*	0.0559*	-0.0576*
Renda04	0.9109*	-0.0048	0.1414*	-0.0104***	0.0341	-0.0263*
Renda05	1.0233*	-0.0063	0.2959*	-0.0035	-0.0150	-0.0299*
CSE	-0.0530**	0.0202	-0.0104	0.0455*	0.0116	0.0464*
Infra	-0.0540	0.1207*	-0.0006	0.0914*	0.0150	0.0299*
Constante	5.5466*	6.3368*	6.5254*	6.4567*	6.4987*	6.5563*

Fonte: Elaboração Própria

Nota (*) significante a 1; (**) significante a 5; (***) significante a 10.

Tabela B.2.3 - Estimativas da Regressão Quantílica Incondicional por Escola - CN

Variáveis	0.1		0.5		0.9	
	EEEEP	Particulares	EEEEP	Particulares	EEEEP	Particulares
Idade	-0.0148*	-0.0111*	-0.0208*	-0.0074*	-0.0085*	-0.0011*
Sexo	0.0300*	0.0091***	0.0564*	0.0409*	0.0453*	0.0363*
Raça	-0.0042	0.0019	0.0089***	0.0106*	0.0022	0.0183*
Estado civil	0.0509	-0.0451	0.0251	-0.0119	0.0093	-0.0210***
Anos_Fund	0.0149*	0.0218*	0.0138*	0.0389*	0.0190*	0.0232*
Anos_Médio	0.0049	0.0066	0.0025	0.0061	-0.0000	0.0145*
E_Pai1	0.0436	-0.0296	0.0323	-0.0582*	-0.0184	-0.0375*
E_Pai2	0.0449	-0.0325*	0.0405***	-0.0457*	-0.0223	-0.0396*
E_Pai3	0.0517	-0.0286*	0.0474**	-0.0370*	-0.0198	-0.0381*
E_Pai4	0.0464	-0.0120	0.0517**	-0.0042	-0.0075	-0.0053
E_mae1	0.0182	-0.0971**	-0.0156	-0.0113	-0.0140	-0.0117
E_mae2	0.0230	-0.0222**	-0.0211***	-0.0367*	-0.0116	-0.0267*
E_mae3	0.0218	-0.0076	-0.0171	-0.0226*	-0.0025	-0.0167**
E_mae4	0.0268	0.0030	-0.0122	-0.0005	-0.0034	0.0025
Renda00	-0.1142*	-0.0559	0.1368*	-0.1103*	0.0216	-0.1152*
Renda01	-0.1138*	-0.0659*	0.1526*	-0.1034*	0.0237*	-0.1081*
Renda02	-0.1075*	-0.0297*	0.1816*	-0.0892*	0.0498*	-0.1114*
Renda03	-0.0920*	-0.0046	0.1864*	-0.0481*	0.0679*	-0.0938*
Renda04	-0.0813*	-0.0027	0.1653*	-0.0193**	0.0090	-0.0526*
Renda05	-0.0117	-0.0120	0.1891	-0.0120	0.2761	-0.0336**
CSE	-0.0133	0.0566*	-0.0416*	0.0552*	0.0026	0.0681*
Infra	-0.0084	0.1288*	-0.0109	0.1705*	0.0111	0.0368**
Constante	6.2495*	6.2488*	6.3473*	6.3909*	6.4610*	6.5434*

Fonte: Elaboração Própria

Nota (*) significante a 1; (**) significante a 5; (***) significante a 10.

Tabela B.2.4 - Estimativas da Regressão Quantílica Incondicional por Escola - LC

Variáveis	0.1		0.5		0.9	
	EEEP	Particulares	EEEP	Particulares	EEEP	Particulares
Idade	-0.0291*	-0.0134*	-0.0210*	-0.0057*	-0.0101*	-0.0010*
Sexo	-0.0265*	-0.0347*	-0.0035	-0.0167*	0.0018	-0.0017
Raça	0.0044	0.0057	0.0051	0.0060**	0.0096**	0.0069*
Estado civil	0.0715	0.0148	0.0347***	-0.0004	-0.0036	-0.0097
Anos_Fund	0.0116**	0.0241*	0.0156*	0.0198*	0.0123*	0.0114*
Anos_Médio	-0.0069	0.0118**	0.0016	0.0015	-0.0014	0.0009
E_Pai1	-0.0454	-0.0593**	-0.0093	-0.0438*	-0.0020	-0.0393*
E_Pai2	-0.0204	-0.0273*	-0.0116	-0.0411*	0.0011	-0.0374*
E_Pai3	-0.0111	-0.0220*	0.0044	-0.0274*	0.0087	-0.0302*
E_Pai4	-0.0354	-0.0100	-0.0027	-0.0027	0.0033	-0.0140**
E_mae1	-0.0166	-0.0335	0.0055	0.0055	-0.0059	0.0028
E_mae2	-0.0022	-0.0234**	-0.0014	-0.0014	-0.0117	-0.0093**
E_mae3	0.0146	-0.0101	0.0109	0.0109	-0.0052	-0.0077***
E_mae4	0.0048	-0.0020	0.0012	0.0012	-0.0085	0.0037
Renda00	0.8101*	-0.0216	0.1198*	0.1198*	0.0121***	-0.0690*
Renda01	0.8470*	-0.0484*	0.1353*	0.1353*	0.0378*	-0.0514*
Renda02	0.8742*	-0.0125	0.1634*	0.1634*	0.0652*	-0.0512*
Renda03	0.8804*	0.0009	0.1753*	0.1753*	0.0728*	-0.0399*
Renda04	0.8960*	0.0028	0.1512*	0.1512*	0.0449**	-0.0191**
Renda05	0.9740*	0.0001	0.1610	0.1610	0.0073	-0.0258*
CSE	-0.0246	0.0473*	-0.0065	-0.0065	-0.0002	0.0389*
Infra	0.0588***	0.0636*	0.0423**	0.0423**	0.0368**	0.0086
Constante	5.6827*	6.3425*	6.4370*	6.4370*	6.4977*	6.4973*

Fonte: Elaboração Própria

Nota (*) significativa a 1; (**) significativa a 5; (***) significativa a 10.

Tabela B.2.5 - Estimativas da Regressão Quantílica Incondicional por Escola - MT

Variáveis	0.1		0.5		0.9	
	EEEP	Particulares	EEEP	Particulares	EEEP	Particulares
Idade	-0.0270*	-0.0155*	-0.0428*	-0.0140*	-0.0204*	-0.0023*
Sexo	0.0537*	0.0534*	0.1180*	0.0966*	0.1058*	0.0833*
Raça	0.0021	0.0034	0.0001	0.0126**	-0.0020	0.0239*
Estado civil	0.0270	-0.0806	-0.0197	-0.0647**	-0.0159	-0.0423**
Anos_Fund	0.0089	0.0199*	0.0278*	0.0605*	0.0264*	0.0458*
Anos_Médio	0.0029	0.0090	-0.0084	0.0238*	0.0012	0.0135**
E_Pai1	0.0085	-0.0139	-0.0202	-0.0738*	0.0128	-0.0431***
E_Pai2	0.0125	-0.0214***	-0.0026	-0.0688*	0.0238	-0.0736*
E_Pai3	0.0214	-0.0034	0.0163	-0.0548*	0.0178	-0.0603*
E_Pai4	0.0454	0.0068	0.0255	-0.0167	0.0544	-0.0089
E_mae1	-0.0127	-0.0289	-0.0374	-0.0562	-0.0277	-0.0577*
E_mae2	0.0061	-0.0446*	-0.0260	-0.0629*	-0.0166	-0.0590*
E_mae3	0.0198	-0.0178***	0.0018	-0.0405*	0.0054	-0.0527*
E_mae4	0.0030	-0.0130	-0.0160	0.0017	0.0165	-0.0059
Renda00	-0.1262*	-0.0261	0.1580*	-0.1573*	0.0300	-0.1518*
Renda01	-0.1390*	-0.0810*	0.2055*	-0.2002*	0.0259**	-0.1607*
Renda02	-0.1123*	-0.0537*	0.2546*	-0.1646*	0.0826*	-0.1751*
Renda03	-0.1134*	-0.0144	0.2892*	-0.0947*	0.0993*	-0.1529*
Renda04	-0.1426*	-0.0087	0.2182*	-0.0484*	0.0775	-0.0840*
Renda05	-0.0267	0.0011	0.2556	-0.0169	-0.0450	-0.0488***
CSE	0.0261	0.0433***	-0.0239	0.0828*	0.0157	0.1490*
Infra	-0.0087	0.1327*	0.0093	0.2219*	0.0343	0.0665*
Constante	6.4286*	6.2523*	6.6696*	6.5513*	6.6641*	6.7093*

Fonte: Elaboração Própria

Nota (*) significativa a 1; (**) significativa a 5; (***) significativa a 10.

Tabela B.2.6 - Estimativas da Regressão Quantílica Incondicional por Escola - R

Variáveis	0.1		0.5		0.9	
	EEEP	Particulares	EEEP	Particulares	EEEP	Particulares
Idade	-0.0749*	-0.0605*	-0.0356*	-0.0103*	-0.0227*	-0.0034*
Sexo	-0.0602*	-0.0998*	-0.0471*	-0.0835*	-0.0402*	-0.0589*
Raça	0.0485*	0.0263**	0.0037	0.0121**	-0.0041	0.0132**
Estado civil	0.0853	0.0314	0.0605	0.0078	0.0606	-0.0434***
Anos_Fund	0.0505*	0.0429*	0.0264*	0.0390*	0.0229*	0.0269*
Anos_Médio	0.0263***	0.0230***	0.0000	0.0136**	-0.0120	0.0131**
E_Pai1	-0.0352	-0.1461***	-0.0238	-0.0775*	-0.0584	-0.0513**
E_Pai2	0.0117	-0.1112*	-0.0113	-0.0677*	-0.0233	-0.0408*
E_Pai3	0.0477	-0.0769*	0.0078	-0.0352*	0.0008	-0.0288**
E_Pai4	0.0489	-0.0537*	-0.0034	-0.0152	-0.0132	-0.0143
E_mae1	-0.0880	-0.3109**	-0.0125	-0.0202	0.0266	-0.0099
E_mae2	-0.0065	-0.0762*	-0.0042	-0.0380*	0.0146	-0.0470*
E_mae3	0.0204	-0.0073	0.0205	-0.0264*	0.0302	-0.0373*
E_mae4	0.0223	-0.0088	0.0030	-0.0162***	0.0340	-0.0166
Renda00	-0.2363*	-0.1798	0.2622*	-0.0813***	0.0601**	-0.1281*
Renda01	-0.2192*	-0.1468*	0.2844*	-0.0867*	0.1211*	-0.1045*
Renda02	-0.1581*	-0.0932*	0.3227*	-0.0583*	0.1689*	-0.0966*
Renda03	-0.1710*	-0.0148	0.3301*	-0.0273**	0.2225*	-0.0682*
Renda04	-0.3329*	-0.0224	0.3209*	-0.0025	0.2209*	-0.0353***
Renda05	0.0329	-0.0071	0.3123***	-0.0114	0.0324**	-0.0643*
CSE	0.0946***	0.1780*	0.0073	0.1106*	-0.0335	0.0959*
Infra	0.0535	0.2588*	-0.0392	0.2269*	-0.0403	0.1097*
Constante	7.2347*	7.0221*	6.6280*	6.5642*	6.8421*	6.8747*

Fonte: Elaboração Própria

Nota (*) significante a 1; (**) significante a 5; (***) significante a 10

CAPÍTULO 3

ENSINO PROFISSIONALIZANTE E INSERÇÃO NO MERCADO DE TRABALHO

Introdução

Avaliação de políticas públicas direcionadas para a inserção de jovens no mercado de trabalho tem sido foco de muitos pesquisadores devido às altas taxas de jovens que não estudam e nem trabalham. Bassi et al. (2012) verificam que, no Brasil e em outros países da América Latina, aproximadamente, 15 dos jovens entre 16 e 24 anos inserem-se nessa categoria. Segundo relatório do Banco Mundial (2016), essa proporção aumentou para 19 nesta região. Além disso, nessa fase existem elevadas taxas de empregos informais, indicando que estes jovens não possuem habilidades exigidas em empregos formais de qualidade (ARAÚJO et al., 2014). No Brasil, cerca de 4 milhões de jovens trabalham em atividades informais, das quais 90 correspondem a menos de um salário mínimo (SIMÕES, 2010).

Dessa forma, segundo Guimarães e Almeida (2013), este grupo torna-se um público vulnerável, que enfrentam maiores dificuldades de encontrar emprego e tendem a trabalhar em ocupações mais precárias, já que não possuem formação educacional adequada no período anterior à entrada no mercado de trabalho. Araújo et al. (2014) afirmam que, alguns estudos realizados para países da América Latina evidenciam que grande parte dos jovens egressos do ensino médio apresenta condições piores de trabalho, como, por exemplo, baixa experiência, menores salários e menos estabilidade, em comparação aos trabalhadores mais experientes. Segundo os autores, isso ocorre porque a maioria dos jovens inicia a trajetória profissional no mercado de trabalho informal e, posteriormente, aprimoram as habilidades até conseguirem um emprego formal, porém essas habilidades deveriam ser ensinadas na etapa de ensino de conclusão da educação básica, o nível médio.

Dentre os tipos de políticas que visam amenizar essa situação, encontra-se a educação profissional, o qual proporciona ao jovem, um ensino específico direcionado ao mercado de trabalho. Segunda Silva et al. (2013), os jovens procuram os cursos técnicos visando uma formação de qualidade superior, se profissionalizarem em áreas com melhores *status* e garantir uma empregabilidade. Diversos autores buscam verificar os efeitos desse tipo de

ensino sobre a inserção no mercado de trabalho (SEVERNINI; ORELLANO, 2010; ASSUNÇÃO; GONZAGA, 2010; ARAÚJO et al, 2014), porém, não se constatou por parte dos autores, trabalhos para o Brasil que utilizem informações longitudinais, os quais permitam identificar o jovem pós ensino médio no mercado de trabalho.

Severnini e Orellano (2010) questionou se os programas de qualificação profissional realmente contribuem para os treinados obterem uma melhoria de bem-estar e se esse ensino contribui para aumentar a probabilidade de inserção do indivíduo no mercado de trabalho brasileiro. Para tanto, os autores investigaram se este tipo de ensino aumentou a probabilidade de inserção no mercado de trabalho e a elevação de renda dos egressos até meados da década de 1990, comparativamente aos que não cursaram esse tipo de ensino. Por meio de microdados da Pesquisa sobre Padrões de Vida (PPV) de 1996 do IBGE encontraram que os egressos de cursos profissionalizantes de nível básico tinham renda esperada 37 maior que a de indivíduos que não fizeram esse tipo de curso no ensino fundamental. Por outro lado, para os egressos do ensino profissional de nível tecnológico, observa-se uma redução de 27 da renda esperada, comparativamente aos que não participaram desse tipo de curso no ensino superior.

