

IDENTIFICAÇÃO PARCIAL DO EFEITO DAS ESCOLAS PRIVADAS BRASILEIRAS¹

Leandro Costa²

Ronaldo Arraes³

Esse estudo tem o objetivo de analisar a diferença no desempenho dos alunos de escolas brasileiras públicas e privadas, com base no questionamento das suposições que fundamentam as técnicas usuais de estimação. Devido aos fatores socioeconômicos interferirem significativamente na escolha dos pais sobre a escola dos seus filhos (viés de seleção), questiona-se os resultados dos estudos que consideram as suposições de identificação da metodologia de pareamento no escore de propensão (PEP) e regressões lineares na investigação da diferença de resultados entre as escolas públicas e privadas. A metodologia de identificação parcial (Manski, 1989) permite estimar limites com suposições menos restritivas e verificar se as estimações pontuais das metodologias usuais não ultrapassam os limites do intervalo. Utilizando as informações do Sistema de Avaliação da Educação Básica (SAEB) de 2005 (INEP, 2007) para os estudantes do 5º e 9º anos do ensino fundamental, comprova-se que a aplicação das metodologias usualmente utilizadas, que se baseiam nas suposições de ignorabilidade e imputação, sobreestima o efeito das escolas privadas. Todavia, ao utilizar hipóteses menos restritivas, esse efeito permanece significativamente positivo. Esses resultados devem advir do forte viés de seleção relacionado à restrição do nível de renda dos pais, que contribui para explicar a escolha da escola do filho e dificulta a determinação de um aceitável contrafactual.

Palavras-chave: identificação parcial, educação privada, viés de seleção.

JEL: C14; C52; I21

1 INTRODUÇÃO

Têm-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). Em paralelo a esse resultado há a expectativa de uma família representativa optar por matricular seus filhos em um sistema de ensino mais eficiente, supostamente o privado, desde que atenda a determinadas restrições

1. Os autores agradecem os comentários de Marcos Vera-Hernandez. Leandro Costa agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de doutorado e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) pela bolsa de estágio de doutorado no exterior. Ronaldo Arraes agradece ao CNPq pelo apoio financeiro à pesquisa. Os erros remanescentes são de responsabilidade dos autores.

2. Doutor em economia pela Universidade Federal do Ceará (UFC) e economista do Banco Mundial. *Email:* lcosta@worldbank.org

3. Professor-associado do curso de Pós-Graduação em Economia (CAEN) da UFC. *Email:* ronald@ufc.br

orçamentárias; caso contrário, a matrícula ocorreria na escola pública.⁴ 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 maioria das famílias brasileiras apresenta perfis socioeconômicos desfavoráveis e só tem a escola pública como única opção.⁵ A existência do viés de seleção relacionado às condições socioeconômicas das famílias pode tornar incomparáveis os desempenhos gerados nas duas redes de ensino, impossibilitando isolar o efeito da eficiência das escolas privadas em obter melhores performances dos alunos.

As suposições que possibilitam determinar um contrafactual são questionadas quando se busca determinar o efeito médio no desempenho dos alunos de frequentar uma determinada rede de escolas, contornado o viés de seleção e, conseqüentemente, a endogeneidade intrínseca ao processo educacional. No Brasil, pressupõe-se que os rendimentos familiares, a capacidade cognitiva dos alunos e a motivação das famílias em relação à importância dada à educação como fator de mobilidade social têm um papel significativo no processo de escolha da escola, conseqüentemente, tornando-se infactível sustentar os pressupostos de que as escolas de uma rede são homogêneas (suposição de homogeneidade) e que as variáveis relacionadas à escolha da escola não estão relacionadas ao desempenho dos alunos (suposição de ignorabilidade do tratamento). Como reconhece Webbink (2005), devido ao 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 busca comparar estudantes dos sistemas público e privado de ensino (Morgan e Winship, 2007). Daí a necessidade de se introduzir ajustes por meio de variáveis instrumentais para corrigir tal problema.

4. 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 no papel da qualidade e da quantidade de escolas. Os resultados apontam que o baixo nível de habilidades cognitivas alcançado pelos países latino-americanos pode responder pelo seu fraco desempenho no crescimento desde 1960 e que o desempenho nos exames de proficiência pode explicar as diferenças de crescimento inter e intrarregional.

5. Essa diferença nos resultados se agravou na última década, pois a universalização do Ensino Fundamental, incentivado por programas de transferência de renda como o Programa de Erradicação do Trabalho Infantil (Peti) e o Programa Bolsa Família (PBF), permitiu que crianças de famílias com baixos níveis socioeconômicos tivessem mais acesso às escolas públicas.

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 de correção. Entretanto, Altonji, Elder e Taber (2005) criticam esses instrumentos utilizados para identificar o efeito das escolas católicas americanas, pois normalmente os instrumentos apresentam alguma correlação com os resultados ou com as variáveis explicativas dos desempenhos dos alunos. A ocorrência disso acarretaria sérios danos às inferências realizadas.

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, os resultados apontam a gestão privada como a mais eficiente (Vandenbergh e Robin, 2004; Somers, McEwan e Willms, 2004; França e Gonçalves, 2009; Dronkers e Avram, 2010).

Em termos gerais, a literatura nacional e a internacional identificam a endogeneidade decorrente de variáveis não observadas e o viés de seleção como os principais problemas de análise, razão pela qual se tem buscado abordagens metodológicas alternativas, tais como variáveis instrumentais, pareamento no escore de propensão (PEP) 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. Visando contribuir com o debate, este estudo se distingue por sua abordagem e conclusões diferenciadas de trabalhos aplicados no Brasil.

Nesse contexto, pretende-se questionar as suposições de identificação da metodologia de PEP e de regressões lineares na investigação sobre a diferença de resultados entre as escolas públicas e privadas brasileiras. Conseqüentemente, a suposição de ignorabilidade do tratamento e imputação dessas metodologias é questionada para a realidade do mecanismo de seleção das escolas privadas. Para isso, será aplicada a metodologia de identificação parcial, desenvolvida em Manski (1989, 1990a, 1990b, 1997,

2008), com o objetivo de inferir limites para as estimativas do efeito médio do aluno estar matriculado em uma escola privada sob suposições menos restritivas e mais críveis, embora menos informativas.

Em vista destes questionamentos, propõe-se, neste artigo, aplicar diferentes ferramentas no conjunto de dados disponíveis que melhor se ajustem 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 estrutura em mais quatro seções, iniciando com uma revisão da literatura. Nas duas seções seguintes, descrevem-se a base de dados e a metodologia de identificação parcial. Os resultados encontram-se na seção cinco, seguidos pelas considerações finais.

2 REVISÃO BIBLIOGRÁFICA

Esta seção visa esclarecer as abordagens feitas sobre o assunto por meio dos trabalhos mais relevantes da literatura em conexão com o presente, que tratam do efeito do sistema privado de ensino sobre o desempenho dos estudantes com distintas abordagens, as quais incluem alternativas de correções de problemas amostrais. Nesse sentido, apresentam-se as estratégias de identificação desse efeito, focando essencialmente as metodologias de estimação empregadas, a abrangência dos dados e os principais resultados encontrados, sem a rigidez da ordem cronológica. Os desafios com que se depararam os pesquisadores centram basicamente na escolha de metodologias julgadas adequadas para estimar os efeitos de causalidade em educação, *vis-à-vis* os problemas inerentes nas bases de dados.⁶

Destaca-se inicialmente Webbink (2005), que apresenta uma síntese da recente literatura sobre os efeitos causais das políticas educacionais no desempenho dos estudantes. Ao considerar o processo educacional como endógeno, o autor alerta sobre a adequação de algumas metodologias para

6. Com o objetivo de abordar os avanços metodológicos, não serão apresentados os clássicos artigos como Friedman (1955) e Coleman, Hoffer e Kilgore (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, teria maior efeito sobre o desempenho dos estudantes de menor nível socioeconômico e das minorias.

se determinar os efeitos de causalidade em educação. Esse debate lança dúvida sobre inúmeros resultados encontrados na literatura, oriundos da função de produção educacional. Webbinck afirma que o principal problema das pesquisas educacionais é a existência de inúmeros fatores que podem mascarar o efeito da intervenção estudada, dificultando a obtenção de resultados convincentes quando não se isolam adequadamente os resultados de uma intervenção do efeito de outros fatores não observados pelo pesquisador. Para isso, novas metodologias têm sido aperfeiçoadas por meio da utilização de variações exógenas produzidas por experimentos controlados ou os ditos naturais, os quais são a forma mais aceitável de se encontrar variações exógenas ou variáveis instrumentais para produzir ambientes comparáveis aos experimentos.

