

## Avaliação do Ensino Profissionalizante: Aplicação em Dados Longitudinais

Francisca Zilania Mariano<sup>1</sup>  
Ronaldo de Albuquerque e Arraes<sup>2</sup>  
Rafael Barros Barbosa<sup>3</sup>

### 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 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 através dos métodos Lasso e pós-Lasso, recentemente desenvolvidos por Belloni et al (2015), no processo de seleção de variáveis para aplicação do modelo de escore de propensão. Feito o pareamento pelo método de Kernel, estimou-se o efeito médio do tratamento, 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 16,8% a mais na nota para um aluno da EEEP. Ao se isolar o efeito líquido da escola, conclui-se que as EEEP demonstram possuir diretrizes mais eficazes em potencializar habilidades do conhecimento cognitivo relativo ao ensino regular, resultado que contribui para subsidiar futuras decisões dos formuladores de 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 to measure the VET effects, some contributions of the paper shall be considered. First, longitudinal dataset application, 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 for variable selection, recently developed by Belloni et al. (2015), followed by the application of the CEM algorithm in order to isolate the prior better performing students and estimate the propensity scoring model. By matching through Kernel method, estimates of the average treatment effect turned out to be positive and significant in all areas of knowledge of ENEM, especially Writing, whose grade effect was 16, 8% higher. Concerning the school net effect, it is concluded that VET school demonstrate to be more effective in enhancing skills of cognitive knowledge compared to regular education, therefore, this may subsidize decision makers for educational policy guidelines in the near future.

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

Classificação JEL: C81, I21, I28

---

<sup>1</sup> Professora UFC/Campus de Sobral; Doutoranda em Economia, UFC/CAEN; [zilania@ufc.br](mailto:zilania@ufc.br)

<sup>2</sup> Professor UFC/CAEN, PhD; [ronald@ufc.br](mailto:ronald@ufc.br)

<sup>3</sup> Professor UFC/DEA, Doutor; [rafaelbarrosbarbosa@gamil.com](mailto:rafaelbarrosbarbosa@gamil.com)

## 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, 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 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. 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 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; MARIANO et al, 2015), 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 do Ceará sobre o desempenho no ENEM 2013, bem como o impacto desse sobre as chances de os alunos estarem trabalhando. Estes trabalhos utilizam dados em cross-section, os quais não permitem o acompanhamento dos alunos durante o processo de maturação desse tipo de ensino, lacuna esta a ser preenchida com o estudo de caso para o Ceará, através da aplicação em dados longitudinais.

O projeto de criação de Escolas Estaduais de Educação Profissional (EEEP) no Ceará em 2008 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 (SEDUC, 2015). 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. 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 foi possível gerar uma amostra com informações longitudinais através do cruzamento de diferentes bases, a saber, SPAECE 2011, 2012; Censo Escolar 2011 e ENEM 2014, até então desconhecida em outros trabalhos pelos autores deste. 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. E, objetivando isolar o efeito-aluno<sup>4</sup>, 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 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

---

<sup>4</sup> Conhecimento e capacidades adquiridos pelos alunos no período anterior a entrada nas escolas profissionalizantes.

determinação do tratamento. Após esses procedimentos, será encontrado 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. 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 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.

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). 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.

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 Algoritmo CEM**

Dado que existe um processo seletivo para a entrada dos alunos nas EEEP, podem-se existir 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 trabalho<sup>5</sup>.

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,  $\mathcal{L}_1(f^m, g^m) \leq \mathcal{L}_1(f, g)$ .

### 3.2 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).

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).

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.

#### 3.2.1 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

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

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:  $P(X) = P(D = 1 | X)$  (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 tratamento<sup>6</sup> pode ser encontrada ou pelo pareamento do valor encontrado no primeiro passo ou utilizando algum esquema de pesos (CALIENDO; KOPEINIG, 2005). 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. Ravallion (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

<sup>6</sup> Esse procedimento se dá dentro da região de suporte comum.



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.

### 3.3 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, 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 agrupado 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.

#### 3.3.1 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:

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

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 variáveis<sup>7</sup> 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.

