

Identificação Parcial do Efeito das Escolas Privadas Brasileiras

Leandro Oliveira Costa (IPECE –CAEN/UFC)

Ronaldo de Albuquerque e Arraes (CAEN/UFC)

Marcos Vera-Hernandez (*University College London*)

Resumo

Este artigo tem o objetivo de analisar a diferença no desempenho dos alunos de escolas brasileiras públicas e privadas, a partir da análise de Identificação Parcial ou Limites introduzida por Manski (1989). No contexto do forte viés de seleção decorrente de fatores socioeconômicos, a comparação com as estimativas da metodologia de Pareamento no Escore de Propensão e a tradicional análise de regressão busca verificar se os limites estimados sob suposições menos restritivas permanecem informativos. Utilizando as informações mais recente de microdados do Sistema de Avaliação da Educação Básica (SAEB) de 2005 para os estudantes do 5º ano do Ensino Fundamental, os resultados apontam a sobreestimação do efeito das escolas privadas com as metodologias que se baseiam nas suposições de ignorabilidade e imputação, mas o efeito ainda permanece significativamente positivo. A possível explicação desse viés pode ser atribuído ao forte viés de seleção relacionado ao baixo nível de renda dos pais dos alunos de escolas públicas, que limita a escolha da escola do filho e impossibilita a determinação de um aceitável contrafactual.

Palavras-chave: Identificação Parcial, Educação Privada, Viés de Seleção.

Abstract

This article aims to compare student performance in Brazilian public and private schools, from an analysis of Partial Identification or Bounds introduced by Manski (1989). In the context of strong selection bias resulting from socioeconomic factors, it is compared the estimates from the Propensity Score Matching with the ones from regression analysis, to verify whether or not the estimated limits under less restrictive assumptions are informative. Using most recent data from the Basic Education Assessment System (SAEB) of 2005 for students in the 5th year of elementary school, the results indicate a significant bias on the estimated effects by the methodologies that are rely on assumptions of ignorability and imputation. A possible explanation to this bias can be attributed to the strong selection bias related to low income level of parents of students in public schools, making it impossible to determine an acceptable counterfactual.

Keywords: Partial Identification, Private Education, Selection Bias.

1. Introdução

Tem-se constatado, nos últimos anos, significativas diferenças nos resultados das avaliações a favor dos estudantes brasileiros de escolas privadas em relação aos de escolas públicas baseadas no Sistema de Avaliação da Educação Básica (SAEB) e no *Programme for International Student Assessment* (PISA). Isso pode estar refletindo, tanto a fuga da classe média do ensino público, que busca nas escolas particulares um sistema de ensino mais eficiente, como o maior acesso à educação pública das famílias mais carentes¹. Entretanto, é difícil separar nesta diferença o efeito da eficiência do sistema privado do efeito das condições socioeconômicas das famílias que podem optar por escolas privadas para seus filhos; principalmente, considerando que a grande maioria das famílias brasileiras apresenta perfis socioeconômicos desfavoráveis e só tem a escola pública como única opção². Esse imensurável viés de seleção relacionado às condições socioeconômicas das famílias pode tornar incomparável o desempenho dessas duas redes de ensino, impossibilitando isolar o efeito da eficiência das escolas privadas em obter melhores performances dos alunos.

Na busca de determinar o efeito médio da frequência regular na escola e o efeito que ela incide no desempenho dos alunos, contornado o viés de seleção e, conseqüentemente, a endogeneidade intrínseca ao processo educacional, as suposições que possibilitam determinar um contrafactual são facilmente questionadas. No Brasil, pressupõe-se que os rendimentos familiares, a capacidade cognitiva dos alunos e a motivação das famílias relacionada a importância dada à educação como fator de mobilidade social têm um papel significativo no processo de seleção da escola, conseqüentemente, torna-se ineficaz sustentar os pressupostos de homogeneidade e ignorabilidade do tratamento. Como reconhece Webbink (2005), devido o processo educacional ser a princípio endógeno, inúmeros resultados encontrados na literatura da função de produção educacional sobre o efeito dos fatores escolares podem não ser corretos, especialmente quando se preocupa busca comparar estudantes dos sistemas público e privado de ensino (MORGAN e WINSHIP, 2007). Daí a necessidade de se introduzir correções por variáveis instrumentais para corrigir tal problema.

Na educação americana, alguns estudos utilizaram variáveis instrumentais como fonte exógena para identificar o efeito do ensino privado no desempenho do estudante. Por exemplo, Evans e Schwab (1995), primeiros a abordarem o problema do viés de seleção na análise do hiato entre escolas públicas e privadas americanas, utilizaram a vertente religiosa da família católica como variável instrumental. Similarmente, Hoxby (1994) analisa as questões da competitividade e da diminuição da demanda devido à baixa qualidade do ensino público, utilizando a composição religiosa das escolas como instrumento. Entretanto, Altonji et al (2002) criticam esses instrumentos utilizados para identificar o efeito das escolas católicas americanas, pois esses instrumentos normalmente apresentam alguma correlação com os resultados ou com as variáveis explicativas dos desempenhos dos alunos. A ocorrência disso acarretaria sérios danos às inferências realizadas.

¹ Hanushek e Woessmann (2008) analisam o papel das habilidades cognitivas, avaliadas a partir dos exames de proficiência, sobre o desenvolvimento econômico dos países, com foco especial sobre o papel da qualidade e da quantidade de escolas. Seus resultados apontam que o baixo nível de habilidades cognitivas alcançado pelos países latino-americanos podem responder pelo seu fraco desempenho no crescimento desde 1960 e que o desempenho nos exames de proficiência podem explicar as diferenças de crescimento inter e intra-regional.

² Na última década essa diferença nos resultados se agravou, pois com a universalização do Ensino Fundamental, incentivado por programas de transferência de renda como PETI e Bolsa Família, pode ter permitido acesso às escolas públicas de crianças de famílias com baixos níveis socioeconômicos.

Utilizando diferentes metodologias para contornar o viés de seleção devido a variáveis não observadas, vários estudos analisaram a educação brasileira no sentido de comparar o desempenho dos estudantes dos sistemas público e privado. Admitindo a comparabilidade entre os estudantes destes diferentes sistemas de ensino, seus resultados apontam a gestão privada como a mais eficiente (VANDENBERGHE e ROBIN, 2004, SOMERS et al, 2004, FRANÇA e GONÇALVES, 2010, DRONKERS e AVRAM, 2010).

Em termos gerais, a literatura nacional e internacional identifica como os principais problemas de análise a endogeneidade decorrente de variáveis não observadas e o viés de seleção, razão pela qual se tem buscado abordagens metodológicas alternativas, tais como, variáveis instrumentais, pareamento no escore de propensão e diferença-em-diferença. Dronkers e Avram (2010) afirmam que, apesar da expressiva quantidade de trabalhos realizados em nível mundial, os resultados permanecem inconsistentes e inconclusivos. O presente estudo pretende contribuir com o debate, distinguindo-se por sua abordagem e conclusões diferenciadas relativos a trabalhos aplicados no Brasil.

Nesse contexto, pretende-se questionar as suposições de identificação da metodologia de Pareamento no Escore de Propensão e de regressões lineares na investigação sobre a diferença de resultados entre as escolas públicas e privadas brasileiras. Consequentemente, a suposição de ignorabilidade do tratamento e imputação dessas metodologias é questionada para a realidade do mecanismo de seleção das escolas privadas brasileiras. Para isso, será aplicada a metodologia de Identificação Parcial, desenvolvida em Manski (1989, 1998, 2001, 2007), com o objetivo de estimar limites para as estimativas do efeito médio do aluno estar matriculado em uma escola privada sob suposições menos restritivas e mais críveis, embora nem sempre muito informativas.

Em vista destes questionamentos, o objetivo desta pesquisa centra-se em aplicar diferentes ferramentas no conjunto de dados disponíveis que melhor se ajustam à educação brasileira para testar se a escola particular é mais eficiente ou se este é um resultado do viés de seleção devido à significativa estratificação do sistema educacional. Com isso, evidencia-se a complexidade dos processos de escolhas das escolas pelos pais e os fatores que influenciam a eficiência escolar presente no sistema educacional brasileiro.

O trabalho se divide em quatro seções além dessa introdução. Na seção dois faz-se uma revisão da literatura. Na seção seguinte apresenta-se uma análise descritiva da base de dados. Na seção quatro discute-se a metodologia de Identificação Parcial. Os resultados encontram-se na seção cinco, seguido pelas considerações finais.

2. Revisão Bibliográfica

Esta seção visa esclarecer as abordagens feitas sobre o assunto através dos trabalhos mais relevantes da literatura que tratam do efeito do sistema privado de ensino sobre o desempenho dos estudantes. Com isso, tenta-se apresentar as estratégias de identificação desse efeito, listando os principais problemas focados pelos autores em suas bases de dados e nos principais resultados encontrados³.

³ Com o objetivo de focar nos avanços metodológicos, não serão apresentados os clássicos artigos que iniciaram com esta pesquisa, como Friedman (1955) e Coleman et al (1982). Friedman iniciou a discussão sobre a competitividade entre as escolas para melhorar a qualidade da educação global, considerando que o sistema escolar privado é mais eficiente do que o público. O mundialmente conhecido Relatório de Coleman aponta que a qualidade do ensino nas escolas privadas católicas, mesmo controlando por diferenças nas características familiares, teriam maior efeito sobre o desempenho dos estudantes de menor nível socioeconômico e das minorias.

Inicia-se com Webbink (2005) ao resumir os recentes desafios enfrentados e algumas metodologias adequadas para determinar efeitos de causalidade em educação. Segundo, apresentamos Evans e Schwab (1995) um dos primeiros a se preocupar com o viés de seleção na análise do efeito das escolas privadas nos EUA. Hoxby (1994) discute as questões da competitividade e da diminuição da demanda devido à baixa qualidade do ensino público. O quarto artigo da sequência é devido a Vandenberghe e Robin (2004) que, em uma análise entre países, inclusive o Brasil, comparam os resultados de diferentes metodologias que visam contornar a possibilidade de variáveis não observadas. Altonji et al (2002) fazem uma crítica sobre os instrumentos utilizados para identificar o efeito das escolas católicas americanas. Somers, Mcewan e Willms (2004), em seguida, analisam a efetividade das escolas privadas em países da América Latina, e enfatizam os contextos institucionais e níveis de renda os quais bem diferentes dos estudos sobre as escolas americanas. Nguyen, Taylor e Bradley (2006), utilizam a metodologia combinada de pareamento no escore de propensão e diferença-em-diferença, tentam controlar o viés de seleção devido às características observadas e não observadas. França e Gonçalves (2010), aplicando a metodologia de pareamento no escore de propensão nos dados do SAEB, encontram um efeito significativo das escolas privadas brasileiras. Por fim, Dronkers e Avram (2010) tentam separar os processos de escolha da escola e a eficácia escolar a partir de uma metodologia de pareamento em dois estágios.

