

**Avaliação do Ensino Profissionalizante: Aplicação dos Métodos Lasso e Pós-Lasso em
Dados Longitudinais**

Ronaldo de Albuquerque e Arraes

Professor UFC/CAEN, PhD

ronald@ufc.br

Francisca Zilania Mariano

Professora UFC/Campus de Sobral

Doutoranda em Economia, UFC/CAEN

zilania@ufc.br

Rafael Barros Barbosa

Professor UFC/DEA, Doutor

rafaelbarrosbarbosa@gmail.com

Av. da Universidade, 2700

60020-181 Fortaleza – CE

Área 3 - Economia do Trabalho, Economia Social e Demografia.

Avaliação do Ensino Profissionalizante: Aplicação dos Métodos Lasso e Pós-Lasso em Dados Longitudinais

Área 3 - Economia do Trabalho, Economia Social e Demografia.

Resumo

Este trabalho procura ampliar o debate existente 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. Para a mensuração do efeito do ensino profissionalizante, 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, condução da avaliação através dos métodos Lasso e pós-Lasso, recentemente desenvolvidos por Belloni et al (2015), no processo de seleção das variáveis explicativas, seguida da aplicação do algoritmo CEM para aplicação do modelo de escore de propensão. Através do pareamento pelo método de *Kernel*, mensurou-se o Efeito Médio do Tratamento (EMT), 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 17,8% a mais na nota para um aluno da EEEP. Ao se isolar o efeito líquido da escola, conclui-se que as EEEP demonstram ser mais competentes em potencializar habilidades do conhecimento cognitivo relativo ao ensino regular, resultado que contribui para direcionar futuras políticas educacionais.

Palavras-Chave: Desempenho no Ensino Profissionalizante, Escola Pública, Dados Longitudinais, Métodos Lasso e Pós-Lasso.

Abstract

This work aims to expand the current literature on public education policy by evaluating the performance of students from upper-secondary state school with vocational education and training (VET) in the national examination (ENEM). Based upon the methodological procedures applied to measure the VET effects, some contributions of the paper shall be considered. First, application longitudinal dataset created from crossing three different databases, namely SPAECE 2011, 2012, School Census 2011 and ENEM 2014. Second, the evaluation was carried out through the application of lasso and post-lasso methods, recently developed by Belloni et al. (2015), in the selection of the explanatory variables, followed by the application of the CEM algorithm in order to perform the propensity scoring model. By matching through Kernel method, estimates of the average treatment effect regarding the grade performance of the students in the VET state schools turned out to be positive and significant in all areas of knowledge of ENEM, especially Writing, whose grade effect was 17, 8% higher. Concerning the school net effect, it is concluded that VET school demonstrate to be more competent in enhancing skills of cognitive knowledge compared to regular education, therefore, this may contribute do guide educational policies in the future.

Keywords: Vocational Education and Training, Public School, Student Performance, Longitudinal Data, Lasso e Post-Lasso Methods

Classificação JEL: C81, I21, I28

1. 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.

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, 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 pública 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 erro padrão melhor 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).

Dentre os tipos de políticas voltados para a melhoria da educação pública no ensino médio encontra-se a concepção do ensino profissionalizante, que 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, porém, Polidano e Tabasso (2016) observaram que na Austrália esse tipo de combinação de ensino reduz as chances de o estudante entrar em uma universidade.

Através da 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 autores buscam verificar os

efeitos desse tipo de ensino sobre a inserção no mercado de trabalho (NEUMAN; ZIDERMAN 1989; CHEN; WEKO, 2009; SEVERNINI; ORELLANO, 2010; ASSUNÇÃO; GONZAGA, 2010; ARAÚJO et al, 2014), todavia, ainda, são poucos os estudos que avaliam os impactos do ensino profissionalizante sobre a educação pública brasileira, destaque 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. Por outro lado, a revisão da literatura aqui feita não constatou algum trabalho em nível estadual com foco de avaliação de política sobre escolas profissionalizantes, lacuna esta a ser preenchida com o estudo de caso para o Ceará.

O Governo Estadual do Ceará iniciou em 2008 um projeto de criação de Escolas Estaduais de Educação Profissional (EEEP), que incluem, 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, pela Secretaria de Educação, atende a realidade local da cidade onde a instituição é implantada (SEDUC, 2015). 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. Silva (2013) observa o desempenho das escolas estaduais no Exame Nacional do Ensino Médio (ENEM), através 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 sobre o papel das escolas profissionais públicas do Ceará em nível educacional.

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

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 a 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 a seleção dessas variáveis, será aplicado o algoritmo CEM desenvolvido por Iacus, King, e Porro (2009), visando melhorar o pareamento ao reduzir os dados para um suporte comum antes da aplicação do modelo de score propensão, que será utilizado para se obter o Efeito Médio de Tratamento (EMT).