Com o intuito de extrair o viés de estimação, Hoxby (1994) introduz procedimentos com variáveis instrumentais ao investigar se a competição entre escolas públicas e privadas americanas, e entre estas últimas, melhora a qualidade da educação. Para tanto, utiliza a composição religiosa das escolas privadas como instrumento, o qual representa variações exógenas sobre sua disponibilidade e seus custos. Com base em análise para escolas públicas, conclui haver uma relação direta entre competitividade e qualidade (desempenho educacional), e entre salários dos professores e taxas de conclusão dos estudantes do ensino médio.

Nessa mesma linha, Evans e Schwab (1995) detectaram problemas causados por endogeneidade dos fatores explicativos e a necessidade de eliminar o viés nas estimações, fato que os motivou a aperfeiçoar a metodologia de Coleman, Hoffer e Kilgore (1982) para analisar o efeito sobre a probabilidade de um aluno ingressar na universidade por meio da diferença de qualidade de ensino entre as escolas públicas e católicas dos Estados Unidos. Para remediar o problema gerado pelo viés de seleção, estimam um modelo *biprobit*, tendo como instrumento a vertente católica da família como fonte exógena de variação para identificar o efeito das escolas católicas. Concluiu-se que os estudantes do segundo grau de escolas católicas têm probabilidade 13% maior de ingressar na universidade.

Morgan (2001) inova ao introduzir a metodologia de PEP como uma estratégia alternativa aos modelos de regressão direcionados às relações de causalidade nos estudos dos efeitos das escolas. Reconhecendo a falta de robustez nas suposições da metodologia de pareamento, os resultados encontrados

revelam que o efeito da escola católica americana é mais forte sobre os estudantes menos prováveis a frequentá-la, condicionado às suas características observáveis. Muito embora sejam reconhecidos os danos causados pelo viés de variável omitida, o autor realça que raramente se discute a fragilidade dos dados observados, principalmente quando o latente viés de seleção é mais significativo. Essa é a razão posta para que a metodologia PEP seja uma alternativa plausível para estimativas do efeito de escolas no contexto de causalidade e avaliação de efeitos do tratamento.

Vandenberghe e Robin (2004) utilizam dados do Pisa de 2000 para estimar o efeito da educação privada e pública sobre o desempenho dos alunos em uma amostra de países-membros da Organização para Cooperação e Desenvolvimento Econômico (OCDE). Ao considerar o potencial viés gerado pela existência de fatores não observados, os autores constata importantes divergências entre os resultados dos testes paramétricos e não paramétricos ao aplicarem três métodos de estimação: variáveis instrumentais (VI), a abordagem de Heckman de dois estágios e o PEP. Os resultados revelam que a educação privada não gera benefícios sistemáticos para a maioria dos países, além de sugerir que se interprete o resultado a partir de características culturais, e não de eficiência das escolas privadas.

Com o intuito de inferir sobre a validade dos instrumentos amplamente utilizados na literatura como fontes de variação, Altonji, Elder e Taber (2005) aplicam análise de sensibilidade para identificar o efeito da escola católica americana sobre várias proposições formuladas. Os referidos autores mostram que as tentativas para corrigir o viés de seleção por intermédio da metodologia de VI não são factíveis 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. Usando diversos métodos para testar o viés induzido por estes instrumentos, os autores 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. Nessa mesma linha de pesquisa, Nguyen, Taylor e Bradley (2006) adotam metodologias diferenciadas – PEP e diferença em diferença – para controlar o viés de seleção, devido a características

observadas e não observadas, cujo resultado revela um efeito menor da escola católica americana, comparado ao obtido por Altonji, Elder e Taber (2005).

Com aplicação para países da América Latina, Somers, McEwan e Willms (2004) comparam a efetividade das escolas privadas, inserindo contextos institucionais e níveis de renda diferenciados dos encontrados em estudos que avaliam o efeito das escolas católicas dos Estados Unidos. Utilizando a metodologia de regressão de multiníveis, os autores abordam a caracterização do efeito dos pares, que pode tornar o real impacto da eficiência das escolas privadas viesado, e destacam que os efeitos observados para o Brasil situam-se entre os maiores.

Com base em dados do SAEB 2003, França e Gonçalves (2009) analisam a diferença entre os resultados do sistema da educação pública e privada brasileira, utilizando a metodologia PEP 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 do ensino fundamental denotam 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 alunos.⁷ Vale apontar, todavia, que essa estratégia possivelmente mascara a comparabilidade 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) fazem uma análise comparativa entre países empregando a metodologia PEP visando avaliar a diferença da efetividade escolar entre os sistemas público e privado independente.⁸ Segundo os autores, este aspecto torna os resultados significativamente diferentes do sistema privado dependente, 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

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

8. Dronkers e Avram (2010) fazem essa mesma análise para escolas privadas dependentes, isto é, financiadas total ou parcialmente por recursos públicos. Ademais, Dronkers e Robert (2008) abordam as mesmas questões utilizando a metodologia de regressões em multiníveis.

de uma escola bem equipada e de boa qualidade. Concluem que, na maioria dos países, as performances mais elevadas nos exames de leitura favorecem os alunos matriculados em escolas privadas independentes, comparados aos de escolas públicas. Particularmente para o Brasil, essa diferença em favor de escolas privadas é significativamente maior.

Curi e Menezes-Filho (2010) argumentam que a escolha da rede de ensino pelas famílias é resultado de uma combinação de fatores que tanto afetam a demanda das famílias pela escola quanto essa demanda é afetada pela oferta do número de vagas disponibilizadas na rede pública. Aplicando regressões de variáveis binárias aos dados da Pesquisa Nacional por Amostra de Domicílios, do Instituto Brasileiro de Geografia e Estatística (PNAD/IBGE), e da Pesquisa de Orçamentos Familiares (POF), esses autores extraem dos resultados que a educação da mãe, a renda familiar, a oferta de escolas públicas, o custo da educação e a região de moradia são os principais fatores que influenciam a decisão dos pais na escolha por escolas privadas, a qual decorre da baixa qualidade de ensino na rede pública e de sua limitada oferta de vagas.⁹

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: modelo de variáveis instrumentais, PEP e de diferença em diferença. Apesar da expressiva quantidade de pesquisas realizadas, Dronkers e Avram (2010) enfatizam que 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

A implementação metodológica para verificação empírica decorre da aplicação dos microdados do SAEB de 2005 (INEP, 2007), pesquisa conduzida pelo Ministério da Educação (MEC), cuja avaliação é realizada bianualmente por amostragem das redes de ensino em cada Unidade da Federação (UF), 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 e português, com foco, respectivamente, na

9. Zoghbi, Menezes e Felício (2010) oferecem uma explicação adicional ao investigar se a produtividade relativa dos setores públicos e privados da educação é levada em conta no processo de escolha das escolas pelas famílias.

resolução de problemas e 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 repetidas seções transversais abrangendo representativas escolas e estudantes de escolas públicas e privadas.¹⁰

A amostra do SAEB selecionou 3.004 escolas, as quais proveram 160.929 alunos do 5º e 9º anos do ensino fundamental para a realização do exame nas disciplinas de matemática e português. Do total de alunos do 5º ano, 41.783 fizeram a prova de matemática e 42.146 a de língua portuguesa; e no 9º ano, 33.189 e 33.164, respectivamente. Desse total, 29,7% dos estudantes são provenientes 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 federados.¹¹ Em relação à dependência administrativa das escolas públicas, 35,4% são municipais e 34,5% são estaduais. Cabe salientar que foram descartados da amostra 0,65% de alunos das escolas públicas federais, uma vez que estas 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 filhos, pois as 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 e 9º anos do ensino fundamental.

10. Em Instituto Nacional de Estudos e Pesquisa (INEP, 2007) são apresentados os aspectos metodológicos para extração da amostra, os instrumentos utilizados na pesquisa, os exames e os questionários.