Webbink (2005) apresenta uma síntese da recente literatura sobre efeitos causais das políticas educacionais no desempenho dos estudantes, e considera que o processo educacional é endógeno. Esse debate lança dúvida sobre inúmeros resultados encontrados na literatura da função de produção educacional. Ele afirma que o principal problema das pesquisas educacionais é que normalmente existem inúmeros fatores que podem mascarar o efeito da intervenção estudada. Ou seja, é difícil produzir resultados convincentes quando não se procura isolar os resultados de uma intervenção do efeito de outros fatores não observados pelo pesquisador. Para isso, novas metodologias têm sido aperfeiçoadas através da utilização de variações exógenas produzidas por experimentos controlados ou naturais, os quais são a forma mais aceitável de se encontrar variações exógenas, embora sua implementação exija custos muito elevados. Já os experimentos naturais tentam, a partir de variações da intervenção, como mudanças de leis e variações naturais, utilizar variáveis instrumentais para produzir ambientes comparáveis aos experimentos. Entretanto, não há nenhum relato da metodologia de pareamento no trabalho deste autor.

Hoxby (1994) investiga se o aumento nas possibilidades de escolha das escolas americanas, aumentando a competição, tanto entre as públicas e privadas como entre as privadas, melhora a qualidade da educação, onde utiliza a composição religiosa das escolas privadas como variável instrumental, a qual representa variações exógenas sobre custos e disponibilidade. Como a baixa qualidade das escolas públicas tende a elevar a demanda pelo ensino privado, a autora enfatiza que não se pode simplesmente comparar resultados dos estudantes de escolas públicas em áreas com ou sem uma alternativa das escolas privadas, pois isso criaria um viés sobre o efeito da competitividade em virtude do maior número de matrículas em escolas privadas. Os resultados atestam haver uma relação direta entre competitividade e qualidade das escolas públicas, mensurado através do desempenho educacional, salários e taxas de conclusão do ensino médio dos estudantes de escolas públicas.

Evans e Schwab (1995), aperfeiçoam a metodologia de Coleman et al (1987) e procuram investigar o efeito sobre a probabilidade de um aluno terminar o ensino médio e entrar na universidade através da diferença de qualidade de ensino entre as escolas públicas e

católicas de ensino médio dos Estados Unidos⁴. Os autores enfatizam o problema gerado pelo viés de seleção, verificando que os estudantes com melhores habilidades cognitivas ou de famílias que atribuem maior valor a educação tendem a ser atraídos para o ensino católico, o que superestimaria o efeito da educação católica. Portanto, um modelo composto de uma única equação não seria adequado, pois necessitaria levar em conta o problema da endogeneidade, fazendo requerer a especificação de mais uma equação. Com isso, estimaram um modelo biprobit com a variável instrumental se a família era católica como fonte exógena de variação para identificar o efeito das escolas católicas. Como principal conclusão, estudantes do segundo grau de escolas católicas aumentam em 13% a probabilidade de ingressarem na universidade.

Morgan (2001) analisa o efeito das escolas católicas nos Estados Unidos pela introdução da metodologia de Pareamento no Escore de Propensão como uma estratégia alternativa aos modelos de regressão no foco das relações de causalidade nos estudos dos efeitos das escolas. Reconhecendo a falta de robustez nas suposições da metodologia de pareamento, seus resultados revelam que o efeito da escola católica americana é mais forte sobre os que são menos prováveis a frequentar uma escola católica, dado suas características observáveis. Ele reconhece que, embora o viés de variável omitida seja levado em conta na análise, raramente se discute sobre a fragilidade dos dados observacionais, principalmente quando o latente viés de seleção é mais significativo. Por esta razão, o autor coloca a metodologia de Pareamento no Escore de Propensão como uma alternativa para plausíveis estimativas do efeito das escolas católicas no contexto de causalidade e avaliação de efeitos do tratamento.

Vandenberghe e Robin (2004) empregam dados do PISA de 2000 para estimar o efeito da educação privada e pública sobre o desempenho dos alunos nos países da Organização para Cooperação e Desenvolvimento Econômico (OCDE). Considerando o potencial viés gerado pela existência de fatores não observados, e relatando importantes divergências entre os resultados dos testes paramétricos e não-paramétricos, utilizam três métodos: variáveis instrumentais (VI), a abordagem de Heckman de dois estágios e o Pareamento no Escore de Propensão. O estimador paramétrico de seleção de Heckman é um método utilizado para controlar o viés de seleção devido à presença de variáveis não observadas. Contudo, como a abordagem emprega VI, faz-se necessário incluir uma destas variáveis como fonte de identificação. A localidade foi usada para saber se a escola encontra-se em cidades grandes, isto é, com mais de 100 mil habitantes⁵. Os resultados revelam que a educação privada não gera benefícios sistemáticos para a maioria dos países. Embora não seja constatado para o Brasil, os autores sugerem interpretar o resultado a partir de características culturais e não de eficiência das escolas privadas.

Altonji, Elder e Taber (2002), a partir de uma análise de sensibilidade, analisam a validade dos instrumentos amplamente utilizados como fonte de variação para identificar o efeito da escola católica americana sobre vários resultados. Investigam se o efeito positivo sobre os resultados dos alunos se relaciona com a questão de uma seleção não aleatória dos alunos pelas instituições de ensino católicas, o que implicaria em correlações espúrias entre a seleção da escola e características não observáveis das famílias que são favoráveis à educação. Ou seja, eles buscaram mostrar que as tentativas para corrigir o viés de seleção

⁴ Eles citam que os resultados de Coleman et al (1987) foram questionados por vários outros autores, os quais apontam esses resultados como sensíveis a escolha de outras variáveis independentes, logo há significativos efeitos decorrentes de um viés de seleção.

⁵ Na escolha desta variável como instrumento, os autores se baseiam no argumento de Hoxby (2000a, b) que advoga que características geográfica e topográficas podem ser usadas como instrumentos naturais para as características endógenas do sistema escolar, pois se acredita que a variação na oferta de escolas privadas entre as grandes cidades e outras áreas reflete principalmente fatores histórico que podem ser equiparados as condições de oferta acidentais.

através da metodologia de VI não é factível na aplicação empírica, dada a inexistência de um instrumento que seja positivamente correlacionado com a escolha da escola, mas também seja exógena no que diz respeito aos resultados educacionais. Os instrumentos analisados foram a filiação religiosa, a proximidade da residência do aluno a escola católica e a interação entre esses dois instrumentos. Usando diversos métodos para testar o viés induzido por estes instrumentos, eles concluem que nenhum é útil para identificar o efeito de uma escola católica e demonstram que o efeito estimado das escolas católicas sobre a conclusão do ensino médio e o ingresso na faculdade é viesado positivamente quando utilizada a abordagem de variáveis instrumentais.

Somers, Mcewan e Willms (2004) comparam a efetividade das escolas privadas em países da América Latina, os quais podem apresentar contextos institucionais e níveis de renda bem diferentes dos estudos que avaliam o efeito das escolas católicas dos Estados Unidos. Utilizando a metodologia de regressão de multi-níveis em uma base de dados organizada pela UNESCO, o artigo aborda a caracterização do efeito dos pares que pode viesar o real impacto da eficiência das escolas privadas. Com mesma metodologia e base de dados, os efeitos observados para o Brasil, comparado com outros países, foram um dos maiores.

Nguyen, Taylor e Bradley (2006), explorando dados longitudinais, deram continuidade à pesquisa sobre o efeito das escolas católicas americanas. Com isso, aplicaram a metodologia de pareamento no escore de propensão em combinação com a metodologia de diferença-em-diferença para controlar o viés de seleção devido a características observadas e não observadas constante no tempo. Os resultados apontam para um efeito menor da escola católica americana, comparado ao obtido por Atonji, Elder e Taber (2002).

França e Gonçalves (2010), a partir do SAEB 2003, analisam a diferença entre os resultados do sistema da educação pública e privada brasileira utilizando a metodologia do Pareamento no Escore de Propensão com o intuito de superar o viés de seleção proveniente de características não observadas relacionadas com o nível socioeconômico das famílias. Os resultados para os alunos do 5º ano de ensino fundamental mostraram uma superioridade do efeito da escola privada sobre a pública, bem como uma diferença crescente quanto ao nível socioeconômico. Para avaliar a satisfação da hipótese de balanceamento no escore de propensão foi utilizado o artifício de se criar variáveis que tentam mensurar o nível socioeconômico dos alunos⁶. Vale apontar, todavia, que essa estratégia possivelmente mascara a incomparabilidade dos estudantes que frequentam escolas públicas e privadas, pois a estimação do escore de propensão é derivada de medidas incorretas sobre os determinantes da matrícula em uma escola privada.

Dronkers e Avram (2010), com o objetivo de distinguir entre os processos de escolha da escola e sua eficácia, fazem uma análise comparativa entre países utilizando o PISA 2006 e a metodologia de Pareamento no Escore de Propensão sobre diferença da efetividade escolar entre as escolas do sistema público e as escolas privadas independentes, as quais são mantidas financeiramente através de taxas escolares⁷. Segundo os autores, este aspecto torna os seus resultados significativamente diferentes das escolas privadas dependentes de recursos públicos, podendo neutralizar todo o efeito da escola privada. Os resultados encontrados sintetizam dois padrões gerais de escolha da escola: reprodução da escolha de uma classe social e a busca de uma boa e equipada escola. A partir de um pareamento em dois estágios,

⁶ Essa variável foi criada a partir da metodologia de análise fatorial utilizando variáveis dos questionários aplicados aos alunos sobre o número de televisores, rádios, dvd, geladeira, livros em casa, carros, banheiros, quartos para dormir e o grau de escolaridade dos pais.

⁷ Dronkers e Avram (2010 b) faz essa mesma análise para escolas privadas dependentes, isto é, financiadas total ou parcialmente por recursos públicos. Ademais, Dronker e Robert (2008) abordam as mesmas questões utilizando a metodologia de regressões em multi-níveis.

visando controlar o viés de seleção e o processo de escolha da escola, as mais elevadas performances nos exames de leitura de alunos matriculados em escolas privadas independentes tornam-se comparáveis as dos alunos de escolas públicas na maioria dos países, inclusive o Brasil, onde a diferença é significativamente maior. Eles consideram ainda que esse resultado não pode ser explicado pelos processos de escolha da escola, mas, pode ser a indicação de uma maior eficácia das escolas privadas independentes nesses países.