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. Revisão da Literatura

2.1 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, o ensino profissionalizante é analisado por diversos autores que 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).

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.

No Brasil, Tavares (2012) afirma que o ensino profissionalizante foi criado para atender crianças, jovens e adultos que viviam à margem da sociedade, onde as primeiras escolas que constituíram a Rede Federal de Educação Profissional tinham a função de instruir tais indivíduos através do ensino de um ofício ou profissão e prepara-los para o mercado de trabalho.

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).

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 score 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.

Segundo Martins (2012), pode-se romper a dualidade estrutural entre formação para o trabalho e preparação para a universidade através da integração das duas redes de ensino, profissional e geral. É, pois, nesse contexto que se enquadram as escolas profissionalizantes de ensino.

3 Metodologia

3.1 Métodos de Avaliação de Impacto

Métodos de avaliação de impacto têm por objetivo verificar se um determinado programa social está alcançando os resultados esperados. Dessa forma, busca mensurar as mudanças no bem-estar dos indivíduos que sofreram a intervenção da política ao compará-los a um contrafactual, o qual capta o comportamento desses indivíduos na ausência de tal intervenção.

O procedimento é feito através da observação de uma amostra de dados para uma determinada unidade i . As unidades i são denominadas tratadas ($T_i = 1$) quando recebem o tratamento, enquanto que as que não recebem são definidas como não-tratadas ($T_i = 0$). O efeito causal do tratamento é encontrado pela diferença entre a variável de resultado quando i é tratado (Y_1) e a variável de resultado quando i não é tratado (Y_0), obtendo assim, o efeito médio do tratamento (EMT).

A principal dificuldade existente em se avaliar uma política pública ou programa social reside em conseguir observar uma mesma unidade em duas circunstâncias distintas, com e sem a política simultaneamente. Uma alternativa proposta para se obter o contrafactual é construir um grupo de controle que seja estatisticamente idênticos ao grupo de tratados de forma a evitar o problema de viés de seleção. Assim, a avaliação de impacto pode ser feita através de dados experimentais e não-experimentais. Segundo Heckman (1989), a forma ideal seria a utilização de dados em

experimento, porém quando se trabalha com ciências sociais nem sempre é possível implementar um experimento aleatório.

Se a definição dos grupos não ocorrer de forma aleatória, então devem existir outras características das unidades que determinem a designação para cada grupo, ou seja, o pressuposto da hipótese de identificação requer que existam unidades 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). Assim, algumas hipóteses são necessárias para garantir a estimação da relação causal se as chances para o tratamento não forem aleatórias, como a Hipótese da Independência Condicional (HIC), o qual assume que, os resultados potenciais são independentes da variável binária de tratamento ao se condicionar às variáveis observáveis, X_i . Além disso, assume-se que fatores não observáveis não são fontes de viés.

Através da diferença entre as médias dos resultados dos escores das unidades do grupo de tratamento e do grupo controle, pode-se estimar o efeito médio dos indivíduos do grupo tratado (EMT). 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,

$$\{Y_{0i}, Y_{1i}\} \perp d_i | X_i \quad (1)$$

A variável binária $d_i = 1$ indica que a unidade i recebeu o tratamento; Y_{1i} e Y_{0i} são os resultados de uma unidade após a adesão ao programa e caso esta não fosse beneficiada pelo programa, respectivamente.

Condicionando às variáveis observáveis, X_i , e assumindo HIC, tem-se que:

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

Em termos gerais, o método consiste em especificar, a priori, um grupo de unidades sujeitas a um determinado tratamento, e compara-las com outras isentas do tratamento (grupo de controle), embora com características semelhantes. Em seguida, estimam-se os efeitos do tratamento por meio da diferença entre os resultados médios dos grupos de tratamento e controle.

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 através 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) afirma 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*, onde este 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 tratamento¹ pode ser encontrada ou pelo pareamento do valor encontrado no primeiro passo ou utilizando algum esquema de pesos (CALIENDO; KOPEINIG, 2005). Essa estimativa tem por objetivo encontrar um grupo de controle que seja o mais semelhante possível ao grupo de tratamento em termos do escore de propensão, dadas às características observadas. Assim, pelo teorema do escore de propensão e supondo a HIC, o EMT, resultante do pareamento direto dos valores de propensão entre tratados e não tratados, aplicando a lei de expectativas iteradas sobre X_i , é dado por:

$$EMT = E\{E[Y_i | P(X_i), D_i = 1] - E[Y_i | P(X_i), D_i = 0] | D_i = 1\} \quad (6)$$

Smith e Todd (2005) mostram que o estimador típico PEP possui a seguinte forma:

$$EMT_{PSM} = \frac{1}{N_T} [\sum_{i \in T} Y_{1,i} - \sum_{j \in C} \omega(i, j) Y_{0,j}] \quad (7)$$

Onde $\omega(i, j)$ é o tipo de pesos utilizado para agregar o resultado potencial dos indivíduos do grupo de controle e depende dos escores de propensão do participante $[P(X_i)]$, e não-participante $[P(X_j)]$, e N_T é o número de indivíduos tratados pertencentes à região de suporte comum.