Assunção e Gonzaga (2010) apresentam os resultados de uma análise realizada com base nos microdados do suplemento especial sobre educação profissional da PNAD/2007, cujo objetivo reside em analisar a inserção da população brasileira nos cursos de educação profissional e verificar o impacto da educação profissional sobre o rendimento dos trabalhadores brasileiros. Dos resultados, observaram que a inserção da educação profissional em famílias com renda *per capita* inferior a dois salários mínimos é bem menor que nas demais faixas de renda e que a educação profissional aumenta a produtividade dos trabalhadores. Dessa forma, os autores sugerem que essa educação mereça atenção no desenho de políticas públicas.

Araújo et al. (2014) avaliam os efeitos da educação profissional e tecnológica (EPT) de nível médio sobre a inserção produtiva do aluno. Dos resultados, observaram efeitos significativos e positivos, nos quais, em média, há um aumento entre 1,2 e 1,3 pontos percentuais na probabilidade de estar empregado para os alunos de EPT. Já a probabilidade de trabalhar em atividade para a qual se preparou é de, aproximadamente, 17 pontos

percentuais maior que os alunos que não realizaram EPT. Os efeitos por dependências administrativas permanecem significativos e variam entre 15 a 25.9 pontos percentuais.

Embora não sendo uma avaliação do ensino profissional, Lin et al. (2016) usaram informações de uma pesquisa longitudinal de jovens e dados suplementares para avaliar como o desempenho cognitivo, medido no fim do ensino secundário, está relacionado com resultados no mercado de trabalho. Observam-se cinco resultados principais. i) o desempenho cognitivo está positivamente associada a resultados futuros do mercado de trabalho em todas as idades; ii) os retornos da habilidade cognitiva aumentam com a idade; iii) o efeito sobre os rendimentos totais reflete uma combinação de impactos positivos do desempenho cognitivo tanto para salários/hora e horas de trabalho anuais; iv) os retornos da habilidade cognitiva são maiores para as mulheres que para os homens e para os negros e hispânicos do que para os brancos não-hispânicos; v) os ganhos médios dos rendimentos ao longo da vida previsto para maiores níveis de desempenho cognitivo são, apenas, ligeiramente superiores aos reportados em estudos anteriores.

As Escolas Estaduais de Educação Profissional (EEEP) no Ceará fornecem, além do ensino com as disciplinas básicas do currículo do ensino médio, os cursos técnicos e profissionalizantes voltados para diversas áreas (saúde, agricultura, computação, finanças, etc). Essa ação possibilita ao aluno concluir o ensino médio e se profissionalizar para o mercado de trabalho. Com isso, além de formar o jovem para o mercado de trabalho, o ensino das escolas profissionalizantes também o capacita a ingressar em universidades. No capítulo I desta tese, verificou-se que esse tipo de ensino demonstra possuir diretrizes mais eficazes em potencializar habilidades do conhecimento cognitivo relativo ao ensino regular, porém não se observou se este possui efeito sobre o mercado de trabalho.

Com isso, este trabalho procura ampliar o debate existente na literatura da Educação sobre o ensino profissionalizante, com ênfase nos efeitos gerados no mercado de trabalho. Para tanto, busca-se trazer novas evidências sobre o papel das escolas profissionais públicas do Ceará sobre os rendimentos e os tipos de ocupações dos jovens.

Uma das contribuições deste estudo reside na base de dados utilizada para mensurar este efeito, pois através da construção de identificadores foi possível gerar uma amostra com informações longitudinais através do cruzamento de diferentes bases, a saber, Censo escolar 2011 e 2012, e Relação Anual de Informações Sociais (RAIS) 2013.

Além disso, este trabalho também contribui com a literatura para o Brasil na escolha da opção metodológica para a condução dessa avaliação, pois aplicou-se os procedimentos feitos por Watson e Elliot (2016), os quais combinam o método da entropia, desenvolvido por Hainmueller (2012) com o pareamento por escore de propensão - PEP, visando assim, equilibrar com maior robustez as covariadas, minimizando os desequilíbrios entre os grupos de tratados e controle. Para aplicação deste método, será utilizada uma variável binária para o tratamento, o qual irá permitir encontrar o efeito tratamento médio sobre os tratados das escolas profissionalizantes sobre a remuneração e as ocupações dos concludentes do ensino médio em 2011 e 2012 no mercado de trabalho em 2013.

Durante o ensino profissionalizante, o aluno tem a opção de escolher um curso, o qual pretenda se especializar. Visando encontrar o efeito dessa escolha sobre a remuneração e os tipos de ocupações, este estudo optou por considerar uma nova variável de tratamento, o qual irá assumir múltiplos valores a partir dos tipos de cursos presentes na amostra. Em decorrência dessa variável ser agora definida como categórica, será utilizado a metodologia do pareamento por escore de propensão com múltiplos tratamentos, a qual é abordada por Rosenbaum e Rubin (1983), Imbens (2000), Lechner (2001). Esta difere da análise do efeito de tratamento único nas suposições de identificação, pois estas são modificadas para incorporar os múltiplos valores do tratamento.

Em sequência, o capítulo está organizado com as seguintes seções: abordagem metodológica dividida em Balanceamento por Entropia e Pareamento por Escore de Propensão com múltiplos tratamentos; resultados e conclusões.

2 Metodologia

2.1 Balanceamento por Entropia

Hainmueller e Xu (2013) descrevem o balanceamento por entropia como uma generalização da abordagem do pareamento por escore de propensão (PEP), embora com procedimentos distintos. Enquanto o PEP calcula os escores de propensão através de uma regressão logit ou probit para fazer o pareamento, e verificar-se os pesos estimados equilibram as distribuições das covariadas, a entropia calcula os pesos diretamente para ajustar as distribuições amostrais conhecidas, integrando o balanceamento das covariáveis aos pesos. Embora ambos os procedimentos avaliem efeitos de tratamentos, o balanceamento por entropia também pode ser aplicado para ajustar a amostra, conforme aplicou Watson e Elliot (2016).

Segundo Hainmueller (2012), esse método permite ponderar um conjunto de dados, tais que, as distribuições das variáveis nas observações reponderadas satisfaçam um conjunto de condições especiais de momentos, de forma que exista equilíbrio exato sobre o primeiro (média), segundo (variância) e terceiro (assimetria) momentos das distribuições de variáveis independentes nos grupos de tratamento e controle. Assim, é possível se especificar um nível de equilíbrio desejável para as covariadas, usando um conjunto de condições associados aos momentos da distribuição. A vantagem deste método sobre os algoritmos logit/probit reside na capacidade de implementar diretamente o equilíbrio exato.

Considere as seguintes informações: 1) D é uma variável binária que assume o valor 1 se a unidade pertence ao tratamento, e 0 se pertencer ao controle; 2) X é uma matriz composta pelos elementos x_{ij} , referentes aos valores da variável exógena pré-determinada j na unidade i ; 3) A densidade das covariadas nas amostras de tratamento e controle são dadas por $f_{X|D=1}(x)$ e $f_{X|D=0}(x)$, respectivamente; 4) A variável resultado observado é expressa por $Y = Y(1)D + (1 - D)Y(0)$.

O Efeito Médio Tratamento sobre os Tratados (EMTT) é dado por $\tau = E[Y(1)|D = 1] - E[Y(0)|D = 1]$, onde a primeira média pode ser diretamente identificada do grupo de tratados, mas a segunda corresponde ao contrafactual, o qual não é observado. Rosenbaum and Rubin (1983) mostram que, assumindo seleção nos observáveis, $Y(0) \perp D|X$, e

sobreposição, $\Pr(D = 1|X = x) < 1$ para todo x no suporte de $f_{X|D=1}$, o EMTT é identificado como:

$$\tau = E[Y|D = 1] - \int E[Y|X = x, D = 0] f_{X|D=1}(x) dx \quad (1)$$

Para estimar o contrafactual da equação 1, a distribuição da covariável no grupo de controle necessita ser ajustada para torná-la semelhante à distribuição no grupo de tratamento, tal que o indicador de tratamento D se torne mais perto de ser ortogonal em relação às covariáveis. Uma variedade de métodos de pré-processamento de dados, tais como pareamento pelo vizinho mais próximo, cem, escore de propensão têm sido propostos para reduzir o desequilíbrio na distribuição de variáveis de controle. Uma vez ajustadas as distribuições de variáveis independentes, métodos de análise padrão, tais como a regressão, podem ser posteriormente utilizados para estimar o tratamento com menor erro e modelo de dependência (IMBENS, 2004; RUBIN, 2006; HO et al., 2007; SEKHON, 2009).

Considere o caso mais simples onde o efeito tratamento nos dados pré-processados é estimado usando a diferença nos resultados médios entre os grupos de tratados e controle ajustado, cujo método muito utilizado na literatura é o escore de propensão ponderado (HIRANO; IMBENS; RIDDER, 2003), onde a média contrafactual é estimada como segue:

$$E[Y(0)|\widehat{D} = 1] = \frac{\sum_{\{i|D=0\}} Y_i d_i}{\sum_{\{i|D=0\}} d_i} \quad (2)$$

As unidades de controle recebem um peso dado por $d_i = \frac{\hat{p}(x_i)}{1-\hat{p}(x_i)}$, onde $\hat{p}(x_i)$ na equação 2 é o escore de propensão, o qual é comumente estimado através de uma regressão probit ou logit. Se este modelo estiver corretamente especificado, então o peso estimado d_i assegura que a distribuição da covariável das unidades de controle reponderadas corresponda à distribuição no grupo de tratamento. No entanto, na prática, essa abordagem muitas vezes não consegue equilibrar conjuntamente todas as covariáveis.

O balanceamento por entropia generaliza a abordagem de ponderação do escore de propensão ao estimar os pesos diretamente de um conjunto de restrições de equilíbrio que exploram o conhecimento do pesquisador sobre os momentos de amostra. Considere w_i o peso do balanceamento por entropia escolhido para cada unidade de controle, os quais foram encontrados pelo seguinte esquema de reponderação que minimiza a distancia métrica de entropia:

$$\min_{w_i} H(w) = \sum_{\{i|D=0\}} w_i \log(w_i/q_i) \quad (3)$$

Sujeito as restrições de equilíbrio e normalização

$$\sum_{\{i|D=0\}} w_i c_{ri}(X_i) = m_r \quad \text{com } r \in 1, \dots, R \quad (4)$$

$$\sum_{\{i|D=0\}} w_i = 1 \quad (5)$$

$$w_i \geq 0 \text{ para todo } i, \text{ tal que } D = 0 \quad (6)$$

Onde $q_i = 1/n$ é um peso base, n é o tamanho da amostra das unidades de controle, e $c_{ri}(X_i) = m_r$ descreve um conjunto de R restrições referentes aos momentos das covariadas no grupo de controle reponderados. Inicialmente, escolhe-se a covariada que será incluída na reponderação. Para cada covariada, especifica-se um conjunto de restrições de balanceamento (Eq. 04) para equiparar os momentos das distribuições das covariadas entre os grupos de tratamento e controles reponderados. As restrições de momentos podem ser a média (primeiro momento), a variância (segundo momento), e a assimetria (terceiro momento).

Dessa forma, o balanceamento por entropia procura, para um conjunto de unidades, pesos $W = [w_1, \dots, w_{n_0}]'$ no qual minimiza a equação (3), distancia de entropia entre W e o vetor base de pesos $Q = [q_1, \dots, q_{n_0}]'$, sujeita as restrições de balanceamento na equação (4), restrição de normalização (Equação 5), e restrição de não-negatividade (Equação 6).

Segundo Hainmueller (2012), o método da entropia pode ser combinado com outros métodos de pareamento, tais como, CEM e PEP, visando assim, equilibrar com maior robustez as covariadas. Nesse caso, o autor sugere aplicar inicialmente a entropia e em seguida o PEP, a partir dos pesos encontrados para fazer o balanceamento, conforme aplicado por Watson e Elliot (2016).

Dessa forma, este trabalho irá seguir esse procedimento, minimizando os desequilíbrios entre os grupos de tratados e controle. Para aplicação deste método, será utilizada uma variável binária para o tratamento, o qual irá permitir encontrar o efeito tratamento médio sobre os tratados das escolas profissionalizantes sobre a remuneração e as ocupações dos concludentes do ensino médio em 2011 e 2012 no mercado de trabalho em 2013. Posteriormente, será feita a análise de sensibilidade de Rosenbaum, discutida na seção 3.4

do capítulo 1, para verificação de presença de variáveis omitidas que influenciem o tratamento e as variáveis de resultado.

2.2 PEP com Múltiplos Tratamentos

Tendo em vista que durante o ensino profissionalizante o aluno tem a opção de escolher um curso, dentre vários, para se especializar, há que se adaptar a abordagem metodológica. Assim, visando encontrar o efeito dessa escolha sobre a remuneração e os tipos de ocupações no mercado de trabalho, este estudo optou por considerar uma nova variável de tratamento, o qual irá assumir múltiplos valores a partir dos tipos de cursos disponíveis.

Em decorrência dessa variável ser agora definida como categórica, será utilizado a metodologia do pareamento por score de propensão com múltiplos tratamentos, a qual é abordada por Rosenbaum e Rubin (1983), Imbens (2000), Lechner (2001). Esta difere da análise do efeito de tratamento único nas suposições de identificação, pois estas são modificadas para incorporar os múltiplos valores do tratamento.

A partir do procedimento de Imbens (2000) sobre o PEP, supõe-se que o aluno “*i*” dispõe de múltiplos tratamentos à sua escolha, a partir dos tipos de cursos ofertados, $T_i(t)$, onde o tratamento de interesse $t = 1, 2, \dots, k$ ¹⁹; $T_i(0)$ seria a categoria que o aluno optou por não fazer curso de especialização.

Para garantir uma suposição que possibilite a identificação de múltiplos tratamentos, Rosebaum e Rubin (1983) estabelecem que a Fraca Suposição de Independência Condicional (FSIC) requer somente a independência de cada tratamento com cada um dos resultados potenciais $Y(t)$. Além disso, essa suposição requer a hipótese da independência local entre cada resultado e o tratamento de interesse, o qual é independente de $D(t)$, indicador binário de tratamento, em vez do nível de tratamento T . Assim, se um aluno pertence ao tratamento $T_i(t)$, desconsideram-se as categorias anteriores e leva-se em conta somente a independência do $T_i(t)$ como se fosse uma variável de tratamento binário. Dessa forma, a FSIC é equivalente a:

$$D(t) \perp Y(t) | X \quad (7)$$

¹⁹ Estes cursos serão estabelecidos com base nas estatísticas descritivas da amostra.