Curi e Menezes-Filho (2010), reconhecendo que os fatores que influenciam a escolha da rede de ensino pelas famílias são combinados tanto do lado da demanda das famílias, através da renda familiar e o nível educacional dos pais, quanto pelo lado da oferta, aferida pelo número de vagas disponibilizadas pela rede pública. Investigaram quais variáveis são relacionadas a decisão das famílias em matricular seus filhos em escolas privadas, e qual o nível de gastos dessas famílias brasileiras em educação. Primeiramente, verifica-se que no período de 2000 a 2008 aumentou o percentual de matrículas nas escolas privadas do Ensino Fundamental e reduziu-se no Ensino Médio, acompanhando o crescimento contínuo de renda nas classes mais baixas⁸. Aplicando modelos de escolha binária aos dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) e da Pesquisa de Orçamentos Familiares (POF), concluíram que a educação da mãe, a renda familiar e a oferta de escolas públicas, custo da educação no estado e a região de moradia são os principais fatores que influenciam a decisão dos pais na escolha por escolas privadas. Os autores supõem que tal escolha decorre da baixa qualidade de ensino observada na rede pública, mesmo sendo gratuita, e de sua limitada oferta de vagas. Concluem ainda que as famílias mais pobres matriculam seus filhos em escolas privadas, mesmo comprometendo cerca de 10% da sua renda com mensalidades⁹.

Em suma, os principais problemas listados nos artigos foram a endogeneidade decorrente de variáveis não observadas e viés de seleção. Para isso, diversas estratégias metodológicas foram utilizadas, como: modelo de variáveis instrumentais, pareamento no escore de propensão e de diferença-em-diferença. Entretanto, como afirma Dronkers e Avram (2010), apesar da expressiva quantidade de pesquisas realizadas, os resultados permanecem inconsistentes e inconclusivos, haja vista as diversidades observadas nos períodos dos estudos, dos desenhos das pesquisas e das variáveis incluídas nos modelos, bem como dos métodos estatísticos utilizados.

3. Dados Descritivos

Foram utilizados os dados do Sistema de Avaliação da Educação Básica Brasileira (SAEB), pesquisa conduzida pelo Ministério da Educação Brasileira – MEC, cuja avaliação é realizada bianualmente, por amostragem das redes de ensino, em cada unidade da federação com focalização nas gestões dos sistemas educacionais. Embora as avaliações tenham se iniciado em 1990, o plano amostral atual se mantém desde 2003, e suas principais ferramentas de avaliação são exames de proficiência em Matemática, com foco na resolução de problemas, e Português, com foco em leitura, aplicados a uma amostra representativa dos alunos do 5º e 9º anos do ensino fundamental e do 3º ano do ensino médio. Complementarmente, são aplicados questionários socioeconômicos com os alunos, professores, diretores e escolas. Em relação ao formato da base de dados gerada, o SAEB consiste de repetidos *cross-sections* de representativas escolas e estudantes de escolas

⁸ Neri (2010) investiga os motivos da evasão escolar no Ensino Médio no Brasil.

⁹ Zoghbi et al (2010) também investigam se a produtividade relativa dos setores públicos e privados da educação é levada em conta no processo de escolha das escolas pelas famílias.

públicas e privadas. A edição utilizada neste artigo é o SAEB de 2005, último ano em que os microdados estão disponíveis¹⁰.

A amostra é composta por 83.929 alunos do 5º ano do ensino fundamental, dos quais, 41.783 e 42.146 realizaram os exames de Matemática e Língua Portuguesa, respectivamente, organizados em 3.004 escolas. Desse total, a amostra contém 29,7% estudantes de escolas privadas, o que caracteriza o objetivo principal da pesquisa, que é monitorar o desempenho dos sistemas de ensino, considerando a atuação dos diferentes entes federados¹¹. Em relação à dependência administrativa das escolas públicas, 35,4% são municipais e 34,5% são estaduais. Retirou-se da amostra os indivíduos que frequentam as escolas públicas federais, mesmo que só representem 0,65% da amostra, essas possuem critérios de ingresso baseados em desempenho e, possivelmente, o nível socioeconômico dos seus alunos possibilitaria a escolha de uma escola privada, ao contrário da maioria das outras escolas públicas (municipais e estaduais). Os critérios de seleção dos alunos mais aplicados nas escolas públicas são o local de moradia e ordem de chegada. Ambos restringem a escolha dos pais sobre qual escola pública matricular os seus filhos, pois as escolas mais procuradas, devido à nota de qualidade do ensino, devem ter uma demanda além da capacidade da escola.

As variáveis de resultados utilizadas como medidas de qualidade do aprendizado são as proficiências padronizadas dos exames de Matemática e Língua Portuguesa dos estudantes do 5º ano do ensino fundamental¹². Na tabela 1 são comparadas as médias dos resultados para os sistemas públicos e privados, subdivididas por gênero e raça para tornar evidente a desigualdade desses atributos¹³.

Os resultados mostram significativas vantagens nos resultados dos estudantes das escolas privadas sobre aqueles de escolas públicas, porquanto, entre todos os grupos, as diferenças são maiores que 25%. Em relação à diferença de sexo, observa-se que as mulheres apresentam melhores resultados nos exames de português, o que não se reflete nos exames de matemática. Ao se analisar as desigualdades de cor ou raça, os alunos auto declarados pretos apresentam piores resultados em ambos os tipos de escolas.

Tabela 1 - Resultados nos Exames de Proficiência do SAEB 2005 dos Alunos do 5º ano do Ensino Fundamental

¹⁰ Em INEP (2007) são apresentados os aspectos metodológicos da construção da amostral e dos instrumentos utilizados na pesquisa, os exames e os questionários.

¹¹ Os dados do Censo Escolar de 2005 informam que somente em torno 10% das escolas de 5º ano do ensino fundamental são privadas e que a maioria das públicas (58%) é municipal. Entretanto, nos últimos 5 anos o número de alunos matriculados em escolas privadas no ensino fundamental vem aumentando cerca de 2% ao ano, provavelmente devido ao crescimento da classe média brasileira. Ademais, os programas do Ministério da Educação FUNDEF e FUNDEB, implementados a partir de 1998, vêm incentivando o processo de municipalização das escolas públicas do ensino fundamental. Razo, Fernandes e Soares (2005) avaliam o efeito da municipalização das escolas estaduais do ensino fundamental no Brasil.

¹² A escolha do 5º ano, final da 1ª fase do ensino fundamental, ao invés do 9º ano do ensino fundamental e 3º ano do ensino médio, deve-se a tentativa de minimizar os efeitos negativos de variáveis educacionais e sócio-culturais que, provavelmente, ampliam-se no processo de aprendizado dos alunos mais velhos. Por exemplo, evasão, trabalho infantil e delinquência j.

¹³ Para dados descritivos que contemplam desigualdades regionais e sócio-econômicas, ver INEP (2007).

Exames	Todas	Escola Pública	Escola Privada	Dif.
Matemática				
Mulher	188,7	173,8	220,3	26,7%
Homem	190,9	174,9	225,8	29,1%
Preto	166,3	161,6	196,4	21,5%
Não Preto	191,8	176,5	224,4	27,1%
Português				
Mulher	186,1	173,1	216,0	24,8%
Homem	172,6	158,5	203,7	28,5%
Preto	158,6	154,0	185,4	20,4%
Não Preto	181,1	167,3	211,1	26,2%

Fonte: Elaboração dos autores com base nos microdados do SAEB, 2005.

Os dados descritivos sobre as variáveis sociais e educacionais dos estudantes e suas famílias utilizados na determinação do escore de propensão estão dispostos na tabela 2. A inclusão destas variáveis tem o objetivo de identificar características que podem influenciar tanto o desempenho nos exames como a seleção da rede de ensino dos alunos. Assim, a variável mulher e preto pretendem controlar as características de gênero e cor ou raça da amostra. As variáveis referentes à distorção idade-série, reprovação e pré-escola caracterizam o histórico acadêmico dos estudantes. A escolaridade dos pais visa captar o efeito da herança educacional e ambiente familiar sobre o desempenho dos filhos, enquanto a estrutura familiar busca controlar o efeito da ausência de um dos pais sobre o ambiente educacional dos mesmos.

Tabela 2 - Características Individuais e Socioeconômicas dos Estudantes do 5º ano do Ensino Fundamental (em %)

Características	Matemática		Português	
	Escola Pública	Escola Privada	Escola Pública	Escola Privada
Mulher	48,8	49,5	48,6	49,5
Preto	30,7	46,0	30,6	45,6
Distorção Idade-Série	22,7	3,5	22,7	3,5
Estrutura Familiar	92,1	96,1	90,7	95,7
Reprovado	33,6	9,9	32,7	9,4
Pré-escola	70,3	91,6	68,5	90,9
Escolaridade da Mãe	-	-	-	-
Nem Fundamental	25,9	4,4	25,7	4,1
Fundamental	11,1	6,6	11,4	6,9
Médio	11,7	17,2	11,3	17,9
Graduação	8,2	37,8	7,8	37,6
Escolaridade do Pai	-	-	-	-
Nem Fundamental	21,5	5,2	21,0	4,8
Fundamental	9,4	6,1	9,4	5,8
Médio	8,6	13,2	8,6	13,1
Graduação	8,9	35,8	8,4	35,6

Fonte: Elaboração dos autores com base nos microdados do SAEB, 2005.

Constata-se nesta tabela que o percentual de estudantes que se consideram de cor/raça preta é cerca de 3 vezes maior na escola pública. Em relação à distorção idade-série, alunos com mais de 2 anos de diferença em relação à idade é significativamente maior o percentual de alunos da rede pública, 22,7%, contra 3,5% da escola privada. Os dados sobre reprovação e

frequência à pré-escola contribuem com o informado sobre distorção idade-série, ratificando a desvantagem dos estudantes de escolas públicas. Nas escolas privadas, menos de 10% dos alunos já tiveram alguma reprovação, enquanto na pública mais de 30%. Quanto à frequência no pré-escolar, cerca de 90% dos alunos de escolas privadas e apenas 70% de escolas públicas tiveram a oportunidade de atender esta fase da educação infantil.

Cabe destacar que é extremamente importante a caracterização do nível socioeconômico dos pais dos alunos e, conseqüentemente, o processo de escolha da rede de ensino. Tendo em vista que nos questionários do SAEB não constam essa informação, o processo de pareamento dos estudantes pode se tornar inviável. Entretanto, é possível inferir-se através da Tabela 2 a correlação entre a escolaridade dos pais e a escolha das escolas privadas, uma vez que 55% dos pais dos estudantes das escolas privadas concluíram pelo menos o ensino médio, contra menos de 20% dos pais de alunos de escolas públicas.