O termo  $\lambda$ , 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 (7) pode ser reescrita 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 *tuning point*. A escolha do  $\lambda$  é elemento crucial no processo de seleção das variáveis. Tibshirani (1996) propôs a escolha de  $\lambda$  via correlação cruzada, ou seja, o método é aplicado considerando um intervalo para o valor do  $\lambda$ , 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:  $\lambda = 2 \cdot c\sigma\sqrt{2n\log(2p/\zeta)}$ . Em que,  $c > 1$ ,  $1 - \zeta$  refere-se ao intervalo de confiança para que a probabilidade de rejeição ficar próximo de 1,  $\sigma$  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  $\sigma$  não é observável. Ao contrário, Belloni et al (2010) propuseram um  $\lambda$  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

<sup>7</sup> 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).

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 eleva<sup>8</sup>. 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 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.

### 3.4 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:

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

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  $\gamma$  representa o efeito de  $u_i$  sobre a decisão de participação no programa. Se não existir viés de seleção, então  $\gamma$  será igual a zero e a probabilidade de participação será exclusivamente determinada pelas características observáveis. Entretanto,

---

<sup>8</sup> 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:  $\epsilon$  consiste no erro de aproximar  $y_0$  por  $\hat{f}(x_0)$ . Para maiores detalhes ver, Hastie *et al* (2009).

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  $\gamma$  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 2014<sup>9</sup> 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

<sup>9</sup> Último ano com informações disponíveis sobre o desempenho individual no ENEM.

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 ano<sup>10</sup>, 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)<sup>11</sup>.

## 4 Resultados

### 4.1 Análise Descritiva

Estatísticas descritivas das variáveis dependentes e de algumas variáveis de controles<sup>12</sup> 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.

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%.

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.

<sup>10</sup> 368 alunos repetiram o 9º ano em 2012

<sup>11</sup> Para mais detalhes ver Apêndice

<sup>12</sup> Total de 95 variáveis de controle.

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%.

Tabela 1 - Estatísticas descritivas das variáveis da amostra – 2014

Variáveis		EEEP			Regular		
		Média	Desvio-padrão	C.V	Média	Desvio-padrão	C.V
<b>Dependentes</b>	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
<b>Controles:</b>							
<b>Caract. de Alunos</b>	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
<b>Escolaridade Pai</b>	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
<b>Escolaridade Mãe</b>	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
<b>Cond. Sócioeconôm.</b>	Bolsa_familia	0,72	0,44	0,61	0,80	0,39	0,49
<b>Caract. Escolas</b>	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
	Lixo_coleta	0,99	0,06	0,06	0,99	0,08	0,08
<b>Caract. Escolas</b>	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
	Nº 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 prejudicar 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 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 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 A1 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 0.00249618, 0.9999981. 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. Os gráficos A1 a A3 no apêndice mostra o 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 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	487,55	467,33	20,22	5,75
Notas_CH	548,79	529,16	19,63	4,84
Notas_MT	478,37	452,40	25,96	4,81
Notas_LC	512,47	491,40	21,07	4,19
Notas_Red	531,39	454,60	76,79	7.25
Notas_MG	511,71	478,98	32.73	7.53

Fonte: Elaboração própria

Os diferenciais de notas entre os dois tipos de escola estimados por pareamento com primeiro *kernel* 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 Linguagens e Códigos (LC) e em Ciências da Natureza (CN), aproximadamente, 4% a mais do que um aluno da escola regular pertencente ao grupo de controle. Em Matemática (MT) o diferencial é 25,96 pontos, 5,7% em favor dos alunos de ensino profissional. Menor e Maior efeito foi observado na área de Ciências humanas (CH) e em Redação (Red), de 19,63 pontos (3,7%) e 78,9 pontos (16,8%), respectivamente. Na Média Geral, o impacto das EEEP foi de 32,73 pontos (aproximadamente, 7%). 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 A2 no apêndice mostram que a robustez ao viés de seleção varia entre as variáveis de resultado. Em geral, o efeito do tratamento apresenta-se robusto a uma possível presença de viés de seleção, principalmente em Ciências Humanas (CH), Linguagens e Códigos (LC), Redação (R) e na Média Geral (MG). Para a variável Ciências da Natureza (CN), o nível crítico de  $\Gamma$  é de 1,9, 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 um fator de 1,9, então pode-se questionar o efeito positivo das escolas profissionalizantes em melhorar o desempenho dos alunos nessa área. Além disso, observa-se que Matemática se apresenta a menos robusta à presença de não-observáveis, porém ainda obteve um nível crítico distante da unidade ( $\Gamma = 1,6$ ).