Os principais métodos de pesos utilizados para a estimação do EMT 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 feito o teste de balanceamento com cada um deles para se selecionar o mais adequado na inferência do EMT.

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.

¹ Esse procedimento se dá dentro da região de suporte comum.

Ravaillon (2008) sugere que elas são 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.

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 existência de grandes bases de dados tem proporcionado um desafio adicional. Tendo em vista o vasto potencial de variáveis de controle, os métodos de estimação tradicionais, 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 e Dominici (2012), Wang, Pagamini e Dominici (2012), Lefebvre et al (2014, 2015), 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 EMT 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 EMT, com ou sem variáveis endógenas para a variável de tratamento.

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

$$E(Y_i|X) = \frac{\exp(g(X))}{1 + \exp(g(X))} + \varepsilon_i \quad (8)$$

As técnicas tradicionais de estimação de (8) 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 variáveis² e que $p \leq \dim(X)$. O método Lasso selecionará s ($s \ll p$) parâmetros após a aplicação do seguinte problema de minimização:

$$\beta_L = \arg \min \hat{Q}(\beta) + \frac{\lambda}{n} \|\beta\|_1 \quad (9)$$

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.

O termo λ , chamado de *tuning 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 (9) pode ser reescrita da seguinte forma:

$$\beta_L = \arg \min \sum_{i=1}^n (g_i - X\beta)^2 \quad (10)$$

$$\text{sujeito a } \sum_{j=1}^p |\beta_j| \leq t \quad (11)$$

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 *tuning 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 *tuning 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: $\lambda = 2 \cdot c\sigma \sqrt{2n \log(2p/\zeta)}$

² Estas podem incluir variáveis não transformadas e variáveis que sofrem alguma transformação, tais como, interação entre as variáveis, variáveis polinomiais, *B-Splines* dentre outras. Ver Chen (2007) e Newey (2008).

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 (9).

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. Assim, supondo que existam duas variáveis que possuem alta correlação com a variável explicativa, então uma delas terá seu parâmetro estimado tendendo a zero. O mesmo ocorre se o poder explicativo de determinada variável for relativamente baixo. Segundo, a forma funcional da equação (9) 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 eleva³.

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 (10) por MQO após a seleção de variáveis. Isto é, inicialmente realiza-se a seleção de variáveis, em que s dentre as p possíveis variáveis são escolhidas ($s \ll p$). Em seguida, a equação (9) 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.

³ 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).

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

Um problema associado aos métodos de Pareamento no Escore de Propensão (PEP) é o seu fraco desempenho em encontrar grupos de comparação equilibrados, haja vista que o alvo principal dos métodos de pareamento consiste em eliminar observações de modo a obter um melhor equilíbrio comparativo entre os grupos tratado e controle (STEFANO et al, 2011). Além disso, Iacus, King e Porro (2011) ponderam que tais métodos fazem parte de um grupo conhecido como *Equal Percent Bias Reducing* (EPBR)⁴, os quais não garantem um nível de equilíbrio mínimo entre os grupos, exceto sob hipóteses muito fortes, ou seja, sem sustentabilidade prática.

Para contornar esses problemas Iacus, King e Porro (2011) desenvolveram o *Coarsened Exact Matching* (CEM), algoritmo não exige nenhuma hipótese sobre o processo de geração de dados, com exceção da ignorabilidade, e assegura que os desequilíbrios entre os grupos de tratado e controle após o pareamento não sejam maiores que um limite previamente selecionado. Os autores mostraram por meio de simulações que CEM possui propriedades estatísticas superiores aos métodos EPBR, como menor dependência do modelo e reduzido erro de estimação baseado em estatísticas como o viés, variância e erro quadrático médio. Além disso, o CEM permite maior robustez de análise de outros métodos, como o próprio PEP, razão pela qual será usado neste trabalho⁵. Todavia, cabe frisar que sua aplicação aqui volta-se especificamente para melhorar o desempenho do PEP ao escolher mais adequadamente, com menos desequilíbrio, a região do suporte comum.

3.4 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).

⁴ Fazem parte deste grupo de métodos: PEP e distância de Mahalanobis.

⁵ Alguns autores têm destacado ter tido sucesso em aplicações recentes com o CEM: Datta (2015), Aroca et al (2014), Schurer et al (2015).