A partir tabela 3, elaborada com base nos dados da Pesquisa Nacional por Amostra de Domicílio (PNAD), pode-se estabelecer a suposição de que a escolaridade, devido ser significativa correlacionada com os rendimentos das pessoas no Brasil, venha a ser o principal representante do nível socioeconômico das famílias que influencia no processo de seleção das escolas. Visto que a maioria das pessoas tem menos de 7 anos de estudo, equivalente a não ter concluído o ensino fundamental, não recebem mais de 2 salários-mínimos, e mais de 90% das pessoas que recebem mais de 5 salários mínimos têm mais de 11 anos de estudo, corroborando com Barbosa Filho e Pessôa (2007) que afirmam ser a taxa de retorno da educação no Brasil extremamente elevada.

Tabela 3 – Brasil - Classes de Rendimento Mensal e Anos de Estudos (%)

	Menos 7 anos	8 a 10 anos	11 a 14 anos	15 anos ou mais	Total
Até 1/2 s. m.	7.57	1.62	0.90	0.03	10.12
Mais de 1/2 a 1 s.m.	12.16	3.90	4.15	0.19	20.40
Mais de 1 a 2 s.m.	12.73	5.56	9.35	0.77	28.41
Mais de 2 a 3 s.m.	3.11	1.80	4.28	0.78	9.97
Mais de 3 a 5 s.m.	2.01	1.32	4.16	1.82	9.31
Mais de 5 a 10 s.m.	0.72	0.56	2.37	2.19	5.84
Mais de 10 a 20 s.m.	0.15	0.10	0.59	1.30	2.14
Mais de 20 s.m.	0.05	0.03	0.16	0.60	0.84
Total	48.06	16.37	27.24	8.05	

Fonte: Dados do SIDRAS/PNAD 2005.

Nota: s.m. é a abreviação para salário mínimo que em 2005 era de 300 reais. Menos de 7 anos equivale a não ter concluído o ensino fundamental. Entre 8 e 10 anos equivale a ter concluído o ensino fundamental. Entre 11 e 14 anos equivale a ter concluído o ensino médio. Mais de 15 anos equivale a ter concluído o ensino superior. O total não soma 100 devido às informações omissas sobre educação e renda na PNAD.

Entretanto, quando se utiliza somente a escolaridade dos pais, não é possível identificar se aqueles pais com menores níveis de escolaridade e elevados níveis de renda tendem a investir na escolaridade dos filhos, considerando que a educação é um significativo fator de mobilidade social. Ou seja, não é observada a motivação relacionada à importância dada à educação pelas famílias que estão ascendendo socialmente. Alternativamente, alguns estudos criam variáveis que tentam caracterizar o nível socioeconômico e as características culturais das famílias dos alunos que influenciam o aprendizado e a escolha das escolas. Entretanto, a escolha das variáveis e a metodologia utilizada para gerar essas variáveis podem inserir um adicional viés na estimativa, devido ao erro de medida do efeito da escola privada sobre a performance dos estudantes.

4. Metodologia

A estratégia de determinação do efeito dos estudantes brasileiros estarem matriculados em escolas privadas relativos aos de escolas públicas sobre o desempenho em exames de proficiência é, primeiramente, através de um modelo que possibilita a identificação pontual do efeito médio do tratamento (EMT). Introduzida por Rosenbaum e Rubin (1983a), a metodologia de Pareamento no Escore de Propensão baseia-se em fortes suposições de identificação. Em seguida, visando apresentar uma metodologia que apresente suposições mais flexíveis e realistas para a intervenção analisada, utiliza-se a metodologia de Identificação Parcial do efeito médio do tratamento, também conhecida como Limites, introduzida por Manski (1989)¹⁴.

A metodologia de Pareamento no Escore de Propensão (PEP), amplamente discutida em Heckman, Ichimura e Todd (1998), Imbens (2004), Blundell e Costa-Dias (2007) e Caliendo e Kopeinig (2008)¹⁵, visa determinar um contrafactual a partir da determinação de indivíduos dentro do grupo controle que sejam semelhantes aos tratados dadas as variáveis observáveis. Essa é uma técnica semi-paramétrica de determinação dos efeitos de um programa em um quase-experimento, baseada em algoritmos de pareamento de indivíduos pertencentes a grupos distintos com o objetivo de se julgar os efeitos de uma determinada intervenção, cujas hipóteses de identificação são a Suposição de Independência Condicional (SIC) ou balanceamento nas variáveis pré-tratamento¹⁶.

A Identificação Parcial, ou Limites, surgiu como uma metodologia alternativa às tradicionais análises de inferência quando as suposições não são confiáveis¹⁷. Isso é provável de ocorrer na análise da diferença de qualidade das escolas pública e privada quanto à suposição de independência condicional e de balanceamento, pois, possivelmente, o viés de seleção devido às condições econômicas das famílias e a motivação de alguns pais em relação à educação seja forte o bastante para que a indisponibilidade dessas variáveis mascare a comparabilidade intrínseca a este estudo. Os principais artigos que têm contribuído para o desenvolvimento desta metodologia são Manski (1989, 1990, 1993, 1997, 2007) e Manski e Pepper (2000)¹⁸, cuja técnica não paramétrica é baseada em fracas suposições ou suposições que tentam representar a regra de seleção ou o efeito do tratamento. As principais suposições são a resposta monotônica do tratamento, a seleção monotônica do tratamento e variável instrumental monotônica¹⁹.

4.1 Análise de Identificação Parcial

A análise de identificação parcial relaxa a hipótese de ignorabilidade forte, simplesmente desconsiderando a suposição de independência condicional. As principais

¹⁴ Blundell e Costa-Dias (2008), Heckman (2008) e Imbens e Woodridge (2008) são estudos que fazem uma análise da metodologia de avaliação de programas abordando as condições de identificação dos principais métodos de avaliação.

¹⁵ Um dos mais citados debates sobre a metodologia PEP deve-se a Dehejia Wahba (1999) ao mostrar, utilizando os dados do estudo de Lalonde (1986), que estimadores do PEP estão mais próximos daqueles de estudos experimentais do que os produzidos pelos métodos tradicionais de avaliação. Entretanto, Smith e Todd (2005) mostraram que os estimadores são melhores em apenas uma subamostra muito específicas dos dados Lalonde, gerando dúvidas sobre a generalização dos resultados de Dehejia e Wahba (1999).

¹⁶ Outras notações para a SIC são a seleção nas variáveis observáveis e *unconfoundedness*. Em relação à suposição de balanceamento, pode-se encontrar o termo sobreposição.

¹⁷ No problema fundamental da avaliação de programas, as suposições de ignorabilidade e imputação podem ser consideradas alvo de questionamento dependendo da intervenção avaliada.

¹⁸ Manski (2008) expõe o desenvolvimento da literatura sobre identificação parcial.

¹⁹ O significado de monotônica é o mesmo atribuído em função monotônica, ou seja, uma função que entre dois conjuntos ordenados preserva (ou inverte) a relação de ordem.

aplicações dessa metodologia ocorrem nos casos em que, mesmo em grandes amostras, não seja possível inferir exatamente o valor do parâmetro, devido às suposições dos modelos paramétricos ou semi-paramétricos não serem teoricamente sustentáveis ou aceitáveis. Nesse sentido, uma estratégia seria não buscar a identificação pontual do parâmetro, mas sim um intervalo de possíveis valores do parâmetro, podendo alguns dos quais ser excluídos a priori com credibilidade Imbens e Woodridge (2008).

Com o objetivo de determinar limites para o Efeito Médio do Tratamento, EMT, define-se que para cada aluno “i” há uma função resposta $y_i(\cdot): T \rightarrow Y$ que mapeia os tratamentos $t \in T$ nos potenciais resultados $y_i(t) \in Y$, para $T=[0;1]$, onde os tratamentos $t=1$ e $t=0$ referem-se aos alunos cursarem escolas privada e pública, respectivamente, e y_i é a performance nos exames de proficiência. Para simplificar a notação, doravante o subscrito i será excluído.

O foco central é aferir o efeito médio da mudança de rede de ensino de $t=0$ para $t=1$, ou seja, nas performances dos estudantes da rede pública para a privada. Ou seja,

$$\Delta(0,1) = E[y(1)] - E[y(0)]. \quad (1)$$

Onde $\Delta(\cdot)$ é o efeito individual do tratamento e $E[y(d)]$ é o valor esperado da performance do estudante com o tratamento d.

Ao usar a lei das expectativas iteradas, seguindo a notação, $E[y(t) | x, z = t] = E[y | x, z = t]$, computa-se este valor médio como segue,

$$E[y(t) | x] = E[y | x, t = 1] \cdot P(t = 1 | x) + E[y | x, t = 0] \cdot P(t = 0 | x). \quad (2)$$

Onde x são as características observadas do aluno; $P(t = 1 | x)$ é a probabilidade de o aluno estar na escola privada.

A inviabilidade no cálculo desta média reside no fato de se desconhecer a performance média caso o aluno de escola privada estivesse freqüentando a escola pública, qual seja, $E[y(1) | x, t = 0]$. Portanto, há que se fazerem suposições sobre o que não é observado para se obter os efeitos de interesse.

Manski (1989) mostra que é possível identificar os limites de $E[y(t) | x]$ se o suporte da variável dependente é limitado com extremos inferior e superior, que é o caso com a performance dos alunos. Substituindo $E[y(1) | x, t = 0]$ pelo menor (K_0) e maior (K_1) nível possível de performance determina-se os limites inferiores e superiores de $E[y(t)]$, gerando, assim, os *limites sem suposição*:

$$E(y | x, z = t)P(z = t | x) + K_0P(z \neq t | x) \leq E(y(t) | x) \leq E(y | x, z = t)P(z = t | x) + K_1P(z \neq t | x) \quad (3)$$

Para estreitar estes *limites*, pode-se adicionar a suposição de resposta monotônica do tratamento (RMT) e a suposição de seleção monotônica do tratamento (SMT), derivadas por Manski (1997) e Manski e Pepper (2000).