### 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 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 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 através das 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 prejudicar 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 score de propensão. Esse método selecionou 39 das 449 variáveis criadas a partir das 95 variáveis originais.

Para a mensuração do Efeito Médio do Tratamento (EMT) 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, 16,8%, 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 da Natureza (CN), aproximadamente, 4% a mais do que um aluno da escola regular e em Matemática (MT) o diferencial é 25,96 pontos, 5,7% em favor dos alunos de ensino profissional. Menor efeito foi observado na área de Ciências humanas (CH) de 19,63 pontos (3,7%). Na Média Geral, o impacto das EEEP foi de 32,73 pontos 7%. 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 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](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](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**, 10<sup>th</sup> 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.
- 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.

- 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
- 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](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 (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](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
- 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_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

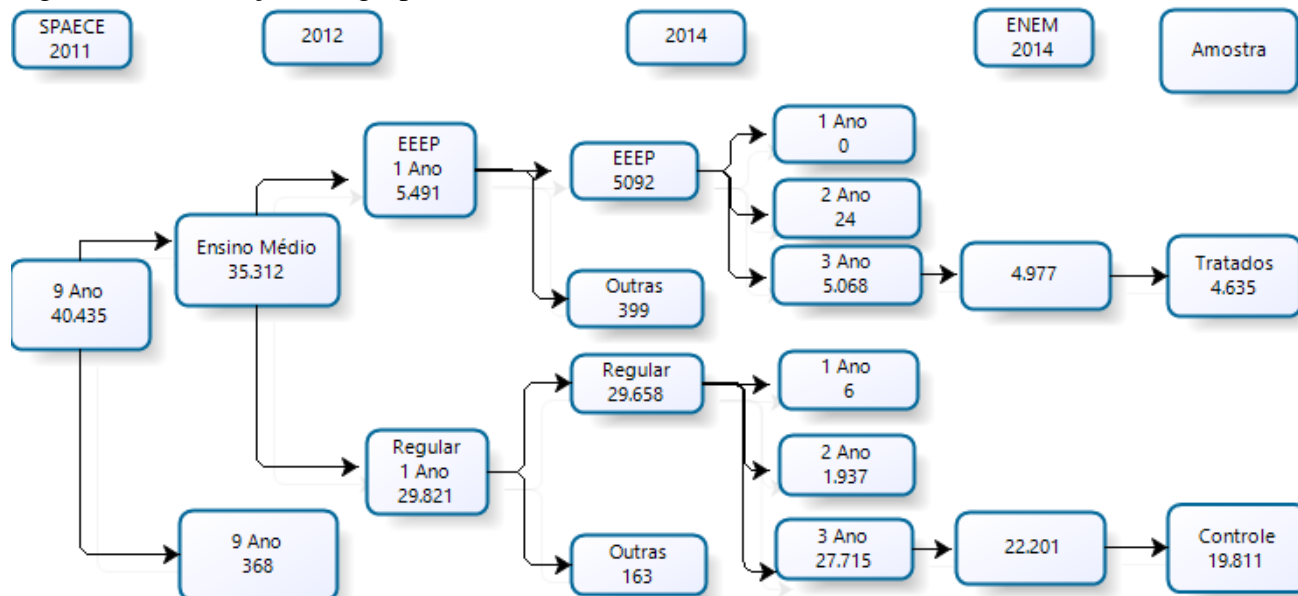
TIBSHARANI, R. Regression Shrinkage and Selection via the Lasso. **Journal of the Royal Statistical Society, Series B-Methodological**, v.58, n.1, p.267-288, 1996.