As informações referentes ao desempenho nas cinco áreas de conhecimento do ENEM 2014⁶ 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.

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 ano⁷, 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 pessoas (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)⁸.

4 Resultados

4.1 Análise Descritiva

Estatísticas descritivas das variáveis dependentes e de algumas variáveis de controles⁹ 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.

⁶ Último ano com informações disponíveis sobre o desempenho individual no ENEM.

⁷ 368 alunos repetiram o 9º ano em 2012

⁸ Para mais detalhes ver Apêndice

⁹ Total de 95 variáveis de controle.

Tabela 1 - Estatísticas descritivas das variáveis da amostra – 2014

Variáveis		EEEP			Regular		
		Média	Desvio-padrão	C.V	Média	Desvio-padrão	C.V
Dependentes	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
Controles:							
Caract. de Alunos	Spaace_MT_2011	268,67	42,35	0,16	244,49	41,70	0,17
	Spaace_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
	Raça	0,21	0,41	1,95	0,19	0,39	2,05
Escolaridade Pai	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. Sócioeconôm.	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
	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.

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 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%.

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

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. Essa é uma técnica “ingênua” comumente usada por não especialistas, 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, 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 melhores 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 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%.

4.2 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 1.517 variáveis a partir das interações lineares entre elas e aplicado o processo de seleção. Destas, o método proposto por Belloni et al (2015) selecionou 20 variáveis consideradas importantes na determinação do tratamento, das quais, 12 são em valores originais e 8 são resultados das interações a partir das variáveis `space_pt`, `sexo`, `radio`, `id_energia_gerador`, `id_energia_outros`, `id_esgoto_rede_publica`¹⁰.

4.3 Estimação do pareamento com escore de propensão

Segundo Blackwell et al (2009), o algoritmo CEM pode ser aplicado a outros métodos, visando melhorar o pareamento ao reduzir os dados para um suporte comum antes da aplicação do modelo de escore de propensão. Com isso, após a seleção das variáveis através do método do Lasso, aplicou-se o algoritmo CEM a estas variáveis, cujos resultados encontram-se na Tabela 2. Esse algoritmo permite identificar os alunos nos dois grupos, controle e tratados, que podem ser comparados. Observa-se que apenas 1.382 e 919 estudantes pertencentes a escolas regulares e profissionalizantes, respectivamente, podem ser pareados a partir das variáveis selecionadas pelo Lasso. Dessa forma, a aplicação do PEP se restringirá somente aos alunos contidos no pareamento e identificados pelo algoritmo CEM.

Tabela 2 – Resultados do algoritmo CEM

	Controle	Tratados
Total	19.811	4.635
Pareado	1.382	919
Não-Pareado	18.429	3.716

Fonte: Elaboração Própria

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 controle¹¹. Posteriormente, descartam-se da amostra aqueles alunos com escore de propensão fora do suporte. Em seguida, o modelo escolhido deverá atender as seguintes propriedades: (1) teste da *balancing property*, ou seja, existir um número suficiente de alunos no ensino regular sob o mesmo intervalo de probabilidade de pertencer a profissional, de forma a se promover um pareamento (terem, em média, mesmas covariáveis); (2) testes de robustez do Pseudo-R² e da análise gráfica da densidade.

Os resultados mostram que nove das dezesseis¹² 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 0.00824325, 0.99926647. Os gráficos A7 e A8 no apêndice apresentam as

¹⁰ Por questão de espaço as estimativas foram omitidas, porém, estão disponíveis à solicitação dos leitores.

¹¹ Idem à nota (11).

¹² O modelo excluiu quatro variáveis por problemas de multicolinearidade

densidades de probabilidade estimadas de pertencer ao ensino profissional entre os alunos para toda a amostra e somente os que estão no suporte comum, após a aplicação do algoritmo CEM respectivamente. 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 (EMT) se dará dentro deste intervalo. Porém, outros métodos fazem-se necessário para se estimar o EMT, 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. A Tabela A1 no apêndice mostra o teste da diferença das médias das variáveis no tratamento e no controle antes e após o pareamento. A hipótese nula do teste é que a diferença das médias é zero, ou seja, as médias dos dois grupos são estatisticamente iguais. Antes do pareamento todas as médias foram estatisticamente diferentes, porém depois de pareados, percebe-se que em quase todos os casos tem-se um pareamento com um bom balanceamento das variáveis.

Embora o pareamento pelo *nearest-neighbor* tenha tido mais variáveis insignificantes ao nível de 5%, o *Pseudo - R²* reduziu apenas de 0,517 para 0,332, enquanto que pelo método *radius* (0,01) e *kernel* (0,01) os resultados foram semelhantes, o *Pseudo - R²* reduz de 0,517 para 0,08, demonstrando que as variáveis explicativas perdem poder para explicar as diferenças entre os alunos. Dessa forma, optou-se por escolher o pareamento pelo *kernel* (0,01) para encontrar o EMT, cujos resultados encontram-se na tabela 3.