A suposição RMT implica que o desempenho de um aluno é fracamente crescente na mudança da escola privada em relação à pública, ou seja:

$$t_2 \geq t_1 \Rightarrow y(t_2) \geq y(t_1). \quad (4)$$

Isso pressupõe a hipótese plausível de que estar freqüentando uma escola privada nunca diminui a performance do aluno, pois, teoricamente, um sistema mais competitivo conduz a um desempenho médio mais elevado dos alunos, embora um efeito zero não seja excluído com esta suposição. Dado que u é o tipo observado da rede de ensino que o aluno freqüenta, a suposição RMT implica o seguinte:

$$\begin{aligned}
\text{Para } u < t & \quad E[y(t) | z = u] \geq E[y(u) | z = u] = E[y | z = u] \\
\text{Então} & \quad E[y(t) | z = u] \in [E[y | z = u], K_1] \\
\text{Para } u > t & \quad E[y(t) | z = u] \leq E[y(u) | z = u] = E[y | z = u] \\
\text{Então} & \quad E[y(t) | z = u] \in [K_0, E[y | z = u]]. \tag{5}
\end{aligned}$$

Com o objetivo de interpretar os limites com a suposição de RMT, considere uma amostra de alunos, distribuídos nas duas redes de ensino, onde, na pública, por ser menos competitivo e eficiente, gera um desempenho médio observado dos alunos menor ou igual daqueles da rede privada. Assim, a partir da proficiência média observada, chega-se ao limite inferior. Usando o mesmo procedimento, a rede privada forneceria o limite superior, que, combinados com os *limites sem suposição*, obtêm-se os limites RMT:

$$\begin{aligned}
& E[y | z < t] \cdot P(z < t) + E[y | z = t] \cdot P(z = t) + K_0 \cdot P(z > t) \\
& \leq E[y(t)] \leq \dots \\
& K_1 \cdot P(z < t) + E[y | z = t] \cdot P(z = t) + E[y | z > t] \cdot P(z > t)
\end{aligned} \tag{6}$$

Para reduzir a amplitude dos limites, pode-se acrescentar a suposição de seleção monotônica do tratamento, SMT, na qual, estudantes de escolas privadas têm fracamente maiores médias de proficiência do que aqueles de escolas públicas, ou seja:

$$u_2 \geq u_1 \Rightarrow E[(y(t) | z = u_2)] \geq E[(y(t) | z = u_1)]. \tag{7}$$

Esta suposição é consistente com o fato de as escolas privadas terem características que podem afetar positivamente (mas não negativamente) o desempenho dos estudantes, a qual pode também ser relacionada com a maior flexibilidade na gestão da escola privada em focar na melhoria do desempenho de seus estudantes. Combinando a suposição de RMT e SMT estabelecem-se as seguintes relações:

$$\begin{aligned}
\text{Para } u < t & \quad E[y(t) | z = u] \geq E[y(t) | z = u] \geq E[y(u) | z = u] \\
\text{Então} & \quad E[y(t) | z = u] \in [E[y | z = u], E[y | z = t]] \\
\text{Para } u > t & \quad E[y(t) | z = u] \leq E[y(u) | z = u] \leq E[y(u) | z = u] \\
\text{Então} & \quad E[y(t) | z = u] \in [E[y | z = t], E[y | z = u]]. \tag{8}
\end{aligned}$$

Para uma melhor compreensão na determinação dos limites, seja uma amostra composta de alunos das duas redes de ensino. Se um aluno se transferir da escola pública para a privada, supõe-se que, devido SMT, seu desempenho médio seria fracamente inferior ao de outro de escola privada. Pode-se, portanto, usar o desempenho médio observado para os alunos das escolas privadas como um limite superior para aqueles de escolas públicas. Combinando as suposições de RMT e de SMT obtêm-se os respectivos limites:

$$\begin{aligned}
& E[y | z < t] \cdot P(z < t) + E[y | z = t] \cdot P(z = t) + E[y | z = t] \cdot P(z > t) \\
& \leq E[y(t)] \leq \\
& E[y | z = t] \cdot P(z < t) + E[y | z = t] \cdot P(z = t) + E[y | z > t] \cdot P(z > t)
\end{aligned} \tag{9}$$

É possível testar as suposições conjuntas de RMT-SMT, pois se pode verificar que:

Para $u_2 \geq u_1$

$$E[y | z = u_2] = E[(y(t) | z = u_2)] \geq E[(y(u_2) | z = u_1)] \geq E[(y(t) | z = u_1)] = E[(y | z = u_1)]$$

Isso implica que o desempenho médio de um aluno deve ser fracamente crescente em relação àqueles matriculados em escolas privadas, caso contrário, a suposição RMT-SMT deve ser rejeitada.

Até aqui foram obtidos apenas os limites sobre $E[y(t)]$, embora o objetivo seja medir o efeito da mudança do aluno da escola pública para a privada, $E[y(t=1)] - E[y(t=0)]$. Para se obter os limites sobre este efeito do tratamento, subtrai-se o limite inferior (superior) de $E[y(t=0)]$ do limite superior (inferior) de $E[y(t=1)]$, assim, os limites estariam definidos. Para os limites utilizando a suposição RMT, o limite inferior da mudança da rede pública para a privada não pode ser negativo e, portanto, é definido como zero.

Suponha-se que são observados, não apenas o desempenho do estudante e a rede a que pertence, mas também uma variável z^* que caracterize indiretamente os alunos de escolas públicas e privadas. Em seguida, divide-se a amostra em subamostras, uma para cada valor de z^* e, para cada uma delas, obtém-se os limites sem suposição com base na equação (3) anteriormente descrita. Isso pode resultar em um diferencial de limites menor para alguma subamostra e maior para outras. Poder-se-ia explorar esta variação nos limites sobre as subamostras se z^* satisfizer a suposição de variável instrumental monotônica (VIM) apresentada em Manski e Pepper (2000), a qual é dada por:

$$m_1 \leq m \leq m_2 \Rightarrow E[y(t) | z^* = m_1] \leq E[y(t) | z^* = m] \leq E[y(t) | z^* = m_2] \tag{10}$$

Onde, m_1 é o limite natural inferior da variável utilizada como instrumento e m_2 é o limite superior.

Então, ao invés de assumir independente em média, a suposição de VIM permite uma fraca relação monotônica entre a variável z e a função do desempenho dos estudantes (MANSKI e PEPPER, 2000). Com isso, pode-se novamente dividir a amostra em subamostras com base em z^* , e obter limites sem suposição para cada subamostra. Da equação (10), segue-se que $E[y(t) | z^* = m]$ não é menor do que o inferior limite sem suposição sobre $E[y(t) | z^* = m_1]$ e não é superior do que o superior limite sem suposição sobre $E[y(t) | z^* = m_2]$. Para a subamostra onde z^* tem o valor m , pode-se derivar um novo limite inferior, que é o limite inferior absoluto sobre todas as subamostras, onde z é inferior ou igual a m . Da mesma forma, obtém-se um novo limite superior, tendo o menor limite superior sobre todas as subamostras com um valor de z superior ou igual a m . Ao repetir o procedimento para todos os possíveis valores de m , e extraindo a média, estabelecem-se os seguintes VIM-limites:

$$\begin{aligned}
& \sum_{m \in M} P(z^* = m) \cdot \left[\max_{m \in M} \left(E[y | z = t, z^* = m] \cdot P(z = t, z^* = m) + K_0 \cdot P(z \neq t | z^* = m) \right) \right] \\
& \leq E[y(t)] \leq \\
& \sum_{m \in M} P(z^* = m) \cdot \left[\min_{m \in M} \left(E[y | z = t, z^* = m] \cdot P(z = t, z^* = m) + K_1 \cdot P(z \neq t | z^* = m) \right) \right]
\end{aligned} \tag{11}$$

O procedimento a ser adotado, evitando a suposição VIM com os limites sem suposição, baseia-se em combinar a suposição VIM com os limites RMT-SMT em cada

subamostra. . Ao substituir os limites sem suposição na equação (11) pelos limites RMT-SMT, então, os limites RMT-SMT-VIM são definidos pelas seguintes relações:

$$\sum_{m \in M} P(z^* = m) \cdot \left[\max_{m_1 \leq m} \begin{pmatrix} E[y | z < t, z^* = m_1] \cdot P(z < t, z^* = m_1) + \\ E[y | z = t, z^* = m_1] \cdot P(z = t, z^* = m_1) + \\ E[y | z > t, z^* = m_1] \cdot P(z > t, z^* = m_1) \end{pmatrix} \right] \leq E[y(t)] \leq \sum_{m \in M} P(z^* = m) \cdot \left[\min_{m_2 \geq m} \begin{pmatrix} E[y | z = t, z^* = m_2] \cdot P(z < t, z^* = m_2) + \\ E[y | z = t, z^* = m_2] \cdot P(z = t, z^* = m_2) + \\ E[y | z > t, z^* = m_2] \cdot P(z > t, z^* = m_2) \end{pmatrix} \right] \quad (12)$$

Como a variável instrumental monotônica (VIM) mais adequada, em bases teóricas, seria a renda familiar do aluno, optou-se, na indisponibilidade desta, por uma *proxy*, notadamente a posse dos pais de pelo um automóvel. Certamente que não seria possível utilizar esta *proxy* como uma variável instrumental convencional, pois é pouco provável que a função do desempenho dos alunos seja independente de alguma variável que reflita a condição econômica da família. Entretanto, ela pode ser usada como uma VIM, pois, para tal, assume-se que a função média do desempenho do aluno seja monotonicamente crescente (ou não decrescente) em relação à propriedade de automóveis pela família.

De posse dos limites que definem o efeito da mudança da rede pública para a privada sobre o desempenho dos alunos, determinam-se os limites RMT-SMT superiores e inferiores para $E[y(t=0)]$ e $E[y(t=1)]$ e, em seguida, toma-se a diferença do limite superior para $E[y(t=1)]$ e do limite inferior de $E[y(t=0)]$ para obter o limite superior do Efeito Médio do Tratamento, $\Delta(0,1) = E[y(1)] - E[y(0)]$. A igualdade, $\Delta(0,1) = E[y(1)] - E[y(0)] = 0$, se justifica devido à suposição de resposta monotônica ao tratamento.

5. Resultados

Devido às escolhas dos pais sobre a escola dos filhos seguir restrições relacionadas com a renda familiar e com a oferta de escolas de qualidade, é improvável que uma simples comparação dos resultados médios do desempenho dos alunos das escolas públicas e privadas produza estimativas confiáveis sobre o efeito causal do sistema privado de ensino sobre o desempenho dos alunos. Pelo pareamento de casos similares do grupo de tratamento e controle, o método de pareamento no escore de propensão busca eliminar o viés de seleção decorrente das variáveis observáveis e, conseqüentemente, parear indivíduos no grupo tratamento e controle que imitem os indivíduos de um experimento aleatório.