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

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 atributos.¹²

Os resultados mostram significativas vantagens no desempenho dos estudantes das escolas privadas sobre aqueles de escolas públicas, porquanto, entre todos os grupos, à exceção dos estudantes pretos do 9º ano, as diferenças são maiores que 20%. Em relação ao gênero, 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 autodeclarados pretos apresentam piores resultados em ambos os tipos de escolas.

TABELA 1

Resultados nos exames de proficiência do SAEB/2005 dos alunos do 5º e 9º anos do ensino fundamental

Exames	5º ano				9º ano			
	Todas	Escola pública	Escola privada	Diferença (%)	Todas	Escola pública	Escola privada	Diferença (%)
Matemática								
Homem	189,4	174,2	225,7	29,6	254,0	233,9	291,7	24,7
Mulher	187,6	173,4	220,2	27,0	244,5	225,3	284,2	26,2
Não preto	191,4	175,8	224,4	27,6	250,7	230,4	288,7	25,3
Preto	166,1	161,4	196,4	21,7	228,6	218,8	271,3	24,0
Português								
Homem	171,2	157,8	203,5	29,0	233,6	218,2	262,3	20,2
Mulher	185,8	172,6	216,0	25,2	246,1	229,4	280,5	22,3
Não preto	180,8	166,6	211,1	26,7	242,0	225,4	272,6	20,9
Preto	158,5	153,8	185,4	20,5	221,3	213,8	253,4	18,5

Fonte: Microdados do SAEB/2005.

Elaboração dos autores.

Os dados descritivos das variáveis sociais e educacionais dos estudantes e suas famílias, utilizadas 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, as variáveis *mulher* e *preto* pretendem controlar as características de gênero e cor ou raça da amostra.

12. Para dados descritivos que contemplam desigualdades regionais e socioeconômicas, ver INEP (2007).

As variáveis denotadas por distorção idade-série, reprovado e pré-escola caracterizam o histórico acadêmico dos estudantes. A escolaridade dos pais visa captar o efeito da herança educacional e do 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 de que os filhos dispõem.

Constata-se, nessa tabela, que, entre os estudantes que se consideram da raça preta, o percentual dos que frequentam escolas particulares é cerca de 50% menor relativo às escolas públicas para as amostras em ambos os exames e séries. O atributo gênero, por sua vez, mostra-se equilibrado entre os sexos para escola, exame e série.

A qualificação educacional dos pais entre os alunos de escolas públicas e privadas é acentuadamente díspar. Em ambas as séries, a proporção de pais sem o ensino fundamental dos alunos de escolas públicas é de quatro a sete vezes maior, conseqüentemente, no outro extremo de educação dos pais, os resultados entre as escolas são invertidos.

TABELA 2

Características individuais e socioeconômicas dos estudantes do 5º e 9º anos do ensino fundamental
(Em %)

Características	5º ano				9º ano			
	Matemática		Português		Matemática		Português	
	Escola pública	Escola privada						
Mulher	48,5	49,5	48,4	49,5	53,3	50,8	54,4	51,9
Preto	14,0	5,0	13,6	5,5	9,6	4,3	9,7	4,5
Família completa	92,1	96,1	90,7	95,6	84,9	91,1	84,5	91,2
Mãe sem EF ¹	34,3	7,2	33,9	6,8	37,7	5,1	38,2	4,6
Pai sem EF	28,2	7,5	27,6	7,0	33,0	5,7	32,4	5,5
Mãe graduada	7,8	37,6	7,5	37,4	5,1	39,7	4,9	39,6
Pai graduado	8,5	35,6	7,9	35,5	4,4	36,4	4,3	37,0
Pré-escola	69,6	91,4	67,8	90,9	29,7	67,0	30,3	68,2
Computador com internet	10,1	51,1	10,1	50,5	12,8	62,0	12,9	62,7
Doméstica	11,3	50,3	11,1	50,4	11,9	57,2	11,8	57,7

Fonte: Microdados do SAEB/2005.

Elaboração dos autores.

Nota: ¹ EF significa ensino fundamental.

Vale 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 consta essa informação, o processo de pareamento dos estudantes pode ser limitado e tornar enviesado a estimação do efeito médio. Visando minimizar esse viés, consideram-se duas *proxies* de renda domiciliar, denotadas por “ter empregada doméstica no domicílio” e “ter um computador com internet”, as quais são julgadas apropriadas para se inferir situações econômicas dos estudantes de escolas públicas. As estatísticas revelam que aproximadamente 51% dos domicílios de alunos do 5º ano das escolas privadas possuíam computador com internet, enquanto nos domicílios de alunos de escolas públicas, apenas 10,1%; para alunos do 9º ano, esse diferencial é acrescido em aproximadamente 10 pontos percentuais (p.p.). Quanto a ter doméstica no domicílio, verifica-se o mesmo padrão de diferencial em favor dos domicílios de alunos de escolas privadas. Também se constatam divergências acentuadas nos níveis de estudo dos pais, pois, conforme a tabela 2, mais de 35% dos pais dos estudantes do 5º ano de escolas privadas concluíram o ensino superior, contra menos de 8,5% dos pais de alunos de escolas públicas. Para alunos do 9º ano, a amplitude dessa diferença é ainda maior.

A partir da análise da escolaridade dos pais, não é possível identificar se aqueles 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 de alunos que influenciam o aprendizado e a escolha das escolas. Contudo, 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 o desempenho dos estudantes.

4 METODOLOGIA

A estratégia metodológica que se segue avalia, em primeira instância, a determinação do efeito de estudantes brasileiros estarem matriculados em escolas privadas em relação aos que estão matriculados em escolas públicas,

sobre o desempenho em exames de proficiência, com base no modelo de PEP introduzido por Rosenbaum e Rubin (1983) e amplamente empregado na literatura. Esse modelo se baseia na identificação pontual do efeito médio do tratamento (EMT), cujo procedimento é conduzido sob fortes suposições. Em contraposição, conduz-se um procedimento alternativo menos restritivo para a intervenção analisada, por meio da metodologia de identificação parcial do EMT (Manski, 1989), o qual é inferido por limites, e não pontualmente, como as demais metodologias.

Amplamente discutida em Heckman, Ichimura e Todd (1998), Imbens (2004), Blundell e Costa-Dias (2009) e Caliendo e Kopeinig (2008), a aplicação da metodologia PEP determina um contrafactual a partir da determinação de indivíduos dentro do grupo de controle que sejam semelhantes aos tratados, dadas as variáveis observáveis. Essa é uma técnica semiparamétrica de determinação dos efeitos de uma intervenção 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-tratamento ou ignorabilidade do tratamento.¹³

A suposição fundamental de identificação do efeito médio do tratamento da metodologia PEP pode ser representada da seguinte forma:¹⁴

$$D \perp \frac{[Y(1), Y(0)]}{X} \text{ ou } E \left[\frac{Y(0)}{D} = 1, X \right] = E \left[\frac{Y(0)}{D} = 0, X \right] D \perp (Y(1), Y(0)) | X$$

Essa suposição implica que, dadas as variáveis observáveis X , o tratamento D é independente dos resultados potenciais, $Y(0)$ e $Y(1)$. Para facilitar a análise de modelos lineares, a segunda expressão reflete que a SIC significa que a esperança condicional do resultado potencial $Y(0)$, em relação às variáveis observáveis X , é independente do *status* de tratamento D . Estas suposições são questionáveis, pois pressupõem que, além das variáveis observadas X , não há características não observadas do indivíduo associadas aos resultados potenciais e ao tratamento.

13. Imbens (2004) apresenta alguns modelos econômicos que implicam a SIC, nas quais os agentes escolhem participar de um programa se os benefícios, igual à diferença dos resultados potenciais, excedem os custos associados à participação. Esta suposição está implícita na independência dos custos e benefícios, condicionada às variáveis observadas.

14. Essa forma de mostrar a condição fundamental de identificação foi determinada por Heckman, Ichimura e Todd (1998), visando incorporar, numa estrutura econométrica, a separabilidade aditiva e as restrições de exclusão.