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

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 A1 – Sobreposição do propensity Score pelo vizinho mais próximo

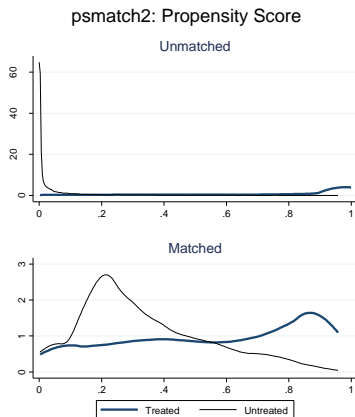


Gráfico A2 – Sobreposição do propensity Score pelo kernel

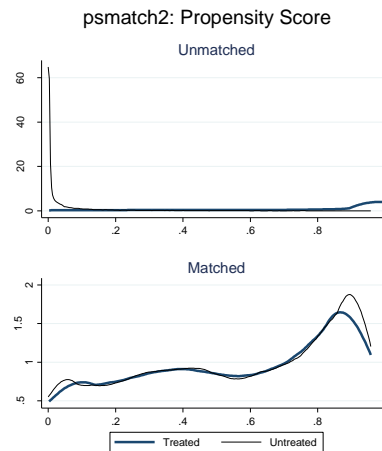


Gráfico A3 - Sobreposição do propensity Score pelo Radius

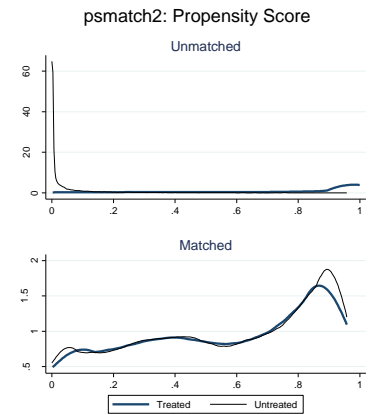


Tabela A1 – Estimação do *Propensity Score* pós Lasso

Variáveis	Coefficientes	p-valor	Variáveis	Coefficientes	p-valor
V01	-1.8178	0.000	V21	-0.4902	0.000
V02	-0.6771	0.000	V22	-1.4019	0.000
V03	-0.5597	0.000	V23	-1.0987	0.000
V04	-1.3684	0.000	V24	1.6349	0.000
V05	-0.2696	0.000	V25	-1.5638	0.000
V06	0.0025	0.000	V26	-1.6719	0.000
V07	-0.0001	0.000	V27	3.6820	0.401
V08	-5.49e-06	0.000	V28	0.0480	0.825
V09	-0.0010	0.000	V29	0.5221	0.000
V10	0.8079	0.000	V30	0.0001	0.000
V11	-0.7612	0.000	V31	-0.0001	0.000
V12	-4.2957	0.000	V32	0.0020	0.168
V13	0.5127	0.000	V33	-0.0017	0.302
V14	0.1093	0.130	V34	-0.7750	0.000
V15	0.1231	0.268	V35	-0.3356	0.000
V16	0.1813	0.147	V36	-0.2806	0.047
V17	0.1101	0.180	V37	0.6220	0.000
V18	0.1147	0.159	V38	-0.2412	0.001
V19	2.0626	0.163	V39	0.0014	0.000
V20	1.9734	0.000	constante	-5.6691	0.000

Fonte: Elaboração Própria

Tabela A2 – Análise de sensibilidade (Rosenbaum Bounds) para as áreas de conhecimento e a média geral do ENEM

$\Gamma$	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	0.003	0.000	0.000	0.000
1.5	0.000	0.000	0.020	0.000	0.000	0.000
1.55	0.000	0.000	0.082	0.000	0.000	0.000
1.6	0.000	0.000	0.223	0.000	0.000	0.000
1.65	0.000	0.000	0.441	0.000	0.000	0.000
1.7	0.000	0.000	0.671	0.000	0.000	0.000
1.75	0.000	0.005	0.846	0.000	0.000	0.000
1.8	0.000	0.021	0.942	0.000	0.000	0.000
1.85	0.004	0.067	0.983	0.000	0.000	0.000
1.9	0.018	0.163	0.996	0.000	0.000	0.000
1.95	0.054	0.316	0.999	0.000	0.000	0.000
2	0.132	0.505	0.999	0.003	0.000	0.000

Fonte: Elaboração Própria