Neste estudo, verificada a assimetria da distribuição dos escores de propensão, aplica-se o algoritmo de Pareamento Linear Local, descrito em Caliendo e Popeign (2008), que faz combinações usando todos os indivíduos na amostra de comparação atribuindo pesos menores para as observações mais distantes. O escore de propensão foi operacionalizado como a probabilidade prevista do aluno estar matriculado na rede privada. As probabilidades foram estimadas a partir de uma regressão de escolha discreta binária²⁰ sob a hipótese *logit* para o aluno frequentar ou não a escola privada sobre as características observadas que são relacionadas com o desempenho dos alunos, como: gênero, raça, pais residentes no domicílio, escolaridade dos pais, distorção idade-série, antecedente de reprovação, frequência a pré-

²⁰ Nestas regressões, o sinal do coeficiente da variável explicativa ser positivo/negativo indica a direção do efeito de aumentar/diminuir a probabilidade de um aluno estar matriculado em uma escola privada.

escola e identificação dos estados²¹. As estimativas das regressões referentes aos escores de propensão para as amostras de estudantes que realizaram os exames de Língua Portuguesa e Matemática estão dispostas na tabela 4.

Para os estudantes que realizaram o exame de Língua Portuguesa, os resultados da regressão apontam que os coeficientes de todas as variáveis explicativas são estatisticamente significantes ao nível de, no máximo, 10%. Quanto às características herança familiar, percebe-se, inicialmente, o efeito negativo dos pais que não concluíram o Ensino Fundamental, o que significa um redutor sobre a probabilidade de o filho estar em uma escola privada. O oposto ocorre quando os pais possuem nível superior, o qual é um indicativo do efeito renda familiar sobre a escolha do sistema privado de ensino no Brasil. Ademais, outros importantes resultados que apontam este viés de seleção oriunda de variáveis relacionadas à renda é sinal negativo das variáveis se o aluno já foi reprovado ou se o aluno apresenta distorção idade-série, e o sinal positivo da variável se o aluno frequentou pré-escola. Ou seja, os alunos que nunca foram reprovados, não apresentam distorção idade-série e frequentaram a pré-escola têm maior probabilidade de estar em uma escola privada. Em relação às *dummies* dos estados, os resultados variam de sinal e foram omitidos para melhor apresentação da tabela. Isso pode estar refletindo as diferentes condições de cada estado em relação ao percentual de alunos matriculados na rede privada de ensino. Os resultados das estimativas para as *dummies* dos estados seguem em Anexo. Com relação aos resultados do exame de Matemática, o padrão de sinais e significância dos coeficientes segue os do exame de Português, porém divergem em suas magnitudes.

Tabela 4 – Estimativas das Regressões Logísticas para os Alunos que Realizaram os Exames de Português e Matemática

Variáveis	Exames de Proficiência	
	Português 5º ano	Matemática 5º ano
Mulher	-0.151* (0.03)	-0.103* (0.03)
Preto	-0.813* (0.05)	0.546* (0.03)
Família completa	0.318* (0.07)	0.310* (0.07)
Mãe sem EF	-1.216* (0.04)	-1.275 (0.05)
Pai sem EF	-0.836* (0.05)	-0.843* (0.05)
Mãe Graduada	1.165* (0.05)	1.195* (0.05)
Pai Graduado	0.928* (0.05)	0.889* (0.05)
Distorção Idade-Série	-0.487* (0.02)	-0.501* (0.02)
Reprovado	-0.686* (0.04)	-0.680* (0.05)
Pré-escola	1.065* (0.04)	1.068* (0.04)
Constante	-0.675* (0.10)	-0.980* (0.10)
N	26807	
Log-likelihood	11584.6	
Pseudo R ²	0.313	

Fonte: Elaboração dos autores.

Nota: p-valores entre parênteses.

²¹ Caliendo e Kopeinig (2008) sintetizam a idéia que há sobre a escolha das variáveis explicativas que devem ser inseridas no cálculo do escore de propensão. Em suma, a quantidade não deve ser muito grande ou nem muito pequena, devem conter variáveis relacionadas tanto com o tratamento como com o resultado e, principalmente, as variáveis devem ser inseridas visando melhorar o balanceamento entre o grupo controle e tratamento.

A metodologia de pareamento pondera a amostra do grupo controle a fim de aumentar a semelhança com os indivíduos do grupo tratamento, visando balancear as características observadas da amostra do grupo tratamento e controle após o pareamento. Uma análise de balanceamento das variáveis utilizadas na estimação do escore de propensão é mostrada na tabela 5. Para isso, apresentam-se as médias dos escores de propensão e de todas as outras variáveis explicativas antes e depois do pareamento, e uma medida do viés entre a amostra do grupo de tratamento e controle²². Analisando os resultados para a amostra de estudantes que realizaram os exames de Língua Portuguesa e Matemática, verifica-se que em quase todos os casos é evidente que as diferenças da amostra dos dados antes do pareamento são significativamente superiores aos pareados. Isso acarreta uma redução dos vieses dessas variáveis observadas e dos escores de propensão. Isto é, o processo de pareamento gera um elevado grau de balanceamento das variáveis entre a amostra do grupo tratamento e do controle que são utilizados no processo de estimação.

Tabela 5 - Balanceamento das Variáveis Utilizadas no Pareamento dos Alunos

Variáveis	Amostra	Língua Portuguesa			Matemática		
		Média		Viés (%)	Média		Viés (%)
		Tratamento	Controle		Tratamento	Controle	
Mulher	Não Pareado	0.499	0.495	0.9	0.497	0.489	1.6
	Pareado	0.499	0.497	0.4	0.497	0.498	-0.2
Preto	Não Pareado	0.055	0.146	-30.7	0.462	0.300	33.9
	Pareado	0.055	0.048	2.4	0.462	0.460	0.5
Família completa	Não Pareado	0.962	0.921	17.4	0.962	0.922	17.5
	Pareado	0.962	0.970	-3.4	0.962	0.968	-2.5
Mãe sem EF	Não Pareado	0.068	0.366	-77.5	0.044	0.276	-66.7
	Pareado	0.068	0.067	0.2	0.044	0.044	0.2
Pai sem EF	Não Pareado	0.070	0.296	-61	0.052	0.227	-52.3
	Pareado	0.070	0.068	0.7	0.052	0.048	1.1
Mãe Graduada	Não Pareado	0.376	0.065	80.9	0.379	0.070	79.5
	Pareado	0.376	0.362	3.7	0.379	0.373	1.5
Pai Graduado	Não Pareado	0.357	0.074	73.4	0.358	0.081	71.2
	Pareado	0.357	0.349	2.1	0.358	0.348	2.6
Distorção Idade-Série	Não Pareado	0.158	0.923	-71.2	0.167	0.904	-69.9
	Pareado	0.158	0.155	0.3	0.167	0.156	1.0
Reprovado	Não Pareado	0.095	0.336	-61.3	0.099	0.339	-60.5
	Pareado	0.095	0.088	1.7	0.099	0.092	1.8
Pré-escola	Não Pareado	0.914	0.676	61.7	0.917	0.697	57.9
	Pareado	0.914	0.918	-0.9	0.917	0.927	-2.9
Escore de Propensão	Não Pareado	0.662	0.289	155.9	0.665	0.309	149.8
	Pareado	0.662	0.662	0.0	0.665	0.665	0.0

Fonte: Elaboração dos autores.

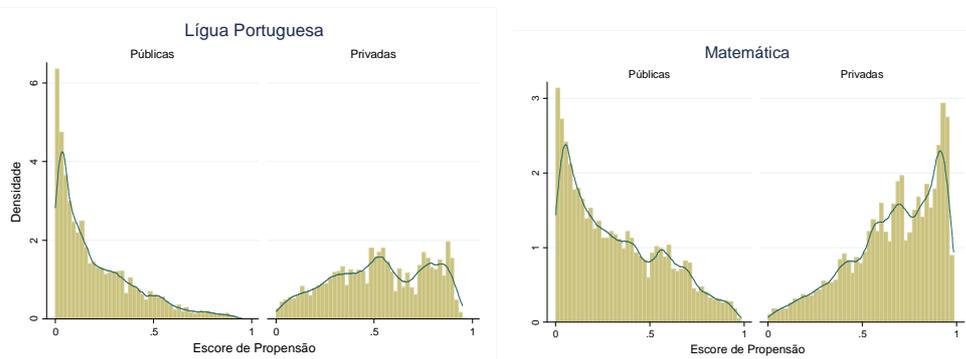
Outra importante fonte de identificação da sobreposição do escore de propensão são os histogramas dos escores de propensão para os indivíduos do grupo controle e tratamento, apresentados no gráfico 1. Percebe-se, visualmente, que as distribuições são assimétricas e não são em torno dos mesmos valores de escore de propensão, ou seja, os indivíduos não apresentam escores de propensão que possibilitam o direto pareamento desses escores. Ademais, verifica-se que em ambas as amostras dos alunos que participaram dos exames de Língua Portuguesa e Matemática os escores de propensão dos alunos distribuem-se de forma

²⁶ A medida de viés sugerida por Rosenbaum (2002), que utiliza a diferença das médias padronizadas das amostras do grupo tratamento e controle, é dada por:

$$Viés = \frac{100(\bar{x}_T - \bar{x}_C)}{\sqrt{\frac{s_T^2 + s_C^2}{2}}}$$

amostrais e \bar{s}_T^2 e \bar{s}_C^2 são as variâncias amostras dos grupos tratamento e controle.

distinta, evidenciando a diferença entre elas. Essa assimetria na distribuição dos escores de propensão confirma a necessidade de se aplicar o algoritmo de Pareamento Linear Local discutido anteriormente.



Fonte: Elaboração dos autores.

Gráfico 1 - Sobreposição do Escore de Propensão

Uma vez que o pareamento no escore de propensão reduz o viés atribuído às variáveis observadas, pode-se usar a diferença nos resultados médios encontrados nas amostras para se obter uma estimativa do Efeito Médio do Tratamento (EMT). A Tabela 6 expõe essas estimativas a partir da metodologia de Pareamento no Escore de Propensão aplicando o algoritmo de Pareamento Linear Local. Nas colunas “Grupos” nesta tabela constam as médias condicionais dos exames de proficiências antes e após o pareamento para o grupo tratamento e controle. A coluna denominada “Diferença” refere-se à diferença entre essas duas médias condicionais, ou seja, entre o viés de seleção e a estimativa do EMT (a diferença que tenta contornar o viés de seleção baseando-se na SIC). As colunas do teste t trazem o erro-padrão e a estatística t das estimativas. Na última coluna estão dispostas as estimativas de Mínimos Quadrados Ordinários utilizando todas as variáveis do escore de propensão como variáveis explicativas.