A identificação parcial, ou limites, surgiu como uma metodologia alternativa às tradicionais análises de inferência, notadamente para buscar corrigir as supostas superestimações dos efeitos decorrentes destas. Com isso, realça sua adequação para se analisar a diferença de qualidade das escolas públicas e privadas quanto à suposição de independência condicional e de balanceamento, pois, possivelmente, o viés de seleção, devido às condições econômicas das famílias e à 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 dessa metodologia são Manski (1989, 1990a, 1990b, 1997, 2008) e Manski e Pepper (2000),¹⁵ cuja técnica não paramétrica é também baseada em suposições, embora menos restritivas, para representar a regra de seleção ou o efeito do tratamento. As principais suposições são a resposta monotônica do tratamento (RMT), a seleção monotônica do tratamento (SMT) e a variável instrumental monotônica (VIM).

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 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 semiparamétricos não serem teoricamente sustentáveis ou aceitáveis. Nesse sentido, uma estratégia seria não apenas obter a identificação pontual do parâmetro mas, sim, um intervalo de possíveis valores do parâmetro, podendo, alguns dos quais, serem excluídos *a priori* com credibilidade. Imbens e Woodridge (2009) enfatizam os recentes avanços da metodologia de identificação parcial que, mesmo alicerçada em suposições, embora menos fortes do que as do PEP, tornam possível restringir o intervalo e obter resultados mais informativos. A utilização de algumas variáveis explicativas aumenta a possibilidade de adicionar suposições que podem estreitar os limites do intervalo. Entretanto, os autores expõem que o método é criticado devido à possibilidade teórica de a identificação de um determinado intervalo ser considerada sem utilidade ou de ser informativo apenas para grandes amostras.

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

Com o objetivo de determinar limites para o EMT, define-se, para cada aluno i , 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 $T = [0; 1]$, em que 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 omitido.

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

$$D(0,1) = E[y(1) | t = 1] - E[y(0) | t = 0] \quad (1)$$

onde $D(0,1)$ é o efeito médio do tratamento e $E[y(t)]$ é o valor esperado da performance do estudante com o tratamento t .

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 e $P(t = 1 | x)$ é a probabilidade de o aluno estar na escola privada.

A inviabilidade no cálculo desta média reside no desconhecimento sobre o desempenho médio caso o aluno de escola privada estivesse frequentando a escola pública ou vice-versa, quais sejam, $E\left[\frac{y(1)}{x}, t = 0\right]$ e $E\left[\frac{y(0)}{x}, t = 1\right]$, respectivamente. Essa desinformação é o fato gerador do viés de seleção na abordagem PEP, cuja magnitude é obtida estendendo-se a equação (1), com a adição e subtração do resultado esperado, para não participantes do tratamento caso participassem, ou outra forma de se obter o contrafactual, qual seja:

$$\Delta(0,1) = E[y(1) | t = 1] - E[y(0) | t = 0] + \{E[y(0) | t = 1] - E[y(0) | t = 0]\} \quad (1')$$

$$\Delta(0,1) = D(0,1) + \{B(0,1)\}$$

O termo B , $\{E[y(0)|t=1] - E[y(0)|t=0]\}$ é a magnitude do viés de seleção que surge ao se inferir $D(\cdot)$ como uma estimativa do efeito médio do tratamento. Desde que se desconhece $E[y(0)|t=1]$, torna-se impossível o cálculo dessa magnitude. Por conseguinte, não se sabe a extensão desse viés na composição de $D(\cdot)$ e, conseqüentemente, não se pode computar a diferença exata nos resultados entre os tratados e o grupo de controle. Portanto, o objetivo básico para se contornar esse problema é encontrar alternativas para eliminar ou reduzir o viés de seleção por meio da formulação de hipóteses 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 o desempenho dos alunos. Substituindo $E\left[\frac{y(1)}{x}, t=0\right]$ e $E\left[\frac{y(0)}{x}, t=1\right]$ pelo menor (K_0) e maior (K_1) níveis possíveis de desempenho determinam-se os limites inferiores e superiores de $E[y(t)]$, gerando, assim, os limites sem suposição:

$$\begin{aligned} E(y|x, z=t)P(z=t|x) + K_0P(z \neq t|x) &\leq \\ \leq E(y(t)|x) &\leq E(y|x, z=t)P(z=t|x) + K_1P(z \neq t|x) \end{aligned} \quad (3)$$

Para estreitar estes limites, pode-se adicionar a suposição de RMT e a suposição de 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 está relacionado à hipótese de que o fato de o aluno estar frequentando uma escola privada nunca diminui seu desempenho, pois, teoricamente, um sistema de ensino mais eficiente conduz a um desempenho médio mais elevado dos alunos, embora um efeito zero não seja excluído

com esta suposição.¹⁶ Entretanto, pode-se criticar essa abordagem em relação à interpretação dos seus resultados, pois ao tentar aferir o efeito médio da mudança de rede de ensino, pode não fazer sentido partir da suposição de que a mudança melhora o rendimento do aluno. Isso, normalmente, é algo a ser estimado e não adotado como suposição.

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

$$u_2 \geq u_1 \Rightarrow E[(y(t) | z = u_2)] \geq E[(y(t) | z = u_1)] \quad (5)$$

Esta suposição baseia-se na possibilidade 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 a melhoria do desempenho de seus estudantes. No entanto, tendo em vista que se espera que haja muitos outros critérios que os pais consideram importantes na escolha das escolas, além do desempenho dos alunos, os resultados devem ser interpretados levando em conta que essa suposição simplifica significativamente um complexo processo de escolha.

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. Utilizando-se 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 qual pertence mas também uma variável z^* , que caracterize indiretamente os alunos de escolas públicas e privadas. Em seguida, obtêm-se subamostras, uma para cada valor de z^* e, para cada uma delas, determinam-se

16. Os autores reconhecem que podem existir outros motivos que levam os pais (ou responsáveis) a escolher uma escola privada, em detrimento de uma escola pública, que não estejam necessariamente relacionados com o desempenho acadêmico e a renda. Por exemplo, os pais podem optar por uma escola pela linha pedagógica, a abordagem religiosa, a proximidade do domicílio, a infraestrutura entre outras razões. Para mais detalhes sobre a escolha dos pais em matricular seus filhos em escolas privadas, ver Curi e Menezes-Filho (2010).

os limites sem suposição com base na equação (3). 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 VIM apresentada em Manski e Pepper (2000), a qual é dada por:

$$m_1 \leq m \leq m_2 \Rightarrow E[y(t)|z^* = m_1] \leq E[y(t)|z^* = m] \leq E[y(t)|z^* = m_2] \quad (6)$$

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

Então, em vez 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 $E\left[\frac{y(t)}{z^*} = m\right]$ não é menor do que o limite inferior sem suposição sobre $E\left[\frac{y(t)}{z^*} = m_1\right]$ e não é maior do que o limite superior sem suposição sobre $E\left[\frac{y(t)}{z^*} = m_2\right]$. Para a subamostra em que z^* tem o valor m , pode-se derivar um novo limite inferior, que é o limite inferior absoluto sobre todas as subamostras, e 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:

$$\begin{aligned} & \sum_{m \in M} P(z^* = m) \cdot [\max_{m \in M} (E[y | z = t, z^* = m] \cdot P(z = t, z^* = m) + \\ & + K_0 \cdot P(z \neq t | z^* = m))] \leq E[y(t)] \leq \sum_{m \in M} P(z^* = m) \cdot \\ & \cdot [\min_{m \in M} (E[y | z = t, z^* = m] \cdot P(z = t, z^* = m) + K_1 \cdot P(z \neq t | z^* = m))] \quad (7) \end{aligned}$$

Em bases teóricas, a VIM mais adequada é a renda familiar do aluno. Entretanto, na indisponibilidade desta, optou-se por uma *proxy* denotada

pela posse dos pais de pelo menos um automóvel.¹⁷ Certamente não seria possível utilizar esta *proxy* como uma variável instrumental convencional, uma vez que 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 EMT, $\Delta(0,1) = E[y(1)] - E[y(0)]$. A ocorrência de $\Delta(0,1) = 0$ justificaria a suposição de resposta monotônica ao tratamento.