Onde X corresponde ao conjunto de variáveis de pré-tratamento. Imbens (2000) relaciona a FSIC à interpretação do problema da inferência causal a partir de *missing data*. Dado que o objetivo é estimar a média do resultado potencial, $E[Y(t)]$, deve-se garantir a representatividade da média dos $Y_i(t)$ na subamostra das observações que recebem o tratamento $D_i(t) = 1$, embora quando $D_i(t) = 0$, a variável $Y_i(t)$ seja sempre desconhecida. Embora não exista uma função direta para o nível de tratamento, quando $D_i(t) = 0$, o que importa são os indivíduos desse grupo de comparação não receberem tratamento “ t ”. Porém, devido aos outros resultados potenciais $Y_i(s)$, para $s \neq t$, não se pode observar os mesmo resultados que as unidades com $D_i(t) = 1$. Assim, dado a condicionalidade das variáveis explicativas, a definição FSIC corresponde à falta de relevância desses outros resultados potenciais, o qual permite estimar o resultado médio condicionado pelo status do tratamento:

A partir das médias condicionais, os resultados médios podem ser estimados com:

$$E[Y(t)] = E[E[Y(t)|X]] \quad (8)$$

Dessa forma, o Escore de Propensão Generalizado (EPG) pode ser definido como a probabilidade condicional de receber um particular nível de tratamento, dada as variáveis observáveis, o qual, segundo Imbens (2000) iguala-se à esperança condicional do tratamento binário e pode ser representado por:

$$e(t, x) = \Pr(T = t|X = x) = E[D(t)|X = x] \quad (9)$$

A partir do EPG e a FSIC, essa suposição pode ser representada por:

$$D(t) \perp Y(t)|e(t, x) \quad (10)$$

Dado o EPG para múltiplos tratamentos, a suposição de sobreposição pode ser expressa por:

$$D(t) \perp X|e(t, x) \quad (11)$$

O argumento condicionante muda com o nível de tratamento e para garantir a independência condicional é necessário condicionar a todo o conjunto de $K + 1$ escores, $\{e(t, x)\} \in T$.

O procedimento da metodologia de pareamento por escore de propensão com múltiplos tratamentos é semelhante ao caso binário, onde inicialmente, estima-se o escore de propensão, $e(t, x)$, a partir de um modelo que incorpore múltiplos tratamentos. Posteriormente, estima-se o efeito médio no nível de tratamento t , $\beta(t) = E\{\beta[t, e(t, x)]\}$, o qual é estimado através do pareamento de *kernel* para cada nível de tratamento relativo aos

demais. Ao analisar os níveis de tratamento individualmente, $D(t) = 1$ significa que determinado aluno pertence a um tipo de tratamento t e $D(t) = 0$ significa que o aluno está em outra categoria.

2.3 Variáveis e Bases de Dados

Para avaliar o efeito da política de profissionalização no ensino médio das EEEP sobre o mercado de trabalho, construiu-se uma amostra com informações longitudinais a partir da junção das seguintes bases de dados: Censo escolar 2011 e 2012, e Relação Anual de Informações Sociais (RAIS) 2013. Dessa forma, foram identificados os alunos concludentes das escolas estaduais do Ceará em 2011 e 2012, e sua inserção no mercado de trabalho em 2013. Vale ressaltar que, como as informações da RAIS são de 2013, optou-se por identificar os egressos do ensino médio em 2011 e 2012, tanto para avaliar o efeito imediato, como permitir uma defasagem temporal para capturar eventual impedimento para o trabalho devido à idade do jovem ou tempo gasto por procura de emprego.

Para compor a amostra, observou-se inicialmente a lista dos concludentes disponibilizada pela SEDUC, dos quais 10.791 concluíram em 2011 e 12.164 em 2012, e fez-se o cruzamento com o censo escolar para obter informações sobre características pessoais e escolares desses alunos, tais como, sexo, raça, nacionalidade, idade, migração, escola de conclusão, turma, curso de profissionalização e localização da escola. Dos 22.955 concludentes, 11.066 (5.378 em 2011 e 5688 em 2012) estudaram nas escolas de ensino profissionalizante.

Em seguida, a partir da RAIS 2013, identificou-se que, 7.302 alunos²⁰ (4.171 em 2011 e 3131 em 2012) estavam no mercado de trabalho, os quais constituem a amostra resultante para análise. Para compor a análise do primeiro modelo (Pareamento por entropia), cuja variável de tratamento é binária, foram criados os seguintes grupos: Tratados – alunos que se profissionalizaram nesse período, estavam presentes no mercado de trabalho em 2013, onde foram observados 4.142 do ensino profissional. Controle – alunos advindos do ensino regular, perfazendo um total de 3.160 alunos.

²⁰ Aproximadamente 32% dos concludentes.

Ademais, dentre os cursos profissionalizantes observados na amostra, destacam-se saúde (32,7), informática (32), gestão e negócios (15,4), Hospitalidade e lazer (10,9) e outros (10). Com isso, para a análise do modelo com múltiplos tratamentos, considerou-se estes cinco tratamentos e computou-se a variável categórica curso, o qual está descrita no quadro 3.1. Neste caso, optou-se pela hipótese da distribuição logística para estimar os escores de propensão, considerando os múltiplos tratamentos.

As variáveis ocupação e remuneração média, descritas no quadro 3.1, serão utilizadas para verificar o impacto das escolas e dos cursos profissionalizante sobre o mercado de trabalho. Tendo em vista o elevado número de 596 ocupações criadas a partir da Classificação Brasileira de Ocupações (CBO 2002), e sua intratabilidade estatística, utilizou-se o procedimento de Jannuzzi (2001, 2003) para reagrupa-las em apenas cinco estratos sócio-ocupacionais, a partir de indicadores de renda, escolaridade e nível de precarização ocupacional, onde estes foram organizados com base no índice socioeconômico posicional (Isep)²¹, na ordenação das ocupações proporcionada pelos índices socioeconômicos, na tipologia de grupos ocupacionais de Valle Silva (1992) e na proposta metodológica de estratificação social segundo grupos ocupacionais de Goldthorpe (1992).

O primeiro estrato compreende as ocupações em posto de comando, de direção ou com especialização técnica superior. Reúne, portanto, as ocupações com rendimento e escolaridade mais elevadas, isto é, de maior status (Isep médio de 96), tais como, médico, Engenheiro, Professor Universitário, Empresários, Gerentes e postos superiores na Administração Pública (Juízes, Promotores, Delegados, Oficiais das Forças Armadas, etc.), geralmente, com rendimento mediano acima de R\$ 3.000, escolaridade de 16 anos ou mais, tempo médio de trabalho acima de 10 anos, baixo risco ao desemprego. O segundo estrato (Isep de 85) é composto pelos pequenos proprietários, chefes e supervisores e empregados qualificados de escritório e técnicos de média especialização na indústria e serviços. Técnicos de contabilidade e administração, Mestre e Contramestres na indústria, Professores de ensino fundamental e médio, Corretores de Imóveis, Inspectores de Polícia, Carteiros, Comerciantes (proprietários) e Agricultores são algumas das ocupações enquadradas neste grupo. Com menor poder de comando ou qualificação, tais ocupações apresentam um rendimento médio

²¹ Corresponde à porcentagem de indivíduos ocupados cujos níveis médios combinados de rendimento e escolaridade são menores ou iguais aos da ocupação considerada.

bem menor, ainda que em relação aos indicadores de escolaridade e precariedade as diferenças não sejam tão expressivas. O terceiro estrato (Isep de 69) reúne a maioria de ocupações no Comércio, Serviços e postos qualificados da Indústria. Ocupações típicas desse grupo são as de Torneiro Mecânico, Montadores de Equipamentos Elétricos, Vendedores, Operadores de caixa, Comerciantes conta-própria, Professores de ensino pré-escolar, Motoristas, Inspetores de alunos, Auxiliares de enfermagem, Auxiliares administrativos e de escritório, Policiais e Praças das Forças Armadas. O quarto grupo socioocupacional (Isep de 45) compreende os empregados em prestação de serviços de baixa qualificação nos serviços, construção civil e indústria tradicional. Nesta categoria estão reunidas as ocupações da Indústria de Alimentos, da Indústria Têxtil, Pedreiros, Pintores, Garçons, Vigias, Porteiros, Estivadores. Alguns indicadores de precarização chegam a ser ainda piores que os do estrato anterior: menos da metade contribui para a Previdência, metade tem sobrejornada e o risco ao desemprego é mais elevado. Na última categoria (Isep de 18), estão agrupados os trabalhadores rurais, além das ocupações urbanas de baixo status, como a de serventes de pedreiro, lavadeiras, empregadas domésticas e lixeiros, ou seja, as que possuem remunerações e escolaridade mais baixas e elevada precariedade no mercado de trabalho²².

Como covariadas utilizou-se as demais variáveis disponíveis no censo escolar, tais como, sexo do aluno, raça, migração, zona de residência, transporte escolar, faixas de idade e se possui deficiência, as quais estão descritas no quadro 3.1.

²² Para mais detalhes ver Jannuzzi (2001, 2002).

Quadro 3.1 - Variáveis na Amostra

Variável	Descrição	Fonte
Tratamentos		
EEEP	1 se o aluno concluiu ensino médio na escola profissionalizante; 0 se o aluno concluiu na escola regular	SEDUC/Censo escolar
Cursos	0 se não participou do curso profissionalizante; 1 se participou do curso de profissionalizante na área da saúde; 2 se participou do curso de profissionalizante na área de gestão; 3 se participou do curso de profissionalizante na área da informação; 4 se participou do curso de profissionalizante na área da Hospitalidade e Lazer e 5 se participou de curso de outras áreas.	SEDUC/Censo escolar
Resultados		
Rend_media	Remuneração anual média	RAIS
Ocupações	Indicador categórico que reflete nível de status socioeconômico e grau de precarização das ocupações, variando de 1 (alto status) a 5 (baixo status)	RAIS
Ocup_socio 1	1 se ocupações assumir valor igual a 1; 0 c.c	RAIS
Ocup_socio 2	1 se ocupações assumir valor igual a 2; 0 c.c	RAIS
Ocup_socio 3	1 se ocupações assumir valor igual a 3; 0 c.c	RAIS
Ocup_socio 4	1 se ocupações assumir valor igual a 4; 0 c.c	RAIS
Ocup_socio 5	1 se ocupações assumir valor igual a 5; 0 c.c	RAIS
Covariadas		
Sexo	1 se masculino; 0 caso contrário (c.c)	Censo escolar
Raça	1 se branca; 0 c.c	Censo escolar
Migrante	1 se não nasceu no estado do Ceará; 0 c.c	Censo escolar
Zona_resid	1 se reside em área urbana; 0 c.c	Censo escolar
Transp_escolar	1 se utiliza transporte público escolar; 0 c.c	Censo escolar
Idade	Anos de idade	Censo escolar
Idade1	1 se o indivíduo possui idade entre 17 e 25 anos	Censo escolar
Idade2	1 se o indivíduo possui idade entre 26 e 40 anos	Censo escolar
Deficiencia	1 se o indivíduo é portador de deficiência; 0 c.c	Censo escolar

Fonte: Elaboração própria

3 Resultados

3.1 Análise Descritiva

Estatísticas descritivas das variáveis de resultados e das variáveis de controles relacionadas aos alunos concludentes das EEEP e das escolas regulares do Ceará para os anos

de 2011 e 2012, extraídas das bases de dados do Censo escolar e RAIS 2013, estão explicitadas na Tabela 3.1.

Observa-se que, embora apresentando diferencial baixo, a remuneração média em 2013 dos alunos advindos das escolas profissionalizantes é superior aos alunos que concluíram nas escolas regulares para os dois anos analisado, aproximadamente 2 em 2011 e 1,15 em 2012 e apresentam variação relativa à média maior as estas de, 49 e 17, respectivamente. Além disso, independentemente do tipo de escola de conclusão, os alunos que terminaram em 2011 possuem rendimentos médios superiores aos que concluíram em 2012.

Referentes aos tipos de ocupações socioeconômicas, pode-se verificar que, os concludentes das EEEP se concentram, principalmente, na categoria 2, 47,7 dos que terminaram em 2011 e 53,4 em 2012, seguida da categoria 3, com 37 e 32, respectivamente. Para os advindos das escolas regulares, a categoria 3 apresenta-se como a de maior proporção, com 39 para 2011 e 41 para 2012, seguida da categoria 4, com 30 e 32 respectivamente. Por outro lado, a categoria 1 concentra as mais baixas proporções de trabalhadores de ambas as escolas, variando entre 2,4 e 3,6 nos dois.

Apesar dessas estatísticas apresentarem valores que favorecem aos alunos que concluíram nas escolas de ensino profissional, não se pode afirmar que o efeito destas sobre a remuneração e os tipos de ocupações no mercado de trabalho sejam positivas, pois o grupo de alunos advindos das escolas regulares não representam necessariamente o contrafactual de não tratamento. Dessa forma, se faz necessário técnicas mais apropriadas que busquem isolar o efeito desta política sobre as variáveis de resultado.

Comparando estes grupos referentes às variáveis de controle, observa-se, ainda na Tabela 3.1, que a proporção de homens advindos das escolares regulares é superior aos das profissionalizantes, tanto em 2011 quanto em 2012, em aproximadamente 19 e 29, respectivamente. Quanto à raça, a proporção de alunos considerados brancos é maior nas EEEP, porém houve uma redução, 17 em 2011 e 15 em 2012, enquanto que nas escolas regulares aumentou de 12,3 para 13,4 neste período.

Ademais, estes grupos apresentam estatística semelhante quanto à variável de migração, com proporção de 5,7 e 5,5, em 2011 e 5,9 e 6,6 em 2012, para as EEEP e regulares, respectivamente e quanto à variável que indica se o aluno apresenta alguma deficiência, não atingindo 1 nos dois tipos de escolas e nos dois anos. Porém, diferem nas

variáveis zona de residência, transporte escolar e faixas de idade, pois aproximadamente 90 dos alunos das profissionalizantes residem em áreas urbanas, enquanto que, apenas 56 dos alunos das escolas regulares residem neste tipo de área, em ambos os períodos. Referente a utilização de transporte escolar, a proporção de alunos é inferior nas EEEP tanto em 2011, quanto em 2012, porém neste último, a diferença é maior, 85 a mais para as escolas regulares, pois essa proporção nas escolas profissionalizantes reduziu 46, enquanto que nas regulares, a redução foi apenas 8. E referente às faixas de idades, observa-se que, aproximadamente, 99 dos alunos concludentes em 2011 e em 2012 nas escolas profissionalizantes possuem idade entre 18 e 25 anos, enquanto que nas regulares, essa proporção reduz para 86 em 2011 e 84 em 2012.