Tabela 3 - Efeito médio do tratamento nas áreas de conhecimento do ENEM, por *kernel* (0,01)

Variáveis	EEEP	Regular	EMT	Estatística t
Notas_CN	484,53	456,07	28,46	5,69
Notas_CH	545,06	514,49	30,56	5,38
Notas_MT	471,11	438,19	32,92	4,32
Notas_LC	509,29	490,74	18,55	2,78
Notas_Red	519,73	440,82	78,9	5,12
Notas_MG	505,94	468,06	37,88	6,27

Fonte: Elaboração própria

Os diferenciais de notas entre os dois tipos de escola estimados por pareamento com primeiro *kernel*¹³ estão dispostos na Tabela 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 Ciências humanas (CH) e em Ciências da Natureza (CN), aproximadamente, 6% a mais do que um aluno da escola regular pertencente ao grupo de controle. Em Matemática (MT) o diferencial é 32,92 pontos, 7,5% em favor dos alunos de ensino profissional. Menor e Maior efeito foi observado na área de Linguagens e Códigos (LC) e em Redação (Red), de 18,55 pontos (3,7%) e 78,9 pontos (17,8%), respectivamente. Na Média Geral, o impacto das EEEP foi de 37,88 pontos 8%. Em suma, o ensino profissional de nível médio no Ceará demonstra ser mais competente em potencializar habilidades do conhecimento cognitivo comparativamente ao ensino regular.

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

¹³

Os diferenciais estimados com base nos métodos *Radius* e *nearest-neighbor* também foram encontrados e podem ser verificados na tabela A2 no apêndice.

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, 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 através 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, através 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 obtida 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 através das informações desses alunos em 2011 foi possível observar os fatores que possivelmente determinaram a entrada destes nas escolas profissionalizantes em 2012.

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 score de propensão. Esse método selecionou 20 das 1.517 variáveis criadas a partir de interações das 95 variáveis originais.

Após a seleção dessas variáveis, aplicou-se o algoritmo CEM desenvolvido por Iacus, King e Porro (2008), o qual visa melhorar o pareamento ao reduzir os dados para um suporte comum antes da aplicação do modelo de score de propensão, assim, esse algoritmo permitiu identificar os alunos nos dois grupos, controle e tratados, que podem ser comparados. Dos resultados, observou-se que apenas 1.382 e 919 estudantes pertencentes a escolas regulares e profissionalizantes, respectivamente, podem ser pareados a partir das variáveis selecionadas pelo método Lasso. Dessa forma, a aplicação do PEP se restringiu somente aos alunos que pertenceram ao suporte identificado pelo algoritmo CEM.

Para a mensuração do Efeito Médio do Tratamento (EMT) realizou-se o *matching* através dos métodos *nearest-neighbor*, *radius* e *kernel*, e escolheu-se o que se mostrou o mais balanceado. Embora com resultados semelhantes, optou-se por escolher o pareamento por *kernel* (0,01) para determinação do efeito e 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, 17,8%, a mais para um aluno da escola profissional. Em média, um aluno pertencente a escola profissional apresenta desempenho superior em Ciências humanas (CH) e em Ciências da Natureza (CN), aproximadamente, 6% a mais do que um aluno da escola regular e em Matemática (MT) o diferencial é 32,92 pontos, 7,5% em favor dos alunos de ensino profissional. Menor efeito foi observado na área de Linguagens e Códigos (LC) de 18,55 pontos (3,7%). Na Média Geral, o impacto das EEEP foi de 37,88 pontos 8%.