5 RESULTADOS

Em vista de as escolhas dos pais sobre a escola dos filhos seguir restrições relacionadas com a renda familiar e diante da 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 do efeito causal do sistema privado de ensino sobre o rendimento dos alunos. Pelo pareamento de casos similares do grupo de tratamento e controle, o método PEP visa eliminar o viés de seleção decorrente das variáveis observáveis e, conseqüentemente, parear indivíduos no grupo de tratamento e controle que imitem os indivíduos de um experimento aleatório.

Então, conforme o procedimento do método, uma vez verificada a assimetria da distribuição dos escores de propensão, aplica-se o algoritmo de Pareamento Linear Local (PLL), descrito em Caliendo e Kopeinig (2008), com combinações que usam 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 de o aluno estar matriculado na rede privada. As probabilidades foram estimadas a partir de uma regressão de escolha discreta binária sob a hipótese *logit* para o

17. A escolha desta *proxy* como medida de riqueza é apoiada em dados da PNAD (IBGE, 2005), uma vez que a renda média das famílias possuidoras de pelo menos um carro excede a das famílias destituídas desse bem em 207%. Além disso, considerando as distribuições acima da mediana, a renda média das famílias com carro excede a das famílias sem carro em 225%; e nas distribuições abaixo da mediana, essa supremacia é de 154%.

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 à pré-escola e identificação dos estados.¹⁸ As estimativas das regressões referentes aos escores de propensão para as amostras de estudantes do 5º e 9º anos que realizaram os exames de língua portuguesa e matemática estão dispostas na tabela 3.

TABELA 3

Estimativas das regressões logísticas para os alunos do 5º e 9º anos que realizaram os exames de matemática e português

Variáveis	5º ano				9º ano			
	Matemática		Português		Matemática		Português	
	Coefficiente	EP ¹	Coefficiente	EP ¹	Coefficiente	EP ¹	Coefficiente	EP ¹
Mulher	0,035 ²	0,027	0,019 ²	0,027	0,099	0,032	0,075	0,033
Preto	-0,760	0,052	-0,653	0,051	-0,574	0,067	-0,473	0,066
Família completa	0,486	0,061	0,479	0,057	0,380	0,050	0,435	0,050
Mãe sem EF	-1,051	0,042	-1,097	0,043	-1,194	0,052	-1,272	0,054
Pai sem EF	-0,660	0,043	-0,706	0,044	-0,858	0,052	-0,802	0,053
Mãe graduada	0,683	0,039	0,675	0,039	0,998	0,047	1,018	0,047
Pai graduado	0,489	0,039	0,584	0,039	0,981	0,050	1,079	0,050
Pré-escola	1,054	0,040	1,034	0,039	0,715	0,033	0,729	0,034
Computador com internet	1,509	0,034	1,411	0,034	1,486	0,038	1,504	0,038
Doméstica	1,265	0,032	1,270	0,032	1,330	0,036	1,346	0,037
Constante	-3,094	0,103	-3,145	0,100	-3,105	0,146	-3,149	0,143
N ³	41.519		41.883		33.026		33.000	
Log-likelihood	16.911,9		17.003,6		17.324,7		17.782,8	
Pseudo-R ²	0,334		0,333		0,411		0,422	

Elaboração dos autores.

Notas: ¹ EP denota erro-padrão.

² Não significante a 1%.

³ Exclui os alunos de escolas públicas federais.

Os resultados das regressões apresentados na tabela 3 atestam que os coeficientes de todas as variáveis explicativas são estatisticamente significantes ao nível de, no máximo, 1%, e os sinais de cada um deles coincidem em ambos os exames e anos. Entretanto, comparando os efeitos decorrentes das magnitudes dos coeficientes de cada variável para o 5º e o 9º anos, sejam positivos ou

18. Heckman, Ichimura e Todd (1998), Heckman, Lalonde e Smith (1999) e Caliendo e Kopeinig (2008) discutem sobre a escolha das variáveis explicativas que devem ser inseridas no cálculo do escore de propensão. Embora não haja uma regra sobre o número de variáveis a serem inseridas, pressupõe-se que elas devem estar relacionadas tanto com o tratamento quanto com o resultado e, principalmente, devem melhorar o balanceamento entre o grupo de controle e de tratamento.

negativos, percebe-se que não há um padrão definido, indistintamente do exame. Considerando pai ou mãe graduados, o efeito positivo é mais forte para os alunos do 9º ano, enquanto, com relação à pré-escola, o efeito é mais acentuado para os alunos do 5º ano; para efeitos negativos, a variável raça tem uma contribuição marginal maior nos alunos do 5º ano, enquanto pais com baixa escolaridade participam com maior efeito para os alunos do 9º ano. Realce-se que este último resultado indica que uma herança educacional familiar em que os pais possuem baixo nível de escolaridade provoca um efeito 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.

A variável se o aluno frequentou a pré-escola, por sua vez, capta o viés de seleção oriunda de variáveis relacionadas à renda, e seu sinal positivo indica maior probabilidade de um estudante estar em uma escola privada; o fato de esse efeito ser mais forte para alunos do 5º ano denota sua maior proximidade temporal em relação aos do 9º ano. O efeito renda familiar, expresso pelos sinais positivos das *proxies* computador com internet e doméstica, reflete a expectativa de direcionamento dos filhos para a rede de ensino particular. Adicionalmente, foram computados efeitos regionais por meio da inclusão de *dummies* dos estados, cujos coeficientes variam de magnitude e sinal (apêndice A). Isso pode estar refletindo as diferentes condições providas em cada estado em relação ao percentual de alunos matriculados na rede privada de ensino.

A metodologia de pareamento pondera a amostra do grupo de controle a fim de aumentar a semelhança com os indivíduos do grupo de tratamento, visando balancear as características observadas da amostra entre os grupos de tratamento e de controle após o pareamento. Uma análise de balanceamento das variáveis utilizadas na estimação do escore de propensão é mostrada na tabela 4, na qual se apresentam, para os alunos de ambas as séries, as médias dos escores de propensão para as variáveis explicativas, antes e depois do pareamento, e uma medida do viés entre a amostra do grupo de tratamento e do de controle.¹⁹

19. A medida de viés sugerida por Rosenbaum (2002), que utiliza a diferença das médias padronizadas das amostras

do grupo de tratamento e de controle, é dada por:
$$\text{Viés} = \frac{100(\bar{x}_T - \bar{x}_C)}{\sqrt{\frac{s_T^2 + s_C^2}{2}}}$$

onde \bar{x}_T e \bar{x}_C são as médias amostrais e s_T^2 e s_C^2 são as variâncias amostrais dos grupos de tratamento e de controle.

TABELA 4
Balanceamento das variáveis utilizadas no pareamento dos alunos do 5º e 9º anos que realizaram os exames de matemática e português