Tabela 6 - Pareamento no Escore de Propensão – Exames de Português e Matemática

Resultado	Amostra	Grupos		Diferença	p-valor	MQO
		Tratamento	Controle			
Exames de Português do 5º ano	Não Pareados	-0,721	-1,563	0,842	0,000	
	EMT	-0,721	-1,362	0,641	0,000	0,541*
Exames de Matemática do 5º ano	Não Pareados	-0,477	-1,396	0,919	0,000	
	EMT	-0,477	-1,208	0,731	0,000	0,628*

Fonte: Elaboração dos autores. Nota: *significante a 1%.

Primeiramente, nota-se que todas as médias padronizadas dos exames de proficiência são negativas, contrárias, portanto, aos sinais positivos e negativos esperados para escolas privadas e públicas, respectivamente (tratamento e controle). Isso ocorre devido ao fato de não serem apresentadas médias condicionais em relação ao escore de propensão, o que gerou médias negativas para ambos os grupos antes e depois do pareamento. Segundo, os resultados apontam para significativos e positivos efeitos das escolas privadas sobre o desempenho dos alunos nos exames de Português e Matemática. Isto é, o efeito médio do aluno estar matriculado em uma escola privada sobre o desempenho nos exames de português e matemática são 0,641% do desvio-padrão e 0,731%, respectivamente, maiores do que alunos matriculados em escolas da rede pública. Comparando com as estimativas da metodologia de Mínimos Quadrados Ordinários (MQO), esses resultados apontam um efeito superior do

sistema privado de ensino, pois os resultados por MQO são 0,541% e 0,628% do desvio-padrão, respectivamente²³.

Os resultados da análise de Identificação Parcial são apresentados nas tabela 7 e 8. Um ponto prático na comparação destes resultados, como os da metodologia PEP e MQO, são as variáveis explicativas que condicionam as estimativas dos efeitos. Tanto nas metodologias não paramétricas de Pareamento como na de Identificação Parcial enfrenta-se o problema gerado pelo número de variáveis explicativas que condicionam as estimativas²⁴. Na primeira, utiliza-se a estratégia proposta por Rosenbaum e Rubin (1983) de pareamento no escore de propensão já citada na seção 4. Na análise de Identificação Parcial, que utiliza o estimador de Polinômio Local de Pesos Kernel (*Kernel-weighted local polynomial*), essa barreira é mais difícil de superar²⁵. Dessa forma, com o intuito de tornar comparável os resultados destas metodologias, incluídas todas as variáveis explicativas no cálculo do escore de propensão, utiliza-se o resíduo previsto da regressão por MQO da proficiência dos alunos sobre as mesmas variáveis explicativas utilizadas no escore de propensão²⁶.

Tabela 7 - Identificação Parcial - Matemática

Matemática		
Suposições	Limite Inferior	Limite Superior
Limites Sem Suposição		
$E[y(0) X]$	-1,877	-0,266
$E[y(1) X]$	-2,539	1,361
$E[y(1)-y(0) X]$	-2,273	3,238
Resposta Monotônicaa ao Tratamento		
$E[y(0) X]$	-1,877	-1,114
$E[y(1) X]$	-1,114	1,361
$E[y(1)-y(0) X]$	0,000	3,238
Resposta Monotônicaa ao Tratamento e Seleção Monotônicaa ao Tratamento		
$E[y(0) X]$	-1,114	-1,114
$E[y(1) X]$	-1,114	-0,692
$E[y(1)-y(0) X]$	0,000	0,422
Variável Instrumental Monotônica, RMT e SMT		
$E[y(0) X]$	-0,438	-0,277
$E[y(1) X]$	-0,153	-0,153
$E[y(1)-y(0) X]$	0,124	0,285

Fonte: Elaboração dos autores.

Tabela 8 - Identificação Parcial – Língua Portuguesa

²³ Os resultados estão medidos em relação ao desvio-padrão devido as variáveis dependentes terem sido padronizadas com média 0 e desvio-padrão 1, como foi informado na descrição das variáveis.

²⁴ Essa dificuldade computacional é conhecida como *curse of dimensionality*.

²⁵ Não faz parte do escopo do estudo discutir o algoritmo da regressão de Kernel utilizado.

²⁶ Essa estratégia tem o objetivo de utilizar o resíduo da regressão por MQO como síntese dos efeitos relativos das outras variáveis, mas reconhecesse que isso insere uma série de questões metodológicas que não são o foco deste estudo.

Língua Portuguesa		
Suposições	Limite Inferior	Limite Superior
Limites Sem Suposição		
$E[y(0) X]$	-1,908	-0,663
$E[y(1) X]$	-2,815	1,096
$E[y(1)-y(0) X]$	-2,152	3,004
Resposta Monotônica ao Tratamento		
$E[y(0) X]$	-1,908	-1,318
$E[y(1) X]$	-1,318	1,096
$E[y(1)-y(0) X]$	0,000	3,004
Resposta Monotônica ao Tratamento e Seleção Monotônica ao Tratamento		
$E[y(0) X]$	-1,318	-1,318
$E[y(1) X]$	-1,318	-0,961
$E[y(1)-y(0) X]$	0,000	0,357
Variável Instrumental Monotônica, RMT e SMT		
$E[y(0) X]$	-0,467	-0,342
$E[y(1) X]$	-0,201	-0,201
$E[y(1)-y(0) X]$	0,140	0,265

Fonte: Elaboração dos autores.

Nas linhas das tabelas 7 e 8 são expostos os limites para as esperanças condicionais dos estudantes tratados e não tratados, $E[y(1)|X]$ e $E[y(0)|X]$, respectivamente. Essas estimativas mostram claramente que os estudantes de escolas privadas, em média, obtêm melhores resultados. Esse resultado é visualizado sob todas as suposições, mas a amplitude dos intervalos varia significativamente entre as suposições. Por exemplo, em ambos os exames, os limites superiores das expectativas condicionais dos tratados, $E[y(1)|X]$, no caso de Limites sem Suposição e Resposta Monotônica ao Tratamento, são positivos. Por outro lado, no caso das suposições de Resposta Monotônica ao Tratamento em conjunto da suposição de Seleção Monotônica ao Tratamento e Variável Instrumental Monotônica todos os limites são negativos, tanto para as expectativas condicionais dos tratados e não tratados.

Ao se analisar os limites do Efeito Médio do Tratamento, $\Delta(0,1) = E[y(1) - y(0) | X]$, aqueles sem suposição são de larga amplitude e não provêm nenhuma informação extra em qualquer dos exames, pois ambas as estimativas do PEP e do MQO se encontram neste intervalo. Isso também é observado para os limites sob a suposição de Resposta Monotônica ao Tratamento. Esse fato já era esperado, visto que todas as aplicações da metodologia de Identificação Parcial apresentaram limites não informativos sob essas suposições, provavelmente devido à fraqueza das restrições das suposições. Entretanto, os intervalos são informativos quando se analisa os limites sob a suposição de Resposta Monotônica ao Tratamento em conjunto da suposição de Seleção Monotônica ao Tratamento. Por definição da suposição de Resposta Monotônica ao Tratamento, os limites inferiores são zero. Os limites superiores para o efeito do aluno estar frequentando a escola privada sob os exames de Língua Portuguesa e Matemática são de 0,35 do desvio-padrão e 0,42, do desvio-padrão respectivamente. Esses intervalos são informativos, porque as estimativas das metodologias de Pareamento no Escore de Propensão e Mínimos Quadrados Ordinários são superiores, ficando acima de 0,50 do desvio-padrão.

Utilizando a variável referente à dotação de veículos na residência do aluno, os limites do Efeito Médio do Tratamento sob as suposições de Variável Instrumental Monotônica, Resposta Monotônica ao Tratamento e Seleção Monotônica ao Tratamento são mais informativos, pois restringem principalmente os limites inferiores a valores não nulos ou negativos. Para o desempenho dos estudantes em Língua Portuguesa os limites inferior e superior do efeito da escola privada são 0,14 e 0,265, respectivamente, enquanto para o exame de Matemática a correspondente amplitude dos limites são 0,124 e 0,285. Portanto, os intervalos obtidos em ambos os exames implicam que as estimativas pela metodologia de

Pareamento no Escore de Propensão e MQO do efeito da escola privada podem estar de 2 a 4 vezes sobreestimados.

Considerações Finais

Diversos estudos recentes realizados em países desenvolvidos têm comprovado a relativa eficiência da escola privada diante das demais. Em relação à realidade brasileira, são questionáveis as suposições impostas nos modelos que identificam pontualmente a representativa diferença de qualidade do ensino da rede pública e privada. A improvável comparabilidade de estudantes que demandam estes dois sistemas, e o provável viés de seleção devido a fatores imensuráveis, são os fatores que alimentam o ceticismo sobre, principalmente, as suposições da metodologia de Pareamento no Escore de Propensão. Seguindo uma estratégia menos restritiva, volta-se para a análise de Identificação Parcial como uma ferramenta de esclarecimento do efeito da escola privada sobre o desempenho educacional, em detrimento da identificação de somente limites do efeito.

Os resultados das estimativas derivadas da aplicação de Pareamento no Escore de Propensão e o tradicional método de Mínimos Quadrados, a partir dos dados do Sistema de Avaliação da Educação Básica Brasileira (SAEB), revelam um significativo efeito das escolas privadas sobre o desempenho dos estudantes, tanto em Língua Portuguesa quanto em Matemática. Através dos dados descritivos ficaram igualmente evidenciadas as diferenças entre as realidades dos estudantes que afetam o processo de escolha da escola pelos pais e a aprendizagem dos alunos. Isso pode estar refletindo a desigualdade de renda brasileira que influencia diretamente na desigualdade de oportunidades dos estudantes brasileiros, pois a possibilidade de escolha da escola e o ambiente educacional propiciado pelos pais dos alunos com maior renda induzem fortes restrições, geradas tanto pela família quanto pela escola, sobre o processo de aprendizagem.

Reconhece-se, contudo, a afirmativa de Morgan (2001), segundo o qual o mais importante para políticas educacionais é que as estimativas derivadas das metodologia de regressão e Pareamento no Escore de Propensão não provêm informações suficientes sobre como os estudantes de escola públicas poderiam se beneficiar caso frequentasse uma escola privada. Esse fato é demonstrado na estrutura de contrafactual e justifica a necessidade da coleta de dados mais informativos e técnicas com suposições mais flexíveis e confiáveis. Assim, a metodologia de Identificação Parcial aponta para a sobreestimação do efeito da escola privada, refletindo a inadequação das fortes suposições que possibilitam a identificação pontual.