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 EMT 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: SENAI.DN, 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
- 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.
- DEHEJIA, R. and WAHBA, S. Propensity Score-Matching methods for nonexperimental causal studies. **The Review of Economics and Statistics**, v.84, n.1, p. 151-161, 2002.
- 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.; ICHIMURA, H.; SMITH, J.; TODD, P. Characterizing selection bias using experimental data. **Econometrica**, v. 66, n. 5, p. 1017-1098, 1998.
- 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>. 2009.
- IACUS, S; KING, G; PORRO, G. Causal Inference without Balance Checking: Coarsened Exact Matching. **Political Analysis**, 2011.
- INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS – INEP. Disponível em:
<http://www.inep.gov.br/> Acesso em: 10 julho 2015.
- KUENZER, A.Z. Exclusão includente e inclusão excludente: a nova forma de dualidade estrutural que objetiva as novas relações entre educação e trabalho. In: SAVIANI, D.; LEE, W. Propensity Score Matching and Variations on the Balancing Test. **Mimeo**. Melbourne Institute of Applied Economics and Social Research. 2006
- LEFBVRE, G.; ATHERTON, J.; TALBOT D. The effect of the prior distribution in the Bayesian Adjustment for Confounding algorithm. **Computational Statistics & Data Analysis** 70 p. 227–240, 2014.
- LEFBVRE, G., ATHERTON, J. and TALBOT D. The Bayesian causal effect estimation algorithm. **Journal of Causal Inference** 3 p. 207–236, 2015.
- 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%C3%A7%C3%A3o%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.
- NIBOM, A.A.R. O processo de implantação e expansão da rede de Escola Estadual de Educação Profissional no Estado do Ceará (2008-2014): um olhar investigativo. Monografia. MBA em Excelência em Gestão de Projetos e Processos. Centro Estadual de Educação Tecnológica Paula Souza, São Paulo, 2014. **Mimeo.**
- POLIDANO, C.; TABASSO, D. **Fully Integrating Upper-Secondary Vocational and Academic Courses: A Flexible New Way?** Discussion Paper n. 9694, Jan. 2016
- RAVALLION, M. Evaluating Anti-Poverty Programs. In: **Handbook of Development Economics** v. 4 2007
- RELATÓRIO DA OECD (2016), “Who are the low-performing students?”, *PISA in Focus*, n. 60, OECD Publishing, Paris. <http://dx.doi.org/10.1787/5jm3xh670q7g-en>
- 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.; RUBIN, R. The Central Role of the Propensity Score in Observational Studies for Causal Effects. **Biometrika**, v.70 n.1, p. 41-55. 1983
- RUBIN, D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. **Journal of Educational Psychology**, v.66 n.5. 1974
- SANFELICE, J.L.; LOMBARDI, J.C. (Org.). Capitalismo, trabalho e educação. 3. ed. Campinas: Autores Associados, 2005. p. 77-96.
- 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> Acesso em: 12 junho 2015

SMITH, J.; TODD, P. Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators? **Journal of Econometrics**, v.125, p. 305-353. 2005

TAVARES, P. A. **Três Ensaio em Economia da Educação**. São Paulo, 2012. Disponível em: <http://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/11542/Tr%C3%AAs%20Ensaio%20em%20Economia%20da%20Educa%C3%A7%C3%A3o.pdf?sequence=1&isAllowed=y>.

Acesso em: 12 junho 2015

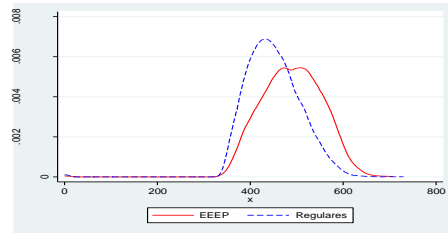
TIBSHARANI, R. Regression Shrinkage and Selection via the Lasso. **Journal of the Royal Statistical Society, Series B-Methodological**, v.58, n.1, p.267-288, 1996.

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.