Variáveis	Amostra	5º ano			9º ano								
		Matemática		Língua Portuguesa		Matemática		Língua Portuguesa					
		Trat.	Cont.	Viés (%)	Trat.	Cont.	Viés (%)	Trat.	Cont.	Viés (%)			
Mulher	Não pareado	0,494	0,485	1,9	0,495	0,484	2,3	0,508	0,533	-4,9	0,519	0,544	-5,0
	Pareado	0,494	0,490	0,9	0,495	0,501	-1,1	0,508	0,489	3,8	0,519	0,485	7,0
Preto	Não pareado	0,050	0,140	-31,1	0,055	0,136	-27,8	0,043	0,096	-20,7	0,045	0,097	-20,5
	Pareado	0,050	0,047	1,2	0,055	0,047	2,7	0,043	0,040	1,4	0,045	0,041	1,6
Família completa	Não pareado	0,961	0,921	17,0	0,956	0,907	19,4	0,911	0,849	18,9	0,912	0,845	20,7
	Pareado	0,961	0,965	-1,5	0,956	0,962	-2,4	0,911	0,921	-3,3	0,912	0,922	-2,9
Mãe sem EF	Não pareado	0,071	0,343	-70,9	0,068	0,339	-71,6	0,051	0,377	-86,5	0,046	0,382	-89,7
	Pareado	0,071	0,069	0,8	0,068	0,064	0,9	0,051	0,045	1,8	0,046	0,041	1,6
Pai sem EF	Não pareado	0,075	0,282	-56,0	0,070	0,276	-56,7	0,057	0,330	-73,8	0,055	0,324	-73,1
	Pareado	0,075	0,072	0,8	0,070	0,071	-0,3	0,057	0,054	0,8	0,055	0,050	1,2
Mãe graduada	Não pareado	0,375	0,078	75,9	0,374	0,075	76,8	0,397	0,051	91,3	0,396	0,049	92,0
	Pareado	0,375	0,382	-1,6	0,374	0,368	1,5	0,397	0,400	-0,8	0,396	0,395	0,4
Pai graduado	Não pareado	0,355	0,085	69,2	0,355	0,079	70,8	0,364	0,044	86,4	0,370	0,043	88,3
	Pareado	0,355	0,351	1,3	0,355	0,348	1,6	0,364	0,355	2,3	0,370	0,365	1,4
Pré-escola	Não pareado	0,914	0,696	57,2	0,909	0,678	59,5	0,670	0,297	80,6	0,682	0,303	82,1
	Pareado	0,914	0,925	-2,8	0,909	0,917	-2,1	0,670	0,686	-3,4	0,682	0,685	-0,5
Computador com internet	Não pareado	0,510	0,101	99,4	0,505	0,101	97,9	0,620	0,128	118,4	0,627	0,129	119,7
	Pareado	0,510	0,502	2,0	0,505	0,503	0,6	0,620	0,624	-0,8	0,627	0,622	1,1
Doméstica	Não pareado	0,503	0,113	93,2	0,504	0,111	94,1	0,572	0,119	108,6	0,577	0,118	110,2
	Pareado	0,503	0,499	1,2	0,504	0,498	1,5	0,572	0,581	-2,0	0,577	0,591	-3,2
Escore de propensão	Não pareado	0,569	0,569	158,4	0,569	0,184	158,4	0,650	0,177	187,5	0,658	0,174	192,4
	Pareado	0,569	0,569	0,0	0,569	0,569	0,0	0,650	0,649	0,0	0,658	0,658	0,0

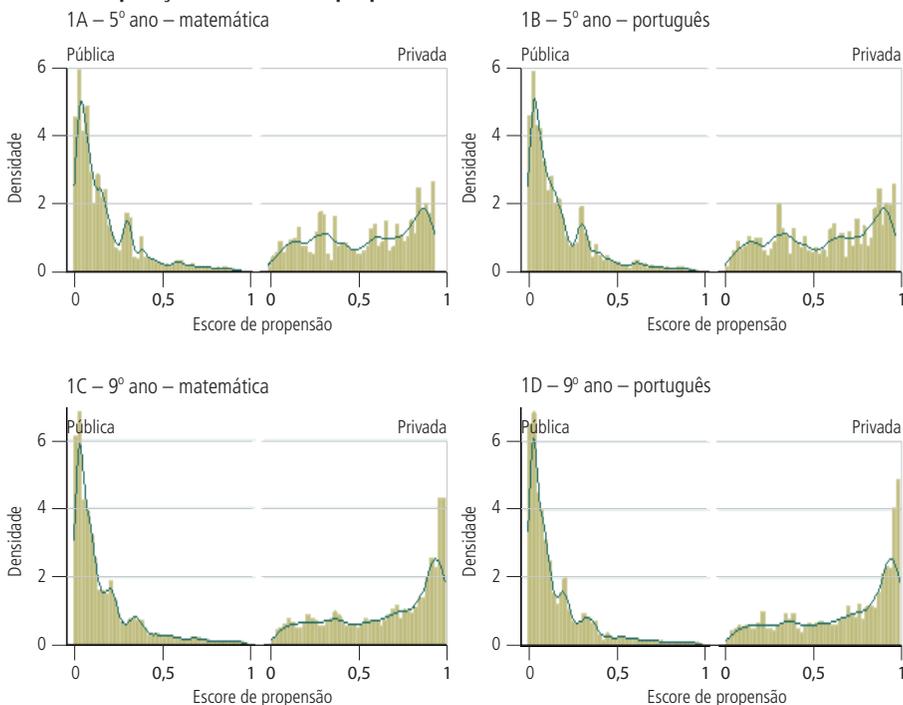
Elaboração dos autores.

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 e mais acentuada para os alunos do 9º ano. 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 de tratamento e do de controle que são utilizados no processo de estimação.

Outra importante fonte de identificação da sobreposição do escore de propensão são os histogramas dos escores de propensão para os indivíduos do grupo de controle e de 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. Ademais, verificam-se, em ambas as amostras dos alunos que participaram dos exames de língua portuguesa e matemática, que os escores de propensão dos alunos distribuem-se de forma distinta, evidenciando a diferença entre elas. Essa assimetria na distribuição dos escores de propensão confirma a necessidade de se aplicar o algoritmo PLL descrito anteriormente.

Uma vez que o método PEP reduz o viés atribuído às variáveis observadas, pode-se usar a diferença nos resultados médios encontrados nas amostras para se obter uma estimativa do EMT. A tabela 5 expõe as estimativas PEP obtidas via algoritmo PLL. Nas colunas “Grupos”, constam as médias condicionais dos exames de proficiências, antes e após o pareamento para o grupo de tratamento e de controle. A coluna denominada “Dif.” refere-se à diferença entre essas duas médias condicionais, ou seja, entre o viés de seleção e a estimativa do EMT (a diferença que contorna o viés de seleção com base na SIC). Na última coluna, estão dispostas as estimativas de mínimos quadrados ordinários (MQO), utilizando todas as variáveis do escore de propensão como variáveis explicativas.

GRÁFICO 1

Sobreposição do escore de propensão

Elaboração dos autores.

TABELA 5

Estimativas PEP e MQO: exames de português e matemática para o 5º e 9º anos

Resultado	Amostra	Grupos		Dif.	Teste t		MQO
		Tratamento	Controle		EP	t	
5º ano							
Matemática	Não pareados	0,706	-0,303	1,009	0,009	106,30	
	EMT	0,706	-0,116	0,822	0,016	44,50	0,722 ¹
Português	Não pareados	0,678	-0,289	0,967	0,010	101,02	
	EMT	0,678	-0,077	0,755	0,019	38,36	0,672 ¹
9º ano							
Matemática	Não pareados	0,701	-0,355	1,056	0,010	104,66	
	EMT	0,701	-0,083	0,784	0,030	25,82	0,679 ¹
Português	Não pareados	0,607	-0,309	0,916	0,010	87,29	
	EMT	0,607	-0,078	0,685	0,022	30,27	0,597 ¹

Elaboração dos autores.

Nota: ¹ Significante a 1%.

Os resultados apontam para significativos e positivos efeitos das escolas privadas em relação às escolas públicas, sobre o desempenho dos alunos nos exames de matemática e português. O efeito médio de o aluno do 5º ano estar matriculado em uma escola privada sobre o desempenho nos exames de matemática e português são 0,822% e 0,755% dos desvios-padrão, respectivamente, enquanto essa supremacia para alunos do 9º ano decai nos respectivos exames para 0,784% e 0,685%. Resultados similares são ratificados pela metodologia econométrica de MQO, embora com as magnitudes dos diferenciais um pouco menores em ambos os exames, ou seja, 0,722% e 0,672% para o 5º ano, e 0,679% e 0,597% para o 9º ano.²⁰

As estimativas da análise de identificação parcial para os alunos de ambos os exames e séries são apresentados nas tabelas 6 (matemática) e 7 (língua portuguesa). Um ponto prático na comparação destes resultados, como os dos métodos 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 estimativas.²¹ Na primeira, utiliza-se a estratégia contida no método PEP, já citada na seção anterior, enquanto na análise de identificação parcial, que utiliza o estimador de polinômio local de pesos kernel (*Kernel-weighted local polynomial*), essa barreira é mais rígida. Dessa forma, com o intuito de tornar comparáveis os resultados destas metodologias, nelas incluídas todas as variáveis explicativas no cálculo do escore de propensão, extrai-se o resíduo previsto da regressão por MQO da proficiência dos alunos sobre as mesmas variáveis explicativas utilizadas no escore de propensão.²²

20. Os resultados estão medidos, em relação ao desvio-padrão, pelo fato de as variáveis dependentes terem sido padronizadas, conforme descrito anteriormente.