A Tabela 3.2 apresenta as estatísticas descritivas das variáveis de resultado no mercado de trabalho em 2013, remuneração e tipos de ocupação, na amostra 2011-2012, considerando os cursos escolhidos pelos alunos durante o 3º ano do ensino médio nas escolas de ensino profissionalizantes. Verifica-se que a área de saúde prowenha a maior remuneração média (R\$810,10), embora o desvio padrão das remunerações médias entre as áreas seja de apenas R\$51,15, e o Coeficiente de Variação das remunerações em Gestão seja mais que o dobro dos demais cursos.

Dentre os cursos profissionalizantes, a área de gestão apresenta a maior proporção de trabalhadores situados na ocupação socioeconômica do tipo 1, o qual é composta por ocupações Empresários, Gerentes e postos superiores na Administração Pública, etc., 5,48, seguido das áreas de informação e saúde, com 4 e 2,5, respectivamente. Referente a ocupação 2, compostas por técnicos de contabilidade e administração, Mestre e Contramestres na indústria, Professores de ensino fundamental e médio, Corretores de Imóveis, Inspetores de Polícia, Carteiros, Comerciantes (proprietários) e etc., destaca-se os cursos da área de saúde, com 56 da proporção dos concludentes das profissionalizantes, enquanto que os demais cursos apresentam proporções que variam entre 40 a 50.

Tabela 3.1 - Estatísticas descritivas das variáveis da amostra por tipos de escolas e por ano.

Variáveis	2011						2012					
	EEEP			Regulares			EEEP			Regulares		
	Média	D.P	CV	Média	D.P	CV	Média	D.P	CV	Média	D.P	CV
Rend_média 2013	768.90	392.36	0.5102	754.4400	258.9200	0.3432	740.9700	396.7300	0.5354	732.6100	335.5400	0.4580
Ocup_socio 1	0.0327	0.1779	5.4404	0.0341	0.1817	5.3284	0.0359	0.1862	5.1866	0.0237	0.1523	6.4262
Ocup_socio 2	0.4766	0.4995	1.0480	0.2611	0.4393	1.6825	0.5338	0.4989	0.9346	0.2299	0.4209	1.8308
Ocup_socio 3	0.3670	0.4820	1.3134	0.3885	0.4875	1.2548	0.3229	0.4677	1.4484	0.4067	0.4913	1.2080
Ocup_socio 4	0.1174	0.3219	2.7419	0.3022	0.4593	1.5199	0.1048	0.3065	2.9246	0.3193	0.4663	1.4604
Ocup_socio 5	0.0061	0.0781	12.8033	0.0138	0.1170	8.4783	0.0023	0.0485	21.0870	0.0202	0.1409	6.9752
Sexo	0.4799	0.4997	1.0413	0.5697	0.4952	0.8692	0.4437	0.4969	1.1199	0.5702	0.4951	0.8683
Raça	0.1689	0.3748	2.2191	0.1233	0.3289	2.6675	0.1496	0.3568	2.3850	0.1341	0.3409	2.5421
Migrante	0.0572	0.2324	4.0629	0.0555	0.2291	4.1279	0.0589	0.2355	3.9983	0.0663	0.2490	3.7557
Zona_resid	0.8850	0.3190	0.3605	0.5651	0.4958	0.8774	0.9004	0.2995	0.3326	0.5688	0.4954	0.8710
Transp_escolar	0.4203	0.4203	1.0000	0.4551	0.4981	1.0945	0.2256	0.4181	1.8533	0.4178	0.4933	1.1807
Idade1	0.9901	0.0986	0.0996	0.8604	0.3466	0.4028	0.9923	0.0872	0.0879	0.8427	0.3641	0.4321
Idade2	0.0077	0.0878	11.4026	0.1233	0.3289	2.6675	0.0053	0.0726	13.6981	0.1348	0.3417	2.5349
Idade3	0.0020	0.0451	22.5500	0.0162	0.1263	7.7963	0.0023	0.0485	21.0870	0.0223	0.1479	6.6323
Deficiencia	0.0045	0.0669	14.8667	0.0081	0.0896	11.0617	0.0005	0.0242	48.4000	0.0020	0.0457	22.8500

Fonte: Elaboração Própria com base nos dados da RAIS (2013) e Censo escolar (2011, 2012)

Tabela 3.2 - Estatística descritiva das variáveis de resultados por cursos, 2013

Cursos	Rem_média		Ocupações									
			1		2		3		4		5	
	Média	DP	Média	DP	Média	DP	Média	DP	Média	DP	Média	DP
Saúde	810.10	312.19	0.0250	0.1564	0.5660	0.4958	0.3217	0.4673	0.0811	0.2732	0.0059	0.0766
Gestão	796.90	722.48	0.0548	0.2278	0.5015	0.5003	0.3354	0.4725	0.1050	0.3068	0.0031	0.0559
Informação	711.24	265.21	0.0407	0.1977	0.4819	0.4998	0.3476	0.4764	0.1251	0.3310	0.0045	0.0671
Hosp_lazer	693.57	325.12	0.0244	0.1545	0.3977	0.4899	0.44	0.4969	0.1355	0.3426	0.0022	0.0471
Outros	750.57	288.69	0.0211	0.1440	0.4542	0.4987	0.3732	0.4845	0.1478	0.3556	0.0035	0.0593

Fonte: Elaboração Própria com base nos dados da RAIS (2013) e Censo escolar (2011, 2012)

Nota: Médias de cada ocupação por curso.

Na ocupação 3, composta por ocupações como, Torneiro Mecânico, Montadores de Equipamentos Elétricos, Vendedores, Operadores de caixa, Professores de ensino pré-escolar, Motoristas, Inspetores de alunos, Auxiliares de enfermagem, Auxiliares administrativos, etc., a área com maior proporção são os cursos da área de Hospitalidade e Lazer, com 44, seguido dos outros tipos de cursos, com 37, o contrário pode ser observado na categoria de ocupação 4, composta por Pedreiros, Pintores, Garçons, Vigias, Porteiros, etc., onde estes apresentam a maior proporção, aproximadamente 15, seguido da área de hospitalidade e lazer, porém, na categoria 5, o qual é composta por trabalhadores rurais, além das ocupações urbanas de baixo status, como a de serventes de pedreiro, lavadeiras, empregadas domésticas, lixeiros, e etc., todas as áreas apresentam proporção semelhante, menos de 1.

Ademais, pode-se verificar que, dos alunos que fizeram os cursos nas áreas de saúde, e gestão, mais de 50 encontram-se nas ocupações de categoria 2, seguido da ocupação 3, com aproximadamente, 32 e 33, respectivamente. Resultado semelhante, porém, com menor proporção na categoria 2, encontra-se a área de informação e outros tipos de cursos, enquanto que a área de hospitalidade e lazer apresenta maior proporção de trabalhadores nas ocupações socioeconômicas 3, seguido da categoria 2.

3.2 Resultados dos modelos

3.2.1 Resultados do Modelo Balanceamento por Entropia

Como primeira etapa deste modelo, considera-se a variável binária de tratamento para os concluintes de escolas profissionalizantes (tratados) e regulares (controle), referentes aos alunos que terminaram o ensino médio em 2011 e 2012. Na segunda etapa, observam-se algumas estatísticas das covariadas nos dois grupos, os quais estão expostas na Tabela 3.3, a fim de verificar, inicialmente, se os grupos estão equilibrados e divergem em características, para em seguida encontrar-se o efeito médio sobre os tratados.

As estatísticas do primeiro, segundo e terceiro momento da distribuição das covariadas e o teste de diferença entre os grupos estão contidas na Tabela 3.3. Observa-se que existem diferenças significativas na média, variância e assimetria dessas variáveis, exceto em migrante, cuja diferença de média não foi significativa. Dessa forma, os grupos não se

apresentam adequadamente balanceados, fazendo-se necessário aplicar o balanceamento por entropia, a fim de equilibrar as características entre os grupos.

O algoritmo busca equilibrar os valores dos momentos especificados na amostra de referência, neste caso, o grupo dos concludentes das profissionalizantes, e procura, a partir de um conjunto de pesos de entropia, ajustar o grupo de controle combinando-o com a amostra de referência.

Dessa forma, esta aplicação tem por objetivo encontrar pesos que minimizem as diferenças entre tratados e controles, considerando os três momentos da distribuição. A convergência ocorre quando todos os momentos são pareados entre os grupos dentro de um determinado número de interações e nível de tolerância²³.

Os resultados após o balanceamento por entropia são apresentados na Tabela 3.4, de onde se pode concluir que, em todas as covariadas, os três momentos da distribuição não diferem entre os grupos. Ou seja, os resultados para a média, variância e assimetria estão praticamente idênticos, portanto, as diferenças entre eles não são significativas. Dessa forma, conclui-se que os grupos estão perfeitamente equilibrados quantos as variáveis utilizadas na amostra, após a entropia.

²³ O número de iteração e o nível de tolerância padrão são 20 e 0.015, respectivamente, e podem ser aumentados se a convergência falhar.

Tabela 3.3 - Condições de Momentos das covariadas antes do balanceamento por entropia

Covariadas	Tratados			Controles			Diferenças			p-valor
	Média	Variância	Assimetria	Média	Variância	Assimetria	Média	Variância	Assimetria	
Ano 2012	0.4098	0.2419	0.3668	0.4531	0.2479	0.1883	-0.0433	-0.006	0.1785	0.000
Sexo	0.4651	0.2488	0.1399	0.57	0.2452	-0.2827	-0.1049	0.0036	0.4226	0.000
Raça	0.1611	0.1352	1.844	0.1282	0.1118	2.224	0.0329	0.0234	-0.38	0.000
Migrante	0.0579	0.0546	3.784	0.0604	0.0568	3.688	-0.0025	-0.0022	0.096	0.650
Zona_resid	0.8913	0.0968	-2.515	0.5668	0.2456	-0.2697	0.3245	-0.1488	-2.2453	0.000
Transp_escolar	0.2277	0.1759	1.299	0.4383	0.2463	0.2489	-0.2106	-0.0704	1.0501	0.000
Idade1	0.9911	0.0088	-10.44	0.8524	0.1258	-1.987	0.1387	-0.117	-8.453	0.000
Idade2	0.0067	0.0067	12.04	0.1286	0.1121	2.219	-0.1219	-0.1054	9.821	0.000
Deficiencia	0.0028	0.0028	18.5	0.0053	0.0053	13.52	-0.0025	-0.0025	4.98	0.094

Fonte: Elaboração Própria

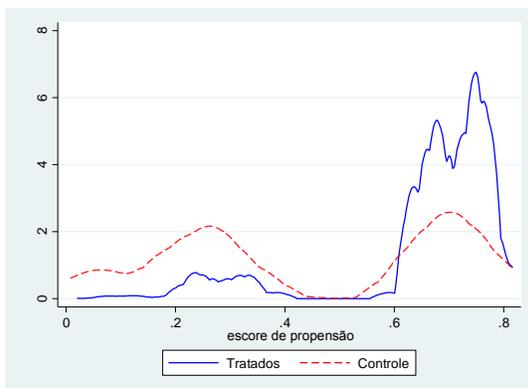
Tabela 3.4 - Condições de Momentos das covariadas após o balanceamento por entropia

Covariadas	Tratados			Controles			Diferenças			p-valor
	Média	Variância	Assimetria	Média	Variância	Assimetria	Média	Variância	Assimetria	
Ano 2012	0.4098	0.2419	0.3668	0.4098	0.2419	0.3669	0	0	-1E-04	0.999
Sexo	0.4651	0.2488	0.1399	0.4651	0.2489	0.14	0	-0.0001	-0.0001	0.999
Raça	0.1611	0.1352	1.844	0.1611	0.1352	1.844	0	0	0	0.998
Migrante	0.0579	0.0546	3.784	0.0579	0.0546	3.784	0	0	0	1.000
Zona_resid	0.8913	0.0968	-2.515	0.89113	0.0968	-2.515	0.00017	0	0	0.999
Transp_escolar	0.2277	0.1759	1.299	0.2277	0.1759	1.299	0	0	0	0.998
Idade1	0.9911	0.0088	-10.44	0.991	0.0089	-10.4	1E-04	-1E-04	-0.04	0.978
Idade2	0.0067	0.0067	12.04	0.0068	0.0067	11.99	-1E-04	0	0.05	0.975
Deficiencia	0.0028	0.0028	18.5	0.0028	0.0028	18.49	0	0	0.01	1.000

Fonte: Elaboração Própria

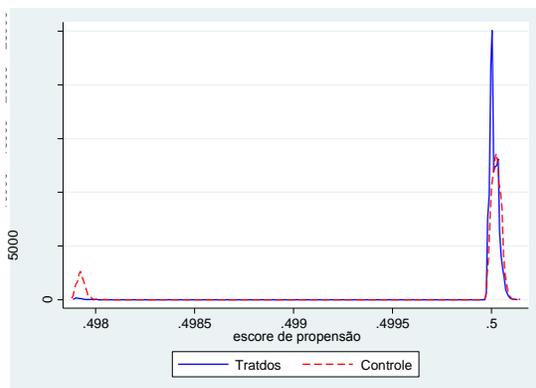
Segundo Hainmueller (2012) este procedimento pode ser combinado com o PEP, onde os escores de propensão são estimados considerando os pesos encontrados pelo balanceamento por entropia. Os gráficos 3.1 e 3.2 apresentam a distribuição da previsão linear desses escores via regressão logit antes e após a aplicação deste algoritmo, os quais demonstram que os escores de propensão tornaram-se balanceados nos dados ponderados via entropia.

Gráfico 3.1- Escore de Propensão antes do balanceamento por entropia



Fonte: Elaboração própria

Gráfico 3.2 - Escore de Propensão após o balanceamento por entropia



Fonte: Elaboração Própria

Considerando os escores de propensão ponderados, foram estimados os efeitos médio de tratamento sobre os tratados para as variáveis de resultado, remuneração e tipo de ocupação socioeconômica no mercado de trabalho em 2013, os quais estão expostos na Tabela 3.5. Verifica-se que o efeito do ensino profissionalizante não foi significativo sobre a remuneração e sobre a categoria socioeconômica 1 dos alunos concludentes em 2011 e 2012, porém, na ocupação do tipo 2, o efeito foi positivo e significativo, com diferencial de mais de 100 comparado ao grupo das escolas regulares. A partir da ocupação nível 3, os efeitos foram significativos e negativos, cujo menor e maior efeito foram observados na categoria 5 e 3, aproximadamente 73 e 12 a menos para as escolas profissionalizante, respectivamente.