Mesmo reconhecendo a sobreestimação das estimativas do efeito da escola privada no Brasil, os resultados da Identificação Parcial advogam a favor do aumento da possibilidade de escolha dos pais, desde que sua renda permita, sobre qual sistema escolar matricular seus filhos, visando aumentar a competitividade no setor. Conseqüentemente, essa mudança poderia diminuir a burocracia do setor e aproximaria a escolha da escola com as características sócio-econômicas dos pais. Por exemplo, escolas dotadas de professores mais qualificados e motivados poderiam se tornar mais eficientes e guiar os pais a tomar decisões sobre o destino da rede de ensino para os filhos.

Referências Bibliográficas

AAKVIK, A. Bounding a matching estimator: the case of a Norwegian training program. **Oxford Bulletin of Economics and Statistics**, v. 63, n. 1, p. 115–143, 2001.

- ALTONJI, J. G.; ELDER T. E.; TABER C. R. Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools. **Journal of Political Economy**, v. 113, n. 1, p. 151-184, 2005.
- BARBOSA FILHO, F. H.; PESSÔA, S.. Retorno da Educação no Brasil. **Pesquisa e Planejamento Econômico**, v. 38, n.1, 2008.
- BECKER, S.O.; CALIENDO, M. Sensitivity analysis for average treatment effect. **Stata Journal**, v. 7, n. 1, p. 71–83, 2007.
- BLUNDELL, R.; DEARDEN, L.; SIANESI, B. Evaluating the impact of education on earnings in the UK: models, methods and results from the NCDS. **Journal of the Royal Statistical Society, Series A**, v. 168, n. 3, p. 473–512, 2005.
- BLUNDELL, R.; COSTA-DIAS, M. Alternative Approaches to Evaluation in Empirical Microeconomics. **Journal of Human Resources**, v. 44, n. 3, 2009.
- CALIENDO, M.; HUJER, R.; THOMSEN, S. The employment effects of job creation schemes in Germany – a microeconomic evaluation. **IZA Discussion Paper**, Nº. 1512, 2007.
- CALIENDO, M.; KOPEINIG, S.. Some Practical Guidance for The Implementation of Propensity Score Matching. **Journal of Economic Surveys**, v. 22, p. 31–72, 2008.
- COLEMAN, J.S.; HOFFER, T.; KILGORE, S.. **High School Achievement: Public, Catholic and Private Schools Compared**. Basic Books. 1982. Disponível em: <http://www.questia.com/PM.qst?a=o&d=100282593>. Acesso em: 21 out. 2010.
- CRUMP, R.; HOTZ V. J.; IMBENS G.; MITNIK, O. Dealing with Limited Overlap in Estimation of Average Treatment Effects. forthcoming **Biometrika**, 2008.
- CURI, A. Z.; MENEZES-FILHO, N. A.. Determinantes dos Gastos com Educação no Brasil. **Pesquisa e Planejamento Econômico**, v. 40, n. 1, 2010.
- DEHEJIA, R.; WABBA, S. Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. **Journal of the American Statistical Association**, v. 94, p. 1053-1062, 1999.
- EVANS, W.; SCHWAB R. M. Finishing High School and Starting College: Do Catholic Schools Make a Difference? **The Quarterly Journal of Economics**, v. 110, n. 4, p. 941-974, 1995.
- FRANÇA, M. T. A.; GONÇALVES, F. O.. Provisão pública e privada de educação fundamental: diferenças de qualidade medidas através de propensity score matching. In: **XXXVII Encontro Nacional de Economia – ANPEC**, 2009.
- HANUSHEK, E; WOESSMANN, L.. The role of cognitive skills in economic development. **Journal of Economic Literature**, v 46, n.3, p. 607-668, 2008.
- HECKMAN, J.; ICHIMURA, H.; TODD, P. Matching as an econometric evaluation estimator. **Review of Economic Studies**, v. 65, n. 2, p. 261–294, 1998.
- HOXBY, C. M. Do Private Schools Provide Competition for Public Schools? **NBER working paper**, no. 4978, 1994.
- HOXBY, C. M. The effects of class size on student achievement: new evidence from population variation. **Quarterly Journal of Economics**, v.116, p. 1239–1286, 2000a.
- HOXBY, C. M. Does competition among public schools benefit students and taxpayers. **The American Economic Review**, v.90, n. 5, p. 1209-1238, 2000b.
- ICHINO, A.; MEALLI, F.; NANNICINI, T. From temporary help jobs to permanent employment: what can we learn from matching estimators and their sensitivity. **IZA Discussion Paper**, Bonn, No. 2149, 2006.
- IMBENS, G. The role of the propensity score in estimating dose–response functions. **Biometrika**, v. 87, n. 3, p. 706–710, 2000.
- _____. Sensitivity to exogeneity assumptions in program evaluation. **American Economic Review**, v. 93, n. 2, p. 126–132, 2003.

- _____. Nonparametric estimation of average treatment effects under exogeneity: a review. **Review of Economics and Statistics**, v. 86, n. 1, p 4–29, 2004.
- IMBENS, J.; WOODRIDGE, J.. Recent Developments in the Econometrics of Program Evaluation. **Journal of Economic Literature**, v. 47, n. 1, p. 5–86, 2009. Disponível em: <http://www.aeaweb.org/articles.php?doi=10.1257/jel.47.1.5> . Acesso em: 21 out. 2010.
- INEP. Primeiros Resultados: Médias de desempenho do SAEB/2005 em perspectiva Comparada. Publicações do INEP, 2007. Disponível em: http://www.inep.gov.br/download/saeb/2005/SAEB1995_2005.pdf. Acesso em: 21 out. 2010.
- KHANDKER, S.; KOOLWAL, G.; SAMAD, H. Handbook on Impact Evaluation. **World Bank**, Washington DC, 2010.
- LECHENER, M. Earnings and employment effects of continuous off-the-job training in East Germany after unification. **Journal of Business Economic Statistics**, v. 17, n. 1, p. 74–90, 1999.
- _____. Identification and estimation of causal effects of multiple treatments under the conditional independence assumption. **Econometric Evaluation of Labour Market Policies**, Heidelberg, p. 1–18, 2001.
- NANNICINI, T. A Simulation-Based Sensitivity Analysis for Matching Estimators. **The Stata Journal**, v. 7, n.3, p. 334-350, 2007.
- MANSKI, C. Anatomy of the Selection Problem. **The Journal of Human Resource**, v. 24, p. 343-360, 1989.
- _____. Nonparametric Bounds on Treatment Effects. **American Economic Review Papers and Proceedings**, v. 80, p. 319-323, 1990.
- _____. Identification of Endogenous Social Effects: The Reflection problem. **Review of Economic Studies**, v. 60, p 531-542, 1990.
- _____. Monotone Treatment Response. **Econometrica**, v. 65, p. 1311-1334, 1997.
- _____. **Identification for Prediction and Decision**. Princeton University Press, Princeton, 2008.
- MANSKI, C.; PEPPER, J. V. Monotone Instrumental Variable: With an Application to the Returns to Schooling. **Econometrica**, v. 68, p. 997-1010, 2000.
- NERI, M. Motivos da Evasão Escolar no Brasil. Disponível em: <http://www.ufgd.edu.br/faed/nefope/publicacoes/pesquisa-motivos-da-evasao-escolar>. Acesso em: 21 out. 2010.
- ROSENBAUM, P.R. **Observational Studies**. Springer, New York, 2002.
- ROSENBAUM, P.R.; and RUBIN, D. The central role of the propensity score in observational studies for causal effects. **Biometrika**, v. 70, n. 1, p. 41–50, 1983a.
- _____. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. **Journal of the Royal Statistical Society, Series B**, v. 45, p. 212–218, 1983b.
- SMITH, J.; TODD, P. Does matching overcome Lalonde’s critique of nonexperimental estimators? **Journal of Econometrics**, v. 125, n. 1-2, p. 305-353, 2005.
- TABER, C.; FRENCH, E.. **Identification of Models of the Labor Market**. Federal Reserve Bank of Chicago, 2010. Disponível no site: http://www.chicagofed.org/digital_assets/publications/working_papers/2010/
- TODD, P. Evaluating Social Programs with Endogenous Program Placement and Selection of the Treated. In: **Handbook of Development Economics**, v. 60, n. 4, p. 3847-389, 2008.
- WEBBINK, D. Causal Effects in Education. **Journal of Economic Surveys**, v. 19, p. 535–560, 2005.
- ZOGHBI, A. C.; MENEZES, R. T; FELÍCIO, F.. Produtividade Relativa dos Setores Público e Privado em Educação: Impacto sobre a Escolha da Escola pela Família. In: **XXXVIII Encontro Nacional de Economia – ANPEC**, 2010.

ANEXOTabela A - Resultados para as *dummies* de estados da Regressão Logística para determinar o Escore de Propensão de estar matriculado na rede privada de ensino

Variáveis	Exames de Proficiência			
	Português 5º ano		Matemática 5º ano	
	Coefficiente	Erro-Padrão	Coefficiente	Erro-Padrão
Acre	-0,880	0,149	-0,879	0,150
Amazonas	-0,477	0,117	-0,408	0,116
Roraima	-1,496	0,174	-1,463	0,170
Pará	0,239	0,107	0,450	0,109
Amapá	-1,534	0,145	-1,525	0,142
Tocantins	-1,041	0,116	-1,030	0,114
Maranhão	0,078	0,101	-0,016	0,100
Piauí	0,559	0,105	0,465	0,103
Ceará	0,394	0,104	0,499	0,103
24 Rio Grande do Norte	0,195	0,104	0,084	0,102
25 Paraíba	0,338	0,110	0,227	0,108
26 Pernambuco	0,080	0,108	0,123	0,108
27 Alagoas	0,165	0,105	0,072	0,105
28 Sergipe	0,512	0,107	0,344	0,104
29 Bahia	0,391	0,105	0,367	0,103
31 Minas Gerais	-0,320	0,096	-0,228	0,096
32 Espírito Santo	-0,790	0,112	-0,826	0,110
33 Rio de Janeiro	-0,137	0,107	-0,157	0,107
35 São Paulo	-0,773	0,103	-0,762	0,104
41 Paraná	-0,603	0,101	-0,587	0,102
42 Santa Catarina	-1,072	0,113	-1,285	0,113
43 Rio Grande do Sul	-0,102	0,102	-0,234	0,103
50 Mato Grosso do Sul	-0,281	0,105	-0,441	0,105
51 Mato Grosso	-1,307	0,117	-1,200	0,117
52 Goiás	-0,030	0,099	-0,028	0,099

Nota: O Estado de Rondônia foi excluído intencionalmente para servir como parâmetro e o Distrito Federal foi excluído devido o pequeno número de observações.