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

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

TABELA 6

Matemática: análise de identificação parcial para alunos do 5º e 9º anos

Suposições	5º ano		9º ano	
	Limite inferior	Limite superior	Limite inferior	Limite superior
Limites sem suposições				
$E[y(0) X]$	-0,800	-0,787	-0,966	0,763
$E[y(0)/X]$	-1,780	2,951	-1,888	2,468
$E\{[y(1) - y(0)]/X\}$	-0,993	3,751	-2,652	3,434
RMT				
$E[y(0)/X]$	0,800	-0,055	-0,966	-0,059
$E[y(0)/X]$	-0,055	2,951	-0,059	2,468
$E\{[y(1) - y(0)]/X\}$	0,000	2,151	0,000	3,434
RMT e SMT				
$E[y(0)/X]$	-0,055	-0,055	-0,059	-0,059
$E[y(0)/X]$	-0,055	0,441	-0,059	0,396
$E\{[y(1) - y(0)]/X\}$	0,000	0,496	0,000	0,455
VIM RMT e SMT				
$E[y(0)/X]$	-0,049	0,110	-0,050	0,098
$E[y(0)/X]$	0,118	0,116	0,190	0,190
$E\{[y(1) - y(0)]/X\}$	0,008	0,165	0,093	0,240

Elaboração dos autores.

Nas linhas de cada tabela constam 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. Isto é visualizado sob todas as suposições, mas a amplitude dos intervalos varia significativamente entre as suposições. Por exemplo, em ambos os exames e séries, os limites superiores das expectativas condicionais dos tratados, $E[y(1)|X]$, no caso de Limites sem Suposição e RMT, são positivos. Por sua vez, no caso das suposições de RMT em conjunto com a suposição de SMT e VIM não há uma padronização nos limites, tanto para as expectativas condicionais dos tratados quanto dos não tratados.

TABELA 7

Língua portuguesa: análise de identificação parcial para alunos do 5º e 9º anos

Suposições	5º ano		9º ano	
	Limite inferior	Limite superior	Limite inferior	Limite superior
Limites sem suposições				
$E[y(0)/X]$	-0,740	0,725	-0,913	0,614
$E[y(0)/X]$	-1,799	2,871	-2,057	2,304
$E\{[y(1) - y(0)]/X\}$	-2,523	3,611	-2,671	3,217
RMT				
$E[y(0)/X]$	-0,740	-0,035	-0,913	-0,046
$E[y(0)/X]$	-0,035	2,871	-0,046	2,304
$E\{[y(1) - y(0)]/X\}$	0,000	3,611	0,000	3,217
RMT e SMT				
$E[y(0)/X]$	-0,035	-0,035	-0,046	-0,046
$E[y(0)/X]$	-0,035	0,448	-0,046	0,419
$E\{[y(1) - y(0)]/X\}$	0,000	0,483	0,000	0,465
VIM RMT e SMT				
$E[y(0)/X]$	-0,042	0,014	-0,048	0,010
$E[y(0)/X]$	0,091	0,091	0,053	0,053
$E\{[y(1) - y(0)]/X\}$	0,077	0,133	0,043	0,101

Elaboração dos autores.

Nos limites do EMT, $\Delta(0,1) = E[y(1) - y(0) | X]$, aqueles sem suposição são de larga amplitude e não proveem nenhuma informação extra em qualquer dos exames, pois ambas as estimativas do PEP e do MQO se encontram nesse intervalo. Isso também é observado para os limites sob a suposição de RMT. 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 analisam os limites sob a suposição de RMT em conjunto da suposição de SMT. Por definição da suposição de RMT, os limites inferiores são zero. Os limites superiores para o efeito de o aluno estar frequentando a escola privada, em qualquer exame ou série, situam-se acima de 0,4 desvio-padrão. Esses intervalos são informativos, porque as estimativas das metodologias de PEP e MQO 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 EMT sob as suposições de VIM, RMT e SMT são

mais informativos, pois restringem principalmente os limites inferiores a valores não nulos ou negativos. Para o desempenho dos estudantes do 5º ano, os limites, inferior e superior, do efeito da escola privada nos exames de matemática e língua portuguesa são $[0,008; 0,165]$ e $[0,077; 0,133]$, respectivamente, enquanto para os alunos do 9º ano, tais respectivos conjuntos de limites são $[0,093; 0,24]$ e $[0,043; 0,101]$. As amplitudes dos limites em ambas as séries se equivalem nos dois exames, embora a amplitude em matemática seja três vezes maior. Portanto, os intervalos obtidos em ambos os exames implicam que as estimativas derivadas dos métodos PEP e MQO, apresentados na tabela 5, do efeito da escola privada encontram-se dentro do intervalo e podem estar de duas a quatro vezes sobre-estimadas. Todavia, deve-se enfatizar que as inferências derivadas da metodologia de identificação parcial dependem de determinadas hipóteses teóricas que podem ou não ser as mais adequadas à realidade do estudo que a aplique.

6 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 entre as redes 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 PEP. 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 do método PEP e o tradicional MQO, a partir dos dados do SAEB (INEP, 2007), revelam um significativo efeito das escolas privadas sobre o desempenho dos estudantes, tanto em língua portuguesa quanto em matemática. As estatísticas descritivas dos dados igualmente evidenciam as diferenças entre as realidades dos estudantes que afetam o processo de escolha da escola pelos pais e a aprendizagem dos alunos. Isso pode ser decorrente da distribuição 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 sobre o processo de aprendizagem gerado tanto pela família quanto pela escola.

Reconhece-se, contudo, a afirmativa de Morgan (2001), segundo a qual o mais importante para políticas educacionais é que as estimativas derivadas das metodologias de regressão e PEP não proveem informações suficientes sobre como os estudantes de escola públicas poderiam se beneficiar caso frequentassem 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 menos restritivas. Assim, a metodologia de identificação parcial aponta para a sobre-estimação do efeito da escola privada, refletindo a inadequação das suposições mais restritivas que possibilitam a identificação pontual. Em suma, supondo a possibilidade de o método de identificação parcial vir a gerar informações sem utilidade ou mesmo ser informativo apenas para grandes amostras, os resultados obtidos nesta pesquisa sugerem haver sobre-estimação nas estimativas pontuais do efeito da escola privada no Brasil.

Em relação à decisão dos pais, considerando a restrição de renda e outras variáveis socioeconômicas, na escolha de qual sistema de ensino – privado ou público – matricular seus filhos, os fatores condicionantes do lado da demanda são determinantes neste processo. Reconhece-se, pela escassez na literatura, que haveria de se considerar também fatores intrínsecos, do lado da oferta, que pudessem contribuir para a explicação dessa demanda. Nesse sentido, uma pesquisa subsequente a esta poderia investigar se determinados fatores, como escolas dotadas de professores mais qualificados e motivados, orientação pedagógica, vertente religiosa, entre outros, seriam relevantes para guiar os pais a tomar decisões sobre o destino da rede de ensino para os filhos.

ABSTRACT

The core aim of this article is to compare student performance in Brazilian public and private schools based upon the methodology of Partial Identification. Since there is selection bias in the school choice derived from socioeconomic factors, identification restrictions in the other methodologies based upon point estimations to infer about the students' performance differences from the two types of schools are questionable. Therefore, it is tested whether the estimated limits provided by the proposed methodology, although under less restrictive assumptions, are as much or more informative than the point estimates generated by other methodologies. Using microdata from the

Basic Education Assessment System (SAEB) of 2005 for students in the 5th and 9th year of elementary school, the results indicate an overestimation of the private school effect from the most applied methodologies that rely on the assumptions of ignorability and imputation. However, the imposition of less restrictive hypotheses has kept such an effect significantly positive. These results shall come from the selection bias related to the parents' earnings constraints which contribute to explain the school choice and make it difficult the determination of an acceptable counterfactual.

Keywords: partial identification, private education, selection bias.