Tabela 3.5 - Impacto das EEEP sobre a Remuneração e Ocupações dos Concludentes no Mercado de Trabalho – 2013

Variáveis de Resultado	EEEP	Regular	EMTT	()	DP	teste t
Valor_Remuneração	757.45	744.33	13.12	-	8.08	1.62
Ocup_socio1	0.0340	0.0294	0.0045	-	0.0041	1.12
Ocup_socio2	0.5001	0.2466	0.2534	102,7	0.0109	23.21
Ocup_socio3	0.3489	0.3970	-0.0480	-12	0.0114	-4.20
Ocup_socio4	0.1122	0.3100	-0.1977	-63,7	0.0095	-20.64
Ocup_socio5	0.0045	0.0168	-0.0122	-72,6	0.0025	-4.86

Fonte: Elaboração Própria

Embora com efeitos insignificantes para a remuneração e para primeira categoria das ocupações, os demais resultados mostram que os concludentes advindos das escolas profissionalizantes se concentram em ocupações com maiores rendimentos e melhores níveis educacionais no mercado de trabalho e com melhores índice socioeconômico posicional, no caso, a categoria 2, enquanto que reduz a proporção destes estarem em ocupações com menores rendimentos e baixa escolaridade, especialmente na categoria 4, cujo efeito foi 63 a menos, comparada aos alunos que concluíram nas escolas regulares.

Os resultados na Tabela 3.6 mostram que a robustez ao viés de seleção varia entre as variáveis de resultado. Referente às variáveis de resultados, remuneração e ocupação socioeconômica 1, embora os coeficientes tenham sido insignificantes, como apresentados na Tabela 3.5, não se pode confiar neles, pois foi identificado pelo teste de sensibilidade (Rosenbaum bounds) que existe problema de variáveis omitidas na estimação dos efeitos.

Tabela 3.6 - Análise de Sensibilidade (Rosenbaum bounds)

Γ	Valor_Rem		Ocup_soc1		Ocup_soc2		Ocup_soc3		Ocup_soc4		Ocup_soc5	
	P^+	P^-										
1	1.000	1.000	0.1495	0.149	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.05	1.000	1.000	0.248	0.080	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.1	1.000	1.000	0.368	0.040	0.000	0.000	0.000	0.013	0.000	0.000	0.000	0.000
1.15	1.000	0.999	0.497	0.019	0.000	0.000	0.000	0.095	0.000	0.000	0.000	0.000
1.2	1.000	0.999	0.432	0.008	0.000	0.000	0.000	0.332	0.000	0.000	0.000	0.000
1.25	1.000	0.979	0.318	0.003	0.000	0.000	0.000	0.362	0.000	0.000	0.000	0.000
1.3	1.000	0.830	0.223	0.001	0.000	0.000	0.000	0.123	0.000	0.000	0.000	0.000
1.35	1.000	0.467	0.149	0.000	0.000	0.000	0.000	0.026	0.000	0.000	0.000	0.000
1.4	1.000	0.139	0.095	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000
1.45	1.000	0.020	0.058	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.5	1.000	0.001	0.034	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.55	1.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.6	1.000	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1.65	1.000	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
1.7	1.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002
1.75	1.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004
1.8	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005
1.85	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007
1.9	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010
1.95	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.013
2	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Fonte: Elaboração Própria

Nas demais, o efeito do tratamento apresenta-se robusto a uma possível presença de viés de seleção, principalmente nas categorias 2, 3, 4 e 5 das ocupações socioeconômicas, apresentando valores críticos apenas na categoria 3, de F é de 1,2, 1,25 e 1,3 significando que se a presença de variáveis não observáveis levarem a uma diferença na *odds ratio* de receber o tratamento entre os grupos de tratamento e controle por esses fatores, então pode-se questionar o efeito das escolas profissionalizantes.

3.2.2 Resultados do Modelo com múltiplos tratamentos

O escore de propensão foi estimado como a probabilidade prevista de um aluno pertencer a uma das categorias da variável curso, descrita no quadro 1, onde esta foi estimada a partir de uma regressão de escolha discreta da variável de tratamento sobre as características observadas dos alunos presentes no censo escolar de 2011 e 2012, tais como, sexo, raça, idade, dentre outras.

O método de pareamento por escore de propensão busca eliminar o viés de seleção originado pelas variáveis observáveis, pareando os alunos nos grupos, através dos escores estimados, que simulem um experimento aleatório. Os resultados do modelo de escolha discreta estão expostos na Tabela 3.7, os quais foram estimados por um modelo multinomial logit. Os sinais dos coeficientes estimados estabelecem efeitos positivos ou negativos das variáveis sobre a probabilidade da variável discreta assumir os valores definidos. Vale ressaltar que, a categoria dos alunos que não participaram de cursos profissionalizantes foi considerada como grupo base para estimação desses resultados.

Observa-se, portanto, que os alunos que concluíram em 2012 reduzem a probabilidade de terem feito curso nas áreas de saúde, informação e hospitalidade, enquanto que aumentam as chances de terem participado de cursos na área de gestão. Referente a outros cursos, essa variável não foi significativa. Resultado semelhante pode ser verificado para os alunos do sexo masculino, embora com sinal negativo em gestão e positivo na área de informação. Apresentando efeitos positivos nas áreas de saúde, gestão e informação, encontram-se os alunos considerados brancos, porém, para a área de hospitalidade e lazer o efeito foi insignificante e para outras áreas, o efeito foi negativo.

Tabela 3.7 - Resultado do Modelo Multinomial Logit – Múltiplos Tratamentos

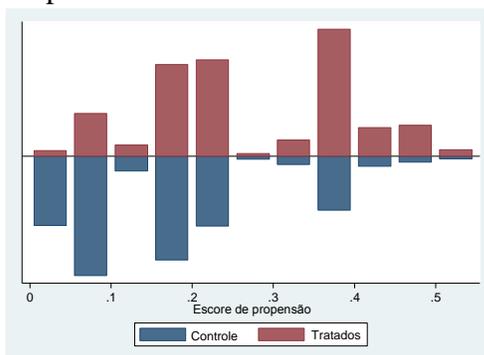
Variáveis	Tratamentos									
	Saúde		Gestão		Informação		Hosp_lazer		Outros	
	Coefficiente	p-valor								
Ano_2012	-0.724	0.000	0.603	0.000	-0.212	0.003	-0.423	0.000	-0.083	0.513
Sexo	-1.183	0.000	-0.289	0.002	0.222	0.002	-0.883	0.000	-0.039	0.758
Raça	0.290	0.003	0.426	0.001	0.273	0.005	-0.015	0.923	-0.476	0.032
Migrante	-0.063	0.661	-0.444	0.034	-0.002	0.986	-0.693	0.009	-0.584	0.068
Zona_resid	2.17	0.000	1.615	0.000	1.835	0.000	2.245	0.000	1.647	0.000
Transp_escolar	0.405	0.000	-0.738	0.000	0.226	0.020	-0.660	0.000	0.084	0.635
Idade1	3.580	0.001	2,139	0.003	3.496	0.001	1.783	0.014	1.258	0.083
Idade2	0.430	0.687	-1.19	0.177	-0.060	0.956	-0.992	0.247	-1.466	0.113
Deficiencia	-0.010	0.986	0.309	0.694	-0.496	0.455	0.792	0.248	-12.34	0.979
Constante	-5.301	0.000	-4.835	0.000	-5.712	0.000	-4.609	0.000	-4.603	0.000

Fonte: Elaboração Própria

Quanto aos alunos migrantes, o efeito foi negativo em todas as categorias, porém não foi significativo nas áreas de saúde e informação, enquanto que as variáveis que indica se a zona de residência é urbana e se os alunos pertencem a primeira faixa de idade foram positivas e significativas em todas as áreas. Além disso, pode-se verificar que alunos que utilizam transporte escolar público aumentam a probabilidade de terem feito cursos nas áreas de saúde e gestão, enquanto que, reduz as chances nas demais áreas, exceto em outras, pois o efeito foi insignificante. E referente as variáveis idade2 e deficiência, estas foram insignificantes para todas as categorias.

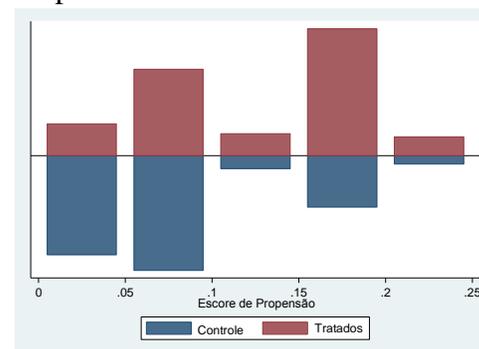
Este estudo utilizou o algoritmo de pareamento de *Kernel* no escore de propensão, desenvolvido em Heckman (1998), o qual combina todos os indivíduos na amostra de comparação ponderando com menores pesos as observações mais distantes. Essa metodologia pondera as observações no grupo de controle visando deixar semelhante ao grupo de tratamento, onde esse pareamento pode ser verificado a partir de uma análise de balanceamento das variáveis utilizadas na estimação do escore de propensão através de histogramas para os indivíduos dos dois grupos, os quais podem ser visualizados nos gráficos 3.3 a 3.7, para as áreas de saúde, gestão, informação, hospitalidade e lazer, e outras áreas, respectivamente. A partir destes, observa-se que as distribuições não diferem nos valores de escore de propensão, ou seja, os indivíduos apresentam escores de propensão que possibilitam o pareamento.

Gráfico 3.3 – Sobreposição do Escore de Propensão - Saúde



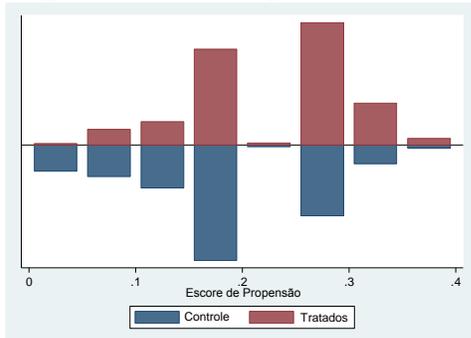
Fonte: Elaboração Própria

Gráfico 3.4 – Sobreposição do Escore de Propensão - Gestão



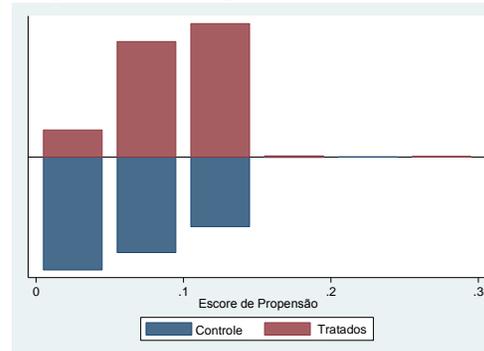
Fonte: Elaboração Própria

Gráfico 3.5 – Sobreposição do Escore de Propensão - Informação



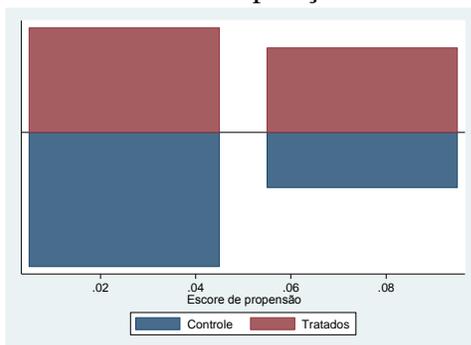
Fonte: Elaboração Própria

Gráfico 3.6 – Sobreposição do Escore de Propensão – Hospitalidade e Lazer



Fonte: Elaboração Própria

Gráfico 3.7 – Sobreposição do Escore de Propensão – Outras áreas



Fonte: Elaboração Própria

Após a estimação do escore de propensão e a eliminação da maior parte do viés atribuído as variáveis observadas, encontrou-se a estimativa do Efeito Médio de Tratamento sobre os Tratados (EMTT) a partir da diferença entre os resultados médios²⁴. As estimações do EMTT para as variáveis de resultado, remuneração e tipos de ocupações socioeconômicas no mercado de trabalho em 2013, estão expostas na Tabela 3.8.

Os resultados apontam efeitos positivos e significativos dos cursos nas áreas de saúde e gestão sobre a remuneração desses alunos, cujo efeito foi 88,05 e 59,15 reais, respectivamente, indicando que os alunos que fizeram estes cursos durante o ensino médio apresentam rendimentos superiores no mercado de trabalho aos demais alunos, aproximadamente, 12 e 8, respectivamente. Enquanto que os cursos voltados para as áreas

²⁴ Este estudo considera como principal efeito a ser investigado, o efeito dos tipos de cursos sobre os alunos participantes. Nesse sentido, o efeito a ser identificado é o Efeito Médio de Tratamento sobre o s tratados em vez do Efeito Médio de Tratamento.

de informação e hospitalidade e Lazer apresentam efeitos negativos sobre esta variável de resultado, ou seja, alunos que se profissionalizaram nessas áreas possuem menores rendimentos comparados ao grupo de controle, cujos efeitos foram -59,59 e -62,03, respectivamente. Referente aos demais cursos, o efeito não foi significativo.

Considerando inicialmente as ocupações socioeconômicas presentes no nível 1 como variável de resultado, observa-se que, apenas para os cursos da área de Gestão, o efeito foi positivo e significativo sobre a proporção de alunos que fizeram este curso e encontram-se em ocupações com maiores rendimentos e níveis mais elevados de escolaridade, tais como, tais como, médico, Engenheiro, Professor Universitário, Empresários, Gerentes e postos superiores na Administração Pública (Juízes, Promotores, Delegados, Oficiais das Forças Armadas, etc.). Ou seja, embora com proporção baixa, apenas 5,4 dos alunos que fizeram o curso de gestão, essa área apresenta maiores proporções de alunos presentes na primeira categoria das ocupações, em aproximadamente 90 superior ao grupo de controle, porém, verificou-se efeitos insignificantes para os demais cursos.