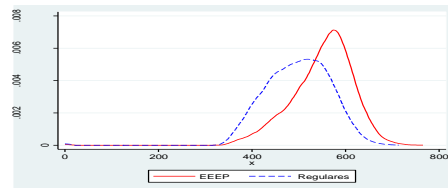
Apêndice

Gráfico A1 - Notas em Ciências Naturais



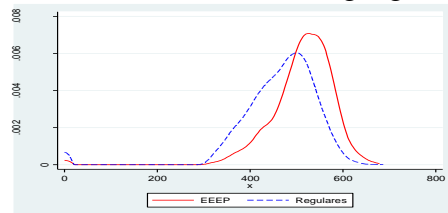
Fonte: Elaboração Própria, ENEM 2014

Gráfico A2 - Notas em Ciências Humanas



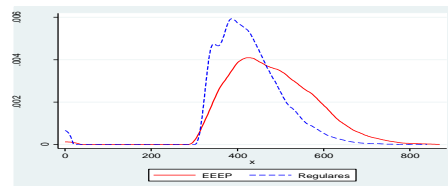
Fonte: Elaboração Própria, ENEM 2014

Gráfico A3 - Nota em Linguagem



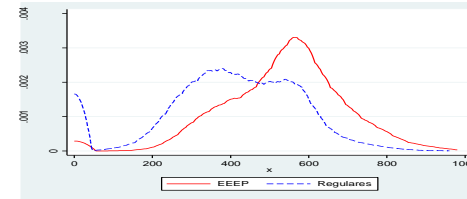
Fonte: Elaboração Própria, ENEM 2014

Gráfico A4 - Notas em Matemática



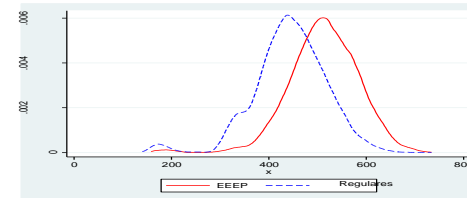
Fonte: Elaboração Própria, ENEM 2014

Gráfico A5 - notas em Redação



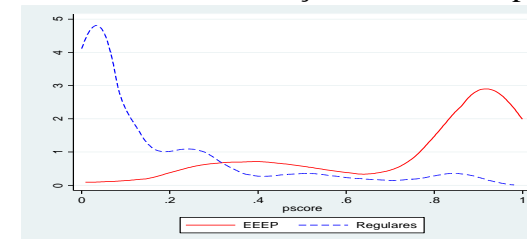
Fonte: Elaboração Própria, ENEM 2014

Gráfico A6 - Notas Média Geral



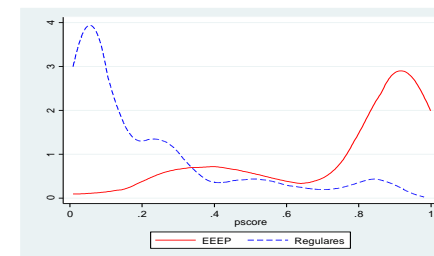
Fonte: Elaboração Própria, ENEM 2014

Gráfico A7 - Distribuição do escore de propensão – toda amostra



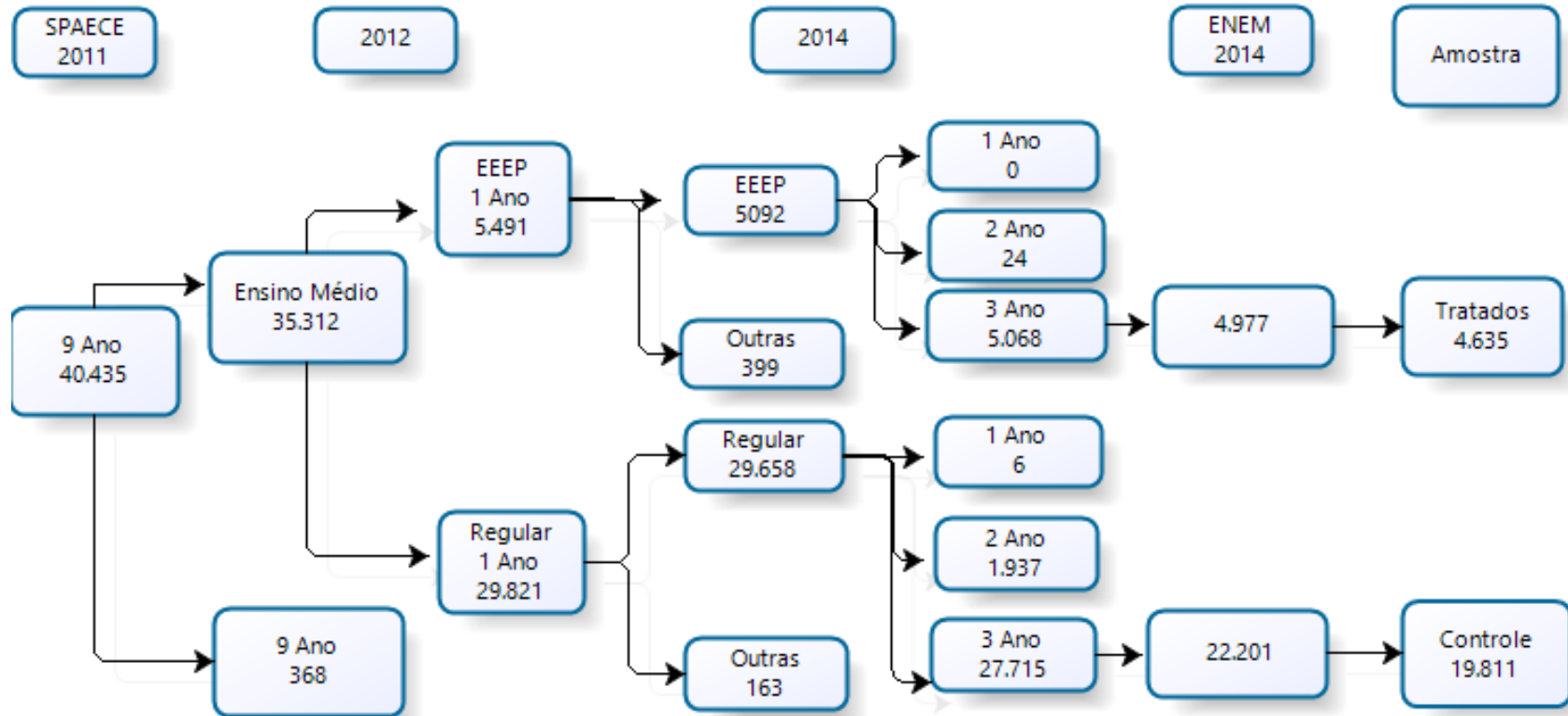
Fonte: Elaboração Própria

Gráfico A8 - Distribuição do escore de propensão –suporte comum



Fonte: Elaboração Própria

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

Tabela A1 - Teste de média dos grupos de EEEP e Regular antes e depois do pareamento por *nearest-neighbor*, *Radius* e *Kernel*