Tabela 3.8 - Impacto dos Múltiplos Tratamentos sobre as Variáveis de resultado – 2013

Cursos		Remuneração	Ocupações				
			1	2	3	4	5
Saúde	Tratados	810.10	0.025	0.566	0.322	0.081	0.005
	Controles	722.04	0.035	0.387	0.396	0.177	0.004
	EMTT	88.05	-0.009	0.178	-0.074	-0.095	0.001
	teste t	(12.19)	(-)	(45.99)	(-18.69)	(-53.67)	(-)
		8.29	-1.79	11.26	-4.89	-9.19	0.40
Gestão	Tratados	796.90	0.054	0.501	0.335	0.105	0.003
	Controles	737.75	0.028	0.404	0.385	0.176	0.005
	EMTT	59.15	0.025	0.097	-0.050	-0.071	-0.002
	teste t	(8.02)	(89.29)	(24.01)	(-12.99)	(-40.34)	(-)
		2.04	2.83	4.59	-2.49	-5.22	-0.87
Informação	Tratados	711.24	0.040	0.481	0.347	0.125	0.004
	Controles	770.83	0.033	0.367	0.387	0.202	0.009
	EMTT	-59.59	0.007	0.114	-0.039	-0.077	-0.004
	teste t	(-7.73)	(-)	(31.06)	(-10.08)	(-38.12)	(-44.44)
		-6.44	1.20	7.33	-2.64	-6.99	-1.91
Hosp_lazer	Tratados	693.57	0.024	0.397	0.44	0.135	0.002
	Controles	755.61	0.031	0.430	0.374	0.158	0.004
	EMTT	-62.03	-0.007	-0.032	0.065	-0.022	-0.002
	teste t	(-8.21)	(-)	(-)	(17.38)	(-)	(-)
		-3.85	-0.96	-1.36	2.69	-1.33	-1.03
Outros	Tratados	750.57	0.021	0.454	0.373	0.147	0.003
	Controles	751.17	0.032	0.391	0.372	0.193	0.009
	EMTT	-0.60	-0.011	0.062	0.0005	-0.046	-0.006
	teste t	(-)	(-)	(15.86)	(-)	(-23.83)	(-)
		-0.03	-1.26	2.08	0.02	-2.13	-1.62

Fonte: Elaboração Própria

Para as ocupações que se encontram no nível 2, efeitos positivos e significativos podem ser verificados em todas as áreas dos cursos, exceto em hospitalidade e lazer. Cabe destacar que o maior impacto foi encontrado na área da saúde, seguido da área da informação, 46 e 31, respectivamente, a mais para os alunos que fizeram esse curso sobre a proporção de alunos trabalhando nessa categoria de ocupação. Assim, alunos que se profissionalizaram nessas áreas durante o ensino médio, apresentam-se em maiores proporções no mercado de trabalho em ocupações com índice socioeconômico posicional de 85, como por exemplo, técnicos de contabilidade e administração, Mestre e Contramestres na indústria, Professores de ensino fundamental e médio, Corretores de Imóveis, Inspetores de Polícia, Carteiros, Comerciantes (proprietários).

Referente aos níveis de ocupações socioeconômicas 3 e 4, se pode verificar efeitos negativos e significativos para os cursos das áreas de saúde, gestão e informação, enquanto que para as áreas de hospitalidade e lazer e outras áreas, o efeito foi positivo e insignificante para a categoria 3 e insignificante e negativo para o nível 4, respectivamente. Quanto ao último nível de ocupações, apenas a área de informação mostrou-se significativa, porém com efeito quase nulo.

Para verificar a robustez dos resultados quanto a presença de variáveis omitidas no modelo, aplicou-se o teste de sensibilidade considerando os níveis de tratamento para as variáveis remuneração e ocupações, cujos resultados estão expostos nas tabelas 3.9 e 3.10. Os resultados mostram que, para a variável de resultado remuneração, os efeitos apresentam-se robustos a uma possível presença de viés de seleção, exceto para a categoria de tratamento gestão. Referente aos tipos de ocupações socioeconômicas, os efeitos dos cursos das áreas de hospitalidade e lazer podem ser questionados quanto a presença de um possível viés de variáveis omitidas.

Tabela 3.9 - Análise de Sensibilidade - Remuneração

Γ	Saúde		Gestão		Informação		Hosp_lazer		Outros	
	P^+	P^-	P^+	P^-	P^+	P^-	P^+	P^-	P^+	P^-
1	0.000	0.000	1.000	1.000	0.000	0.000	0.000	0.000	0.001	0.001
1.05	0.000	0.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.005
1.1	0.000	0.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.012
1.15	0.000	0.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.027
1.2	0.000	0.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.054
1.25	0.000	0.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.094
1.3	0.000	0.000	1.000	0.999	0.000	0.000	0.000	0.000	0.000	0.151
1.35	0.000	0.000	1.000	0.999	0.000	0.000	0.000	0.000	0.000	0.223
1.4	0.000	0.000	1.000	0.999	0.000	0.000	0.000	0.000	0.000	0.308
1.45	0.000	0.000	1.000	0.999	0.000	0.000	0.000	0.000	0.000	0.401
1.5	0.000	0.000	1.000	0.999	0.000	0.000	0.000	0.000	0.000	0.497
1.55	0.001	0.000	1.000	0.997	0.000	0.000	0.000	0.000	0.000	0.590
1.6	0.004	0.000	1.000	0.992	0.000	0.000	0.000	0.000	0.000	0.675
1.65	0.017	0.000	1.000	0.982	0.000	0.000	0.000	0.000	0.000	0.750
1.7	0.048	0.000	1.000	0.964	0.000	0.000	0.000	0.000	0.000	0.813
1.75	0.111	0.000	1.000	0.933	0.000	0.000	0.000	0.000	0.000	0.863
1.8	0.214	0.000	1.000	0.886	0.000	0.000	0.000	0.000	0.000	0.903
1.85	0.353	0.000	1.000	0.823	0.000	0.000	0.000	0.000	0.000	0.932
1.9	0.510	0.000	1.000	0.743	0.000	0.000	0.000	0.000	0.000	0.954
1.95	0.663	0.000	1.000	0.649	0.000	0.000	0.000	0.000	0.000	0.969
2	0.789	0.000	1.000	0.549	0.000	0.000	0.000	0.000	0.000	0.980

Fonte: Elaboração Própria

Tabela 3.10 - Análise de Sensibilidade - Ocupações

Γ	Saúde		Gestão		Informação		Hosp_lazer		Outros	
	P^+	P^-	P^+	P^-	P^+	P^-	P^+	P^-	P^+	P^-
1	0.056	0.056	0.000	0.000	0.033	0.033	0.201	0.201	0.179	0.179
1.05	0.032	0.093	0.001	0.000	0.064	0.016	0.159	0.249	0.149	0.213
1.1	0.017	0.142	0.003	0.000	0.109	0.007	0.124	0.300	0.123	0.249
1.15	0.009	0.203	0.006	0.000	0.171	0.003	0.096	0.352	0.101	0.285
1.2	0.005	0.273	0.011	0.000	0.248	0.001	0.074	0.405	0.083	0.323
1.25	0.002	0.350	0.020	0.000	0.336	0.000	0.057	0.457	0.068	0.360
1.3	0.001	0.431	0.033	0.000	0.431	0.000	0.043	0.508	0.056	0.397
1.35	0.000	0.511	0.051	0.000	0.525	0.000	0.033	0.557	0.045	0.434
1.4	0.000	0.486	0.075	0.000	0.445	0.000	0.024	0.521	0.037	0.469
1.45	0.000	0.413	0.105	0.000	0.360	0.000	0.018	0.477	0.030	0.504
1.5	0.000	0.345	0.142	0.000	0.283	0.000	0.014	0.435	0.024	0.538
1.55	0.000	0.283	0.185	0.000	0.217	0.000	0.007	0.395	0.020	0.570
1.6	0.000	0.229	0.233	0.000	0.162	0.000	0.005	0.357	0.016	0.565
1.65	0.000	0.183	0.286	0.000	0.119	0.000	0.004	0.322	0.013	0.537
1.7	0.000	0.144	0.342	0.000	0.085	0.000	0.003	0.289	0.010	0.510
1.75	0.000	0.112	0.400	0.000	0.060	0.000	0.002	0.259	0.008	0.483
1.8	0.000	0.086	0.458	0.000	0.041	0.000	0.001	0.231	0.006	0.458
1.85	0.000	0.065	0.516	0.000	0.028	0.000	0.001	0.206	0.005	0.433
1.9	0.000	0.049	0.502	0.000	0.018	0.000	0.000	0.182	0.004	0.409
1.95	0.000	0.036	0.447	0.000	0.012	0.000	0.000	0.161	0.003	0.386
2	0.000	0.026	0.394	0.000	0.008	0.000	0.000	0.142	0.002	0.364

Fonte: Elaboração Própria

Conclusões

Este trabalho procura ampliar o debate existente na literatura que relaciona educação e mercado de trabalho. Para tanto, busca-se trazer novas evidências para as escolas de ensino profissionalizante, ao se comparar o desempenho no mercado de trabalho dos alunos dessas escolas com outros do ensino regular. Dessa forma, busca-se encontrar o efeito médio sobre os tratados dessas escolas sobre a remuneração e as ocupações dos concludentes do ensino médio no mercado de trabalho.

Além disso, tendo em vista que durante o ensino profissionalizante o aluno tem a opção de escolher um curso, dentre vários, para se especializar, faz-se necessário adaptar a abordagem metodológica. Assim, visando encontrar o efeito dessa escolha sobre a remuneração e os tipos de ocupações no mercado de trabalho, este estudo optou por considerar uma nova variável de tratamento, o qual irá assumir múltiplos valores a partir dos tipos de cursos disponíveis.

Uma das contribuições deste estudo reside na base de dados utilizada para mensurar este efeito, pois através da construção de identificadores em parceria com a SEDUC foi possível construir uma amostra com informações longitudinais a partir da junção das seguintes bases de dados: Censo escolar 2011 e 2012, e Relação Anual de Informações Sociais (RAIS) 2013. Dessa forma, foram identificados os alunos concludentes das escolas estaduais do Ceará em 2011 e 2012, e sua inserção no mercado de trabalho em 2013.

Para atender ao primeiro objetivo, este trabalho aplicou os procedimentos feitos por Watson e Elliot (2016), os quais combinam o método da entropia, desenvolvido por Hainmueller (2012), com o pareamento por escore de propensão - PEP, visando assim, equilibrar com maior robustez as covariadas, minimizando os desequilíbrios entre os grupos de tratados e controle. Para aplicação deste método, utilizou-se uma variável binária para o tratamento. Posteriormente, em decorrência da mudança da variável de tratamento, agora definida como categórica, utilizou-se a metodologia do pareamento por escore de propensão com múltiplos tratamentos, a qual é abordada por Rosenbaum e Rubin (1983a), Imbens (2000), Lechner (2001). Para verificação de presença de variáveis omitidas que influenciem o tratamento e as variáveis de resultado foi feita a análise de sensibilidade de Rosenbaum.

No primeiro modelo, inicialmente, observou-se que, antes de aplicar o pareamento por entropia, existiam diferenças significativas na média, variância e assimetria entre os grupos

de tratados e controle. Dessa forma, os grupos não se apresentam adequadamente balanceados, porém, após a aplicação desse método, todas as covariadas, nos três momentos da distribuição, não diferiram entre os grupos. Ou seja, os resultados para a média, variância e assimetria passaram a ser praticamente idênticos, portanto, as diferenças entre eles não são significativas.

Seguindo os procedimentos de Watson e Elliot (2016), combinou-se esse método com o PEP, onde os escores de propensão foram estimados considerando os pesos encontrados pelo balanceamento por entropia. Desses resultados, verificou-se que os escores de propensão tornaram-se mais balanceados nos dados ponderados via entropia e com isso, encontrou-se o efeito médio tratamento sobre os tratados, onde foi possível observar que o efeito do ensino profissionalizante não foi significativo sobre a remuneração e sobre a categoria socioeconômica 1 dos alunos concludentes em 2011 e 2012, porém, na ocupação do tipo 2, o efeito foi positivo e significativo, com diferencial de mais de 100 comparado ao grupo das escolas regulares e a partir da ocupação nível 3, os efeitos foram significativos e negativos, cujo menor e maior efeito foram observados na categoria 5 e 3, aproximadamente 73 e 12 a menos para as escolas profissionalizante, respectivamente.

Os resultados da análise de sensibilidade mostraram que a robustez ao viés de seleção varia entre as variáveis de resultado, onde, o efeito do tratamento apresenta-se robusto a uma possível presença de viés de seleção, principalmente nas categorias 2, 3, 4 e 5 das ocupações socioeconômicas.

Para o modelo com múltiplos tratamentos, inicialmente, estimou-se um modelo multinomial logit para se encontrar a probabilidade prevista de um aluno pertencer a uma das categorias da variável curso. As estimativas do Efeito Médio de Tratamento sobre os Tratados (EMTT) encontrada a partir do pareamento de *Kernel* sobre as remunerações mostraram-se positivas e significativas somente para os cursos nas áreas de saúde e gestão, cujo efeito foi 12 e 8, respectivamente, indicando que os alunos que fizeram estes cursos durante o ensino médio apresentam rendimentos superiores no mercado de trabalho aos demais alunos.

Considerando as ocupações socioeconômicas presentes no nível 1 como variável de resultado, observou-se que, apenas para os cursos da área de Gestão, o efeito foi positivo e significativo. Embora com proporção baixa, apenas 5,4 dos alunos que fizeram o curso de

gestão estavam contidos nessa categoria, essa área apresenta proporções superiores ao grupo de controle, em aproximadamente 90, porém, verificou-se efeitos insignificantes para os demais cursos. Para as ocupações que se encontram no nível 2, efeitos positivos e significativos podem ser verificados em todas as áreas dos cursos, exceto em hospitalidade e lazer, cujo maior impacto foi encontrado na área da saúde, seguido da área da informação, 46 e 31, respectivamente, a mais para os alunos que fizeram esse curso sobre a proporção de alunos trabalhando nessa categoria de ocupação. Referente aos níveis de ocupações socioeconômicas 3 e 4, se pode verificar efeitos negativos e significativos para os cursos das áreas de saúde, gestão e informação, enquanto que para as áreas de hospitalidade e lazer e outras áreas, o efeito foi positivo e insignificante para a categoria 3 e insignificante e negativo para o nível 4, respectivamente. Quanto ao último nível de ocupações, apenas a área de informação mostrou-se significativa, porém com efeito quase nulo.

Os resultados do teste de sensibilidade encontrados para verificar a presença de variáveis omitidas no modelo, mostraram que, para a variável de resultado remuneração, os efeitos apresentam-se robustos a uma possível presença de viés de seleção, exceto para a categoria de tratamento gestão. Referente aos tipos de ocupações socioeconômicas, os efeitos dos cursos das áreas de hospitalidade e lazer podem ser questionados quanto a presença de um possível viés de variáveis omitidas.

Esses resultados indicam que, os efeitos sobre a remuneração e os tipos de ocupações socioeconômicas diferem quanto ao tipo de curso escolhido durante o ensino médio, onde o curso na área de saúde se mostrou com maior efeito sobre as remunerações. E sobre os níveis de ocupações, destacam-se as áreas de saúde, gestão e informação, os quais possuem maiores proporções de alunos presentes nas ocupações que exigem mais escolaridade e proporcionam maiores retornos financeiros, 1 e 2, enquanto que apresentam menores proporções nas categorias 3, 4 e 5.

Dessa forma, se o foco do aluno for maiores remunerações e ocupar melhores postos de trabalho após o término do ensino médio, não adianta apenas participar do ensino profissionalizante, é necessário escolher dentre os cursos oferecido aqueles que possuem efeito positivo e significativo sobre essas variáveis de resultado.

Referências

- ARAÚJO, A. J. N; CHEIN, F.; PINTO, C. **Ensino Profissionalizante, Desempenho Escolar e Inserção Produtiva: Uma Análise com dados do ENEM**. 2014. Disponível: <http://www.anpec.org.br/encontro/2014/submissao/files_I/i12-ac3a8f487db438fc6278e117ab468a01.pdf> Acesso em: 08 agosto 2016
- ASSUNÇÃO J; GONZAGA G. Educação Profissional no Brasil: Inserção e retorno. **Série Cenários**, n. 3, Brasília: SENAI/DN, 2010.
- BASSI, M. BUSSO, M. URZÚA, S. VARGAS, J. **Desconectados: Habilidades, educación y empleo en América Latina**, BID, 2012.
- GOLDTHORPE, J.H. **Social mobility and class structure in modern britain**. New York: Oxford University Press, 1992.
- GUIMARÃES, A. Q; ALMEIDA, M. E. Os jovens e o mercado de trabalho: evolução e desafios da política de emprego no Brasil. **Temas de Administração Pública**. v. 8, n. 2, 2013. Disponível em: <http://seer.fclar.unesp.br/temasadm/article/view/6845>. Acesso em: 12 setembro 2016.
- HAINMUELLER J. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. **Political Analysis**, v. 20 n.1, p. 25-46, 2012.
- HAINMUELLER J.; XU, Y. Ebalance: A Stata Package for Entropy Balancing. **Journal of Statistical Software**. v. 54, n. 7. August, 2013.
- HIRANO, K; IMBENS, G; RIDDER G. Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. **Econometrica**, v. 71 n.4, p.1161-1189, 2003
- HO, D., IMAI, K; KING, G. Stuart E, A. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. **Political Analysis**, v.15 n.3, p.199. 2007
- IMBENS, G. The role of the propensity score in estimating dose–response functions. **Biometrika**. v. 87, n.3, p.706–710, 2000.
- IMBENS, G. Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review. **Review of Economics and Statistics**, v.86 n.1, p.4-29, 2004
- JANNUZZI, P. M. Status socioeconômico das ocupações brasileiras: medidas aproximativas para 1980, 1991 e anos 90. **Revista Brasileira de Estatística**, Rio de Janeiro, v. 61, n. 2, p. 47-74, 2001.
- JANNUZZI, P. M. **Indicadores sociais no Brasil**. Campinas: Alínea, 2003, 2.ed.
- LECHENER, M. Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. **Econometric Evaluation of Labour Market Policies**. Heidelberg, p. 1–18. 2001
- LIN, D.; LUTTER, R.; RUHM, C. J. **Cognitive Performance and Labor Market Outcomes**. Discussion Paper, IZA, n. 10075. July, 2016.

ROSENBAUM, P.; RUBIN, D. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. **Journal of the Royal Statistical Society. Series B**, v.45, 212–218. 1983

RUBIN, D. B. **Matched Sampling for Causal Effects**. Cambridge University Press. 2006

SEKHON, J. S. Opiates for the Matches: Matching Methods for Causal Inference. **Annual Review of Political Science**, v. 12, p.487-508, 2009

SEVERNINI, E. R.; ORELLANO, V. I. F. O efeito do ensino profissionalizante sobre a probabilidade de inserção no mercado de trabalho e sobre a renda no período pré-Planfor. **Revista EconomiA**, 2010.

SILVA, M. R.; PELISSARI, L. B.; [STEIMBACH, A. A.](#) Juventude, escola e trabalho: permanência e abandono na educação profissional técnica de nível médio. **Educ. Pesqui.** [online]. v.39, n.2, p.403-417. 2013.

SIMÕES, C. A. Políticas públicas do ensino médio: realidade e desafios. In: FERREIRA, Cristina Araripe (Org.). *Juventude e iniciação científica: políticas públicas para o ensino médio*. Rio de Janeiro: EPSJV; UFRJ, 2010. p. 135-142.

VALLE SILVA, N. **Uma proposta de classificação das ocupações brasileiras**. LNCC, Mimeografado. 1992.

WATSON, S. ELLIOT, M. Entropy Balancing: A maximum-entropy reweighting scheme to adjust for coverage error. **Quality & Quantity**. v. 50, [n. 4](#), p. 1781–1797, July 2016.

CONCLUSÃO GERAL

Essa tese é composta por três artigos que tem como foco a educação pública, com ênfase no ensino profissionalizante. Embora possuam o mesmo tema, cada trabalho utiliza metodologias e base de dados distintas, os quais permitem responder problemas diferentes associados a essa modalidade de ensino.

O primeiro artigo denominado “Avaliação do Ensino Profissionalizante com Dados Longitudinais” avaliou o desempenho dos alunos de Escolas Estaduais de Ensino Profissionalizante (EEEP) nos exames do ENEM. A partir da criação de uma base longitudinal e da aplicação do algoritmo CEM foi possível controlar o efeito-aluno, e através da combinação do método do Lasso e Pós-Lasso com o PEP encontrou-se o efeito escola, o qual mostrou-se positivo e significativo em todas as áreas de conhecimento e na média geral do ENEM, destaque para Redação, cujo efeito foi de 39,9 a mais na nota para um aluno da EEEP.

No segundo artigo, que tem como título “Decomposição Quantílica Incondicional dos Diferenciais de Desempenho entre Ensino Profissionalizante e Privado”, objetivou-se trazer novas evidências sobre os diferenciais de desempenho entre o ensino público e privado, analisando as diferenças nas distribuições de notas no ENEM 2014, para as cinco áreas de conhecimento e na média geral, entre alunos das escolas de ensino profissionalizante e alunos das escolas privadas.

A partir da aplicação do método de Regressão Quantílica Incondicional (RQI) e da decomposição de Oaxaca-Blinder, observou-se diferenciais positivos e significativos, em todas as áreas, a favor dos alunos das escolas particulares em relação aos alunos do ensino profissional, principalmente em Redação, cujos diferenciais foram os mais elevados ao longo de toda distribuição, oscilando entre 12 e 16, aproximadamente. Além disso, verificou-se que, o efeito característico apresentou-se negativo apenas em alguns quantis da parte inferior da distribuição das notas, agindo assim, como redutor dos diferenciais de desempenho. Porém, na maioria dos casos, esse efeito foi positivo e significativo, atuando de forma a contribuir com o diferencial de notas, ou seja, os alunos das escolas privadas apresentam

melhores características que contribuem para um bom desempenho, principalmente nos quantis mais elevados da distribuição. Já, os efeitos não observáveis são todos positivos e significativos em todos os quantis da distribuição e em todas as áreas do ENEM e são responsáveis por explicar a maior parte do diferencial das notas, com destaque para matemática, cujo efeito é crescente ao longo da curva.

No terceiro e último artigo, cujo título é “Ensino Profissionalizante e Inserção no Mercado de Trabalho”, buscou-se comparar o desempenho no mercado de trabalho dos alunos das EEEP com outros do ensino regular. Dessa forma, objetivou-se encontrar o efeito médio sobre os tratados dessas escolas e da escolha do tipo de curso sobre a remuneração e as ocupações dos concludentes do ensino médio no mercado de trabalho. A partir de uma amostra com informações longitudinais, identificou-se os alunos concludentes das escolas estaduais do Ceará em 2011 e 2012, e sua inserção no mercado de trabalho em 2013.

Seguindo os procedimentos feitos por Watson e Elliot (2016), os quais combinam o método da entropia, desenvolvido por Hainmueller (2012), com o pareamento por escore de propensão – PEP, verificou-se que os escores de propensão tornaram-se mais balanceados nos dados ponderados via entropia e com isso, encontrou-se o efeito médio tratamento sobre os tratados, onde foi possível observar que o efeito do ensino profissionalizante não foi significativo sobre a remuneração e sobre a categoria socioeconômica 1 dos alunos concludentes do ensino médio, porém, na ocupação do tipo 2, o efeito foi positivo e significativo, com diferencial de mais de 100 comparado ao grupo das escolas regulares e a partir da ocupação nível 3, os efeitos foram significativos e negativos.

Posteriormente, estimou-se um modelo multinomial logit para se encontrar a probabilidade prevista de um aluno pertencer a uma das categorias da variável curso. Os efeitos sobre as remunerações mostraram-se positivas e significativas somente para os cursos nas áreas de saúde e gestão, cujo efeito foi 12 e 8, respectivamente, indicando que os alunos que fizeram estes cursos durante o ensino médio apresentam rendimentos superiores no mercado de trabalho aos demais alunos. Sobre os níveis de ocupações, destacam-se as áreas de saúde, gestão e informação, os quais possuem maiores proporções de alunos presentes nas ocupações que exigem mais escolaridade e proporcionam maiores retornos financeiros, 1 e 2, enquanto que apresentam menores proporções nas categorias 3, 4 e 5.

Os resultados aqui encontrados permitem afirmar que as escolas profissionalizantes apresentam desempenho superior a escolas regulares, demonstrando assim, possuir diretrizes mais eficazes em potencializar habilidades do conhecimento cognitivo relativo ao ensino regular, resultado que contribui para subsidiar futuras decisões aos formuladores de políticas educacionais. Porém, cabe destacar que ainda apresentam diferenciais positivos e significativos a favor das escolas particulares. O foco das EEEP não se restringe apenas a preparar o jovem para o ingresso no ensino superior, esta também procura capacitar os alunos para a inserção no mercado de trabalho. Dessa forma, após controlar por um conjunto de características observáveis, pode-se afirmar que esse tipo de ensino proporciona melhores remunerações e ocupações no mercado de trabalho, comparados ao ensino regular, porém deve-se levar em consideração a escolha do curso que se pretende fazer durante o ensino médio, pois estes apresentam impactos diferentes sobre o mercado de trabalho.

Embora essa tese tenha procurado responder a problemas pertinentes sobre esse tipo de ensino público, ainda existem questões a serem indagadas e que merecem destaques, tais como, o efeito dessas escolas sobre a evasão e repetência dos alunos; a tomada de decisão de forma simultânea após o ensino médio, o qual permita identificar aqueles jovens que ingressaram em alguma universidade e/ou no mercado de trabalho ou nenhuma dessas; a relação de custo benefício de implantação e permanência dessas escolas comparadas a de ensino regular.

REFERÊNCIAS

- ANGRIST, J., PISCHKE, J.S. **Most Harmless Econometrics: an empiricist's companion**. Princeton, New Jersey: Princeton University Press. 2009
- ARAÚJO, A. J. N; CHEIN, F.; PINTO, C. **Ensino Profissionalizante, Desempenho Escolar e Inserção Produtiva: Uma Análise com dados do ENEM**. 2014 Disponível: <http://www.anpec.org.br/encontro/2014/submissao/files_I/i12-ac3a8f487db438fc6278e117ab468a01.pdf>
- AROCA, P., BRIDA, J.G, VOLO, S. Tourism statistics: correcting data inadequacy using coarsened exact matching. Working Paper. **School of Economics and Management at the Free University of Bozen**, 2014.
- ASSUNÇÃO J; GONZAGA G. Educação Profissional no Brasil: Inserção e retorno. **Série Cenários**, n. 3, Brasília: SENALDN, 2010.
- BACALHAU, P.; MATTOS, E. **A Provisão Pública de Ensino Superior como Mecanismo de Seleção por Habilidade: Evidências para o Brasil**. 2013 Disponível em: http://www.anpec.org.br/encontro/2013/files_I/i12-e15541f20ab44aacc725c19288802a8d.pdf Acesso em: 10 Jan. 2016 .
- BARRO, Robert J. Education and Growth. **Annals of Economics and Finance**, v.14, n.2, p.301-328, 2013.
- BASSI, M. BUSSO, M. URZÚA, S. VARGAS, J. **Desconectados: Habilidades, educación y empleo en América Latina**, BID, 2012.
- BECKER, S., ICHINO, A. Estimation of Average Treatment Effects based on Propensity Scores. **The Stata Journal**, v.2 n.4, p.358-377. 2002
- BELLONI et al. High Dimensional Methods and Inference on Structural and Treatment Effects. **Journal of Economic Perspectives**, v. 28, n2, p. 29-50, 2014.
- BELLONI, A; HANSEN, C.; CHERNOZHUKOV, V. Inference Methods for High Dimensional Sparse Econometric Models. **Advances in Economics and Econometrics**, 10th World Congress of the Econometric Society, 2013.
- BELLONI, A; HANSEN, C.; CHEN, D.; CHERNOZHUKOV, V. **LASSO Methods for Gaussian Instrumental Variable Models**. Working paper, Duke University, 2010.
- BELLONI, A; HANSEN, C.; CHEN, D.; CHERNOZHUKOV, V. Sparse Models and Methods for Optimal Instruments with Application to Eminent Domain. **Econometrica**, v.80, n.6, p. 2369-2429, 2012.
- BELLONI, A; HANSEN, C.; FERNANDEZ-VAL, I; CHERNOZHUKOV, V. **Program Evaluation with High-Dimensional data**. CeMMAP working papers. Centre for Microdata Methods and Practice, Institute for Fiscal Studies, 2015.
- BICKEL, P. J.; RITOV, Y.; TSYBAKOV, A. B. Simultaneous analysis of Lasso and Dantzig selector, **Annals of Statistics**, v.37, n.4, p. 1705-1732, 2009.
- BLACKWELL, M. IACUS, S. KING, G. PORRO, G. CEM: Coarsened exact matching in Stata. **The Stata Journal**. v. 9, n. 4, p. 524-546, 2009.

- BORAH, B. J.; BASU, A. Highlighting Differences Between Conditional and Unconditional Quantile Regression Approaches Through An Application To Assess Medication Adherence. **Health Econ.** v. 22, n.9, p. 1052–1070, 2013.
- CALIENDO, M.; KOPEINIG, S. Some Practical Guidance for the Implementation of Propensity Score Matching. **Journal of Economic Surveys**, v.22 n.1, p. 31-72. 2008
- CASTRO, M.H.G.; TIEZZI, S. A reforma do ensino médio e a implantação do ENEM no Brasil. In: BROCK, C.; SCHWARTZMAN, S. (Org.). **Os desafios da educação no Brasil**. Rio de Janeiro: Nova Fronteira, 2005. p. 119-154. Disponível em: <<http://www.schwartzman.org.br/simon/desafios/4ensinomedio.pdf>> Acesso em 12 junho 2015
- CHEN, X; WEKO, T. US Department of Education NCES 2009. **Aurora**, v. 202, p. 502-7334, 2009.
- CHERNOZHUKOV, V.; FERNANDEZ-VAL, I; MELLY, B. Inference on Counterfactual Distributions. **Econometrica**. v. 81, n.6, p. 2205-2268, 2013.
- COSTA, L.; ARRAES, R. Identificação parcial do efeito das escolas privadas brasileiras. **Pesquisa e Planejamento Econômico (PPE)**. v.44, n.2, agosto de 2014.
- CROUCH, C., D.; SAKO, M. **Are skills the answer?** The political economy of skill creation in advanced industrial countries. New York: Oxford University Press. 1999
- DATTA, N. Evaluating Impacts of Watershed Development Program on Agricultural Productivity, Income, and Livelihood in Bhalki Watershed of Bardhaman District, West Bengal. **World Development**. v. 66, p. 443-456, 2015.
- DINARDO, J; FORTIN, N. M.; LEMIEUX, T. Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach. **Econometrica**, v. 64: p. 1001-1044, 1996.
- DIPRETE, T.; GANGL, M. 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, Abr. 2004. Disponível em: <http://www.wjh.harvard.edu/~winship/cfa_papers/HBprop_021204.pdf>. Acesso em: 12/11/2015.
- DONALD, S. G.; GREEN, D. A.; PAARSCH, H. P. Differences in Wage Distributions between Canada and the United States: An Application of a Flexible Estimator of Distribution Functions in the Presence of Covariates Source. **Review of Economic Studies**, v. 67, p. 609-633, 2000.
- EDOKA; I. P. **Decomposing Differences in Cotinine Distribution between Children and Adolescents from Different Socioeconomic Backgrounds**. HEDG. Working Paper n. 12/29. 2012.
- FARREL, M. Robust Inference on Average Treatment Effects with Possibly More Covariates than Observations. **Journal of Econometrics**, v.189 n.1 p. 1–23, November 2013.
- FISHER, J.; MARCHAND, J. Does the retirement consumption puzzle differ across the distribution? **The Journal of Economic Inequality**, v.12, p. 279-296. 2014.