Variáveis	Sem Pareamento			Com Pareamento								
				<i>nearest-neighbor</i>			Radius (0,01)			Kernel (0,01)		
	Media		p-valor	Media		p-valor	Media		p-valor	Media		p-valor
	Tratados	Controle		Tratados	Controle		Tratados	Controle		Tratados	Controle	
V01	261.24	255.17	0.000	264.02	257.54	0.000	262.12	258.29	0.043	262.12	258.25	0.041
V02	255.24	250.15	0.000	257.91	253.27	0.005	256.25	253.96	0.226	256.25	254.05	0.245
V03	0.7736	0.7149	0.002	0.7841	0.7327	0.001	0.76	0.7692	0.696	0.76	0.7687	0.711
V04	0.1381	0.1280	0.483	0.1505	0.1543	0.826	0.1507	0.2071	0.008	0.1507	0.2045	0.011
V05	0.1207	0.1121	0.526	0.133	0.1367	0.856	0.1307	0.1803	0.014	0.1307	0.1752	0.026
V06	0.9825	0.9848	0.679	0.9799	0.9811	0.856	0.9892	0.9941	0.327	0.9892	0.9942	0.322
V07	1	1	-	1	1	-	1	1	-	1	1	-
V08	0.4798	0.6924	0.000	0.5533	0.7302	0.000	0.5784	0.7212	0.000	0.5784	0.7210	0.000
V09	0.7519	0.8321	0.000	0.7678	0.7979	0.145	0.7615	0.8348	0.001	0.7615	0.8357	0.001
V10	0.8161	0.9059	0.000	0.9209	0.9623	0.000	0.9169	0.9556	0.004	0.9169	0.9518	0.011
V11	10.422	10.012	0.000	10.226	9.845	0.000	10.27	10.47	0.010	10.27	10.498	0.004
V12	4114.1	1780.6	0.000	3735.2	2237.8	0.000	3378.5	3471.2	0.251	3378.5	3470.4	0.256
V13	0.9825	0.9799	0.679	0.9799	0.9811	0.856	0.9892	0.9941	0.327	0.9892	0.9942	0.322
V14	1	1	-	1	1	-	1	1	-	1	1	-
V15	0.8161	0.9059	0.000	0.9209	0.9623	0.000	0.9169	0.9556	0.004	0.9169	0.9518	0.011
V16	111.18	102.84	0.000	107.14	99.33	0.000	107.86	111.45	0.023	107.86	111.86	0.011
V17	280000	86565	0.000	240000	120000	0.000	210000	220000	0.520	210000	220000	0.517
V18	0.16104	0.2221	0.000	0.1857	0.1706	0.432	0.1446	0.2328	0.000	0.1446	0.2375	0.000
V19	160000	170000	0.000	170000	170000	0.857	170000	220000	0.000	170000	220000	0.000
V20	960.33	1585.9	0.000	1045.2	1040.4	0.851	1034.8	1034.7	0.997	1034.8	1033.1	0.946
Ps R2	0.517			0.332			0.083			0.084		
LR chi2	1600.38		0,000	733.04		0,000	150.23		0,000	151.38		0,000

Fonte: Elaboração Própria

Tabela A2 - Efeito médio do tratamento nas áreas de conhecimento do ENEM, por *nearest-neighbor* e *Radius* (0,01)

Variáveis	<i>nearest-neighbor</i>				Radius (0,01)			
	EEEP	Regular	EMT	Estatística t	EEEP	Regular	EMT	Estatística t
Notas_CN	486,26	451,55	34,71	11,13	484,53	455,95	28,58	5,79
Notas_CH	546,85	509,77	37,08	11,27	545,06	514,35	30,70	5,48
Notas_MT	473,78	433,87	39,9	9,13	471,11	438,33	32,77	4,36
Notas_LC	511,10	483,00	28,10	8,54	509,29	490,30	18,99	2,89
Notas_Red	525,07	423,33	101,73	12,41	519,73	439,41	80,31	5,29
Notas_MG	508,61	460,30	48,31	14,7	505,94	467,67	38,27	6,42

Fonte: Elaboração própria