FIRPO, S.; FORTIN, N.; LEMIEUX, T.. Decomposing wage distributions using recentered influence function regressions. **Mimeo**. University of British Columbia, 2007.

FIRPO, S.; FORTIN, N.; LEMIEUX, T.. Unconditional Quantile Regressions, **Econometrica**, v. 77, n.3, p. 953-973, 2009.

FOLEY, P. **The Socio-Economic Status of Vocational Education and Training Students in Australia**. National Centre for Vocational Education Research Ltd. PO Box 8288, Stational Arcade, Adelaide, SA 5000, Australia, 2007.

FORTIN, N M. The Gender Wage Gap among Young Adults in the United States: The Importance of Money vs. People. **Journal of Human Resources**, v.43, p. 886-920, 2008.

FORTIN, N. M; LEMIEUX, T. Rank Regressions, Wage Distributions, and the Gender Gap. **Journal of Human Resources** v. 33, p. 610.643, 1998.

FOURNIER, J.; KOSKE, I. Less income inequality and more growth—Are they compatible? Part 7. The drivers of labour earnings inequality—An analysis based on conditional and unconditional quantile regressions. **OECD Economics Department Working Papers**, n. 930, OECD Publishing, 2012.

FRANÇA, M. T. A.; GONÇALVES, F.O. Provisão pública e privada de educação fundamental: diferenças de qualidade medidas por meio de propensity score. **Economia Aplicada**. v.14, n.4, p.373-390, 2010.

GANIMIAN, A. J.; ROCHA, A. S. **Measuring Up?. How Did Latin America and the Caribbean Perform on the 2009 Programme for International Student Assessment (PISA)?**. Partnership for Educational Revitalization in the Americas (PREAL), 60 p., 2011

GOLDTHORPE, J.H. **Social mobility and class structure in modern Britain**. New York: Oxford University Press, 1992.

GUIMARÃES, A. Q; ALMEIDA, M. E. Os jovens e o mercado de trabalho: evolução e desafios da política de emprego no Brasil. *Temas de Administração Pública*. v. 8, n. 2, 2013. Disponível em: <http://seer.fclar.unesp.br/temasadm/article/view/6845>. Acesso em: 12 setembro 2016.

HAHN, R.; CARVALHO,C.; PUELZ, D. Bayesian Regularized Regressions For Treatment effects estimation from observational data. Working Paper, **Duke University**, 2015.

HAINMUELLER J. Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. **Political Analysis**, v. 20 n.1, p. 25-46, 2012.

HAINMUELLER J.; XU, Y. Ebalance: A Stata Package for Entropy Balancing. **Journal of Statistical Software**. v. 54, n. 7. August, 2013.

HAMPEL, F. R.; RONCHETTI, E. M.; ROUSSEEUW, P. J.; STAHEL, W. A. **Robust Statistics: The Approach Based on Influence Functions**. Wiley; New York: 1986.

HANUSHEK, E. A.; WOESSMANN, L. Schooling, Educational Achievement, and the Latin American Growth Puzzle. **Journal of Development Economics**, v.99, n.2, p.497-512, 2012.

- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learnings: Data Mining, Inference and Prediction**. Springer Series in Statistics, Second Edition, 2012.
- HECKMAN, J.; LALONDE, R.; SMITH, J. The economics and econometrics of active labor market programs. *In*: ASHENFELTER, O., CARD, D. (Eds.) **The Handbook of Labor Economics**. Amsterdam: North Holland. v.3A, part.6, cap.31. 1999.
- HIRSCH, B.T.; WINTERS, J.V. An Anatomy of Racial and Ethnic Trends in Male Earnings in the U.S. **Review of Income and Wealth** v. 60, p. 930-947, 2014.
- HIRANO, K; IMBENS, G; RIDDER G. Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score. **Econometrica**, v. 71 n.4, p.1161-1189, 2003
- HO, D., IMAI, K; KING, G. Stuart E, A. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. **Political Analysis**, v.15 n.3, p.199. 2007
- IACUS, S. M.; KING, G; PORRO, G. Matching for causal inference without balance checking. <http://gking.harvard.edu/files/cem.pdf>. 2008.
- IMBENS, G. The role of the propensity score in estimating dose–response functions. **Biometrika**. v. 87, n.3, p.706–710, 2000.
- IMBENS, G. Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review. **Review of Economics and Statistics**, v.86 n.1, p.4-29, 2004
- INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS – INEP. Disponível em: <http://www.inep.gov.br/> Acesso em: 10 julho 2015.
- JANNUZZI, P. M. Status socioeconômico das ocupações brasileiras: medidas aproximativas para 1980, 1991 e anos 90. **Revista Brasileira de Estatística**, Rio de Janeiro, v. 61, n. 2, p. 47-74, 2001.
- JANNUZZI, P. M. **Indicadores sociais no Brasil**. Campinas: Alínea, 2003, 2.ed.
- JUHN, C.; MURPHY, K. M.; PIERCE, B. Wage Inequality and the Rise in Returns to Skill. **Journal of Political Economy** v. 101, p. 410-442, 1993.
- KASSENBOEHMER, S.C.; SINNING, M.G. Distributional changes in the gender wage gap. **Industrial & Labor Relations Review** v. 67, p. 335-361, 2014.
- KOENKER, R.; BASSETT, J.R. Regression quantiles. **Econometrica: journal of the Econometric Society**, p. 33-50, 1978.
- KUENZER, A. Exclusão incluyente e inclusão excludente: a nova forma de dualidade estrutural que objetiva as novas relações entre educação e trabalho. *In*: SAVIANI, D.; SANFELICE, J.L.; LOMBARDI, J.C. (Org.). **Capitalismo, trabalho e educação**. 3. ed. Campinas: Autores Associados, 2005. p. 77-96.
- LE, H.T.; BOOTH, A.L. Inequality in Vietnamese Urban–Rural Living Standards, 1993–2006. **Review of Income and Wealth** v. 60, p. 862–886. 2013.
- LEE, W. Propensity Score Matching and Variations on the Balancing Test. **Mimeo**. Melbourne Institute of Applied Economics and Social Research. 2006

LECHENER, M. Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. **Econometric Evaluation of Labour Market Policies**. Heidelberg, p. 1–18. 2001

LEME, A.; ESCARDÍBUL, J-O. **The effect of a specialized versus a general upper secondary school curriculum on students' performance and inequality. A difference-in-differences cross country comparison**. Document de treball de l'ieib, n.16, 2016

LEVY, S., SCHADY, N. Latin America's Social Policy Challenge: Education, Social Insurance, Redistribution. **The Journal of Economic Perspectives**, v.27, n.2, p.193-218, 2013.

LIN, D.; LUTTER, R.; RUHM, C. J. **Cognitive Performance and Labor Market Outcomes**. Discussion Paper, IZA, n. 10075. July, 2016.

MACHADO, J. F.; MATA, J. Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression. **Journal of Applied Econometrics**, v.20, p. 445-465, 2005.

MANSKI, C. Anatomy of the selection problem. **The journal of human resource**, v. 24, p. 343-360, 1989.

MARIANO, F. Z.; ARRAES, R. A; SOUZA, N. O. **Desempenho Escolar e Inserção no Mercado de Trabalho: Uma Avaliação das Escolas Estadual de Ensino Profissionalizantes (EEEP) do Ceará**. In. XI Encontro Economia do Ceará em Debate. IPECE 2015. Disponível em:

<http://www2.ipece.ce.gov.br/encontro/2015/trabalhos/Desempenho20escolar20e20inserC3A7C3A3o20no20mercado20de20trabalho.pdf> Acesso em: 10 Jan. 2016

MARTINS, A. P. Pressupostos de Gramsci na educação profissional e tecnológica de nível médio. **Revista de Educação, Ciência e Tecnologia**, v. 1, n. 2, 2012.

MELLY, B. Decomposition of differences in distribution using quantile regression. **Labour economics** v.12, p. 577-590, 2005.

MOEHLECKE, Sabrina. O Ensino Médio e as novas diretrizes curriculares nacionais: entre recorrências e novas inquietações. **Revista Brasileira de Educação**, v. 17, n. 49, Rio de Janeiro. Jan- abril, 2012. Disponível em: http://www.scielo.br/scielo.php?pid=S1413-24782012000100003&script=sci_arttext Acesso em: 10 julho 2015

MORAES, A. G. E.; BELLUZZO, W. O diferencial de desempenho escolar entre escolas públicas e privadas no Brasil. **Nova Economia**. Belo Horizonte. v.24 n.2 p.409-430 maio-agosto, 2014.

MORIN, L.P. Do Men and Women Respond Differently to Competition? Evidence from a Major Education Reform. **Journal of Labor Economics** v. 33, p. 443-491, 2015

NEUMAN, S; ZIDERMAN, A. Vocational Secondary Schools Can Be More Cost Effective than Academic Schools: The Case of Israel. **Comparative Education**, v. 25, n. 2, p. 151-163, 1989.

NGUYEN; H. T. **The Evolution of The Gender Test Score Gap Through Seventh Grade: New Insights From Australia Using Quantile Regression and Decomposition.** Bankwest Curtin Economics Centre. Working Paper. n. 15/7, 2015.

POLIDANO, C.; TABASSO, D. **Fully Integrating Upper-Secondary Vocational and Academic Courses: A Flexible New Way?** Discussion Paper n. 9694, Jan. 2016

RELATÓRIO DA OECD. “Who are the low-performing students?”, *PISA in Focus*, n. 60, OECD Publishing, Paris. 2016 Disponível em: <http://dx.doi.org/10.1787/5jm3xh670q7g-en>. Acesso em: 10 julho 2016

RELATÓRIO DE GESTÃO “O PENSAR E O FAZER DA EDUCAÇÃO PROFISSIONAL NO CEARÁ – 2008 a 2014”. Secretaria da Educação. Coordenadoria de Educação Profissional. – 1. ed. – Fortaleza: Secretaria da Educação, 2014

RELATÓRIO NACIONAL PISA 2012. Resultados Brasileiros, OCDE 2012. Disponível em: http://download.inep.gov.br/acoes_internacionais/pisa/resultados/2014/relatorio_nacional_pisa_2012_resultados_brasileiros.pdf Acesso em: 22 Fev 2016

ROSENBAUM, P. **Observational Studies**, Springer, New York. 2002.

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

ROSENBAUM, P.; RUBIN, D. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. **Journal of the Royal Statistical Society. Series B**, v.45, 212–218. 1983

RUBIL; I. **The Great Recession and the Public-Private Wage Gap: Distributional Decomposition Evidence from Croatia 2008-2011.** MPRA Paper n. 46798. 2013.

RUBIN, D. B. **Matched Sampling for Causal Effects.** Cambridge University Press. 2006

SCHURER, S. ALSPACH, M. MaCRAE, J. MARTIN, G. L., The Medical Care Costs of Mood Disorders: A Coarsened Exact Matching Approach, **IZA Discussion Papers**, 2015.

SECRETARIA DE EDUCAÇÃO DO CEARÁ - SEDUC. Disponível em: <http://www.seduc.ce.gov.br/index.php/educacao-profissional> Acesso em: 15 de Jun. 2015

SEKHON, J. S. Opiates for the Matches: Matching Methods for Causal Inference. **Annual Review of Political Science**, v. 12, p.487-508, 2009

SEVERNINI, E. R.; ORELLANO, V. I. F. O efeito do ensino profissionalizante sobre a probabilidade de inserção no mercado de trabalho e sobre a renda no período pré-Planfor. **Revista EconomiA**, 2010.

SILVA, V. H. O. Análise da Participação das Escolas Públicas Estaduais Cearenses no Exame Nacional do Ensino Médio (ENEM): 2009-2011. **IPECE Informe**, n. 54. Fortaleza, 2013. Disponível em: < http://www.ipece.ce.gov.br/publicacoes/ipece-informe/Ipece_Informe_54_A26_fevereiro_2013.pdf> Acesso em: 12 junho 2015

SILVA, M. R.; PELISSARI, L. B.; [STEIMBACH, A. A.](#) Juventude, escola e trabalho: permanência e abandono na educação profissional técnica de nível médio. **Educ. Pesqui.** [online]. v.39, n.2, p.403-417. 2013.

SIMÕES, C. A. Políticas públicas do ensino médio: realidade e desafios. In: FERREIRA, Cristina Araripe (Org.). Juventude e iniciação científica: políticas públicas para o ensino médio. Rio de Janeiro: EPSJV; UFRJ, 2010. p. 135-142.

TAVARES, P. A. **Três Ensaio em Economia da Educação**. São Paulo, 2012. Disponível em: <http://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/11542/TrC3AAs20Ensaio%20em20Economia20da20EducaC3A7C3A3o.pdf?sequence=1&isAllowed=y>. Acesso em: 12 junho 2015

TIBSHARANI, R. Regression Shrinkage and Selection via the Lasso. **Journal of the Royal Statistical Society, Series B-Methodological**, v.58, n.1, p.267-288, 1996.

VALLE SILVA, N. **Uma proposta de classificação das ocupações brasileiras**. LNCC, Mimeografado. 1992.

WANG, C., PARMIGIANI, G. DOMINICI, F. Bayesian effect estimation accounting for adjustment uncertainty. **Biometrics** 68 p. 661–671, 2012.

WATSON, S. ELLIOT, M. Entropy Balancing: A maximum-entropy reweighting scheme to adjust for coverage error. **Quality & Quantity**. v. 50, n. 4, p. 1781–1797, July 2016.

ZOU, H., HASTIE, T. Regularization and variable selection via elastic net. **Journal of Royal Statistics Society B**, v. 67 n.2, p. 301-320, 2005.