REFERÊNCIAS

- 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.
- BLUNDELL, R.; COSTA-DIAS, M. Alternative approaches to evaluation in empirical microeconomics. **Journal of human resources**, v. 44, n. 3, p. 565-640, 2009.
- 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.
- CURI, A. Z.; MENEZES-FILHO, N. A. Determinantes dos gastos com Educação no Brasil. **Pesquisa e planejamento econômico**, v. 40, n. 1, 2010.
- DRONKERS, J.; AVRAM, S. A cross-national analysis of the relations between school choice and effectiveness differences between private-independent and public schools. **Sociological theory and methods**, v. 25, n. 2, p. 183-205, 2010.
- DRONKERS, J.; ROBERT, P. School choice in the light of the effectiveness differences of various types of public and private schools in 19 OECD Countries. **Journal of school choice**, v. 2, n. 3, p. 260-301, 2008.
- 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*: ENCONTRO NACIONAL DE ECONOMIA, 37., 2009, Foz do Iguaçu, Paraná. **Anais...** Foz do Iguaçu: ANPEC, 2009.
- FRIEDMAN, M. The role of government in public education. *In*: ROBERT, A. S. (Ed.). **Economics and the public interest**. New Brunswick, NJ: University of Rutgers Press, 1955.

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.

HECKMAN, J.; LALONDE, R. J.; SMITH, J. A. The economics and econometrics of active labor market programs. *In*: ASHENFELTER, O.; CARD, D. (Ed.). **Handbook of labor economics**. New York: North Holland, 1999. v. 3, c. 31, p. 1.865-2.097.

HOXBY, C. M. **Do private schools provide competition for public schools?** Cambridge: National Bureau of Economic Research, 1994. (Working Paper Series, n. 4.978).

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Pesquisa Nacional por Amostra de Domicílios**. Rio de Janeiro: IBGE, 2005.

IMBENS, J. 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 – INSTITUTO NACIONAL DE ESTUDOS E PESQUISA. **Primeiros resultados: médias de desempenho do SAEB/2005 em perspectiva comparada**. INEP, 2007. Disponível em: <http://www.inep.gov.br/download/saeb/2005/SAEB1995_2005.pdf>. Acesso em: 21 out. 2010.

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, 1990a.

_____. Identification of endogenous social effects: the reflection problem. **Review of economic studies**, v. 60, p. 531-542, 1990b.

_____. Monotone treatment response. **Econometrica**, v. 65, p. 1.311-1.334, 1997.

_____. **Identification for prediction and decision**. Princeton: Princeton University Press, 2008.

MANSKI, C.; PEPPER, J. V. Monotone instrumental variable: with an application to the returns to schooling. **Econometrica**, v. 68, p. 997-1.010, 2000.

MORGAN, S. Counterfactual, causal effects, heterogeneity, and the catholic school effect on learning. **Sociology of education**, v. 74, p. 341-374, 2001.

MORGAN, S.; WINSHIP, C. **Counterfactual and causal inference**: methods and principles for social science. New York: Cambridge University Press, 2007.

NGUYEN, A. N.; TAYLOR, J.; BRADLEY, S. The estimated effect of catholic schooling on educational outcomes using propensity score matching. **Bulletin of economic research**, v. 58, n. 4, p. 285-307, Oct. 2006.

RAZO, R.; FERNANDES, C.; SOARES, S. **O impacto da municipalização no ensino fundamental**: uma estimativa por escores de propensão utilizando os dados do censo escolar. PUC-RJ, 2005. 37 p. Disponível em: <http://www.econ.puc-rio.br/pdf/seminario/2005/artigo_versao2_cris.pdf>. Acesso em: 3 nov. 2010.

ROSENBAUM, P. R. **Observational studies**. New York: Springer, 2002.

ROSENBAUM, P. R.; RUBIN, D. The central role of the propensity score in observational studies for causal effects. **Biometrika**, v. 70, n. 1, p. 41-50, 1983.

SOMERS, M.-A.; McEWAN, P. J.; WILLMS, J. D. How effective are private schools in Latin America? **Comparative education review**, v. 48, n. 1, p. 48-69, 2004.

VANDENBERGHE, V.; ROBIN, S. Evaluating the effectiveness of private education across countries: a comparison of methods. **Labour economics**, n. 11, p. 487-506, 2004.

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*: ENCONTRO NACIONAL DE ECONOMIA, 38., 2010, Salvador, Bahia. **Anais...** Salvador: ANPEC, 2010.

BIBLIOGRAFIA COMPLEMENTAR

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.

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.

CALIENDO, M.; HUJER, R.; THOMSEN, S. **The employment effects of job creation schemes in Germany**: a microeconomic evaluation. Bonn: Institute for the Study of Labor, 2007. (Discussion Paper Series, n. 1.512).

CRUMP, R. *et al.* Dealing with limited overlap in estimation of average treatment effects. **Biometrika**, v. 96, p. 187-199, 2008. Forthcoming.

DEHEJIA, R.; WABBA, S. Causal effects in nonexperimental studies: reevaluating the evaluation of training programs. **Journal of the American statistical association**, v. 94, p. 1.053-1.062, 1999.

_____. The effects of class size on student achievement: new evidence from population variation. **Quarterly journal of economics**, v. 116, p. 1.239-1.286, 2000a.

_____. Does competition among public schools benefit students and taxpayers. **The American economic review**, v. 90, n. 5, p. 1.209-1.238, 2000b.

ICHINO, A.; MEALLI, F.; NANNICINI, T. **From temporary help jobs to permanent employment**: what can we learn from matching estimators and their sensitivity. Bonn: Institute for the Study of Labor, 2006. (Discussion Paper Series, n. 2.149).

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.

KHANDKER, S.; KOOLWAL, G.; SAMAD, H. **Handbook on impact evaluation**. Washington, D.C.: The World Bank, 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.

NERI, M. 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, 1983.

_____. **Motivos da evasão escolar no Brasil**. Rio de Janeiro: FGV/CPS, 2010. Disponível em: <<http://www.ufgd.edu.br/faed/nefope/publicacoes/pesquisa-motivos-da-evasao-escolar>>. Acesso em: 21 out. 2010.

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**. Chicago: Federal Reserve Bank of Chicago, 2010. (Working Paper Series, n. 2010-08). Disponível em: <<http://www.econstor.eu/bitstream/10419/70596/1/636377997.pdf>>.

TODD, P. Evaluating social programs with endogenous program placement and selection of the treated. *In*: SCHULTZ, T. P.; STRAUSS, J. (Ed.). **Handbook of development economics**, v. 60, n. 4, p. 3.847-3.894, 2008.

(Originais submetidos em dezembro de 2011. Última versão recebida em junho de 2014. Aprovada em julho de 2014.)

APÊNDICE A

TABELA A.1

Resultados para as *dummies* de estados da regressão logística para determinar o escore de propensão de estar matriculado na rede privada de ensino

	5ª Matemática		5ª Português		9ª Matemática		9ª Português	
	Coef.	EP	Coef.	EP	Coef.	EP	Coef.	EP
Acre	-0,903	0,140	-0,555	0,136	-0,010	0,184	-0,007	0,184
Amazonas	-0,195	0,113	-0,053	0,113	0,752	0,166	0,709	0,163
Roraima	-1,246	0,168	-1,255	0,171	-1,414	0,327	-1,609	0,349
Pará	0,688	0,101	0,747	0,102	1,248	0,154	1,135	0,150
Amapá	-1,110	0,137	-1,043	0,138	0,351	0,175	0,192	0,173
Tocantins	-0,114	0,117	-0,077	0,118	-0,037	0,216	-0,086	0,216
Maranhão	0,679	0,098	0,777	0,098	1,194	0,153	1,138	0,149
Piauí	1,077	0,099	1,226	0,100	1,551	0,153	1,562	0,149
Ceará	1,115	0,099	1,205	0,099	1,450	0,156	1,407	0,152
Rio G. do Norte	0,821	0,100	1,005	0,100	1,368	0,153	1,371	0,149
Paraíba	0,761	0,103	0,878	0,103	1,319	0,155	1,265	0,151
Pernambuco	0,674	0,103	0,775	0,103	1,513	0,156	1,483	0,152
Alagoas	0,726	0,102	0,828	0,101	0,890	0,165	0,891	0,161
Sergipe	0,762	0,100	0,949	0,101	0,773	0,164	0,644	0,162
Bahia	0,864	0,100	0,856	0,101	1,297	0,156	1,259	0,152
Minas Gerais	-0,099	0,097	0,078	0,097	0,354	0,156	0,294	0,153
Espírito Santo	-0,028	0,114	0,031	0,115	0,416	0,163	0,472	0,159
Rio de Janeiro	0,554	0,106	0,639	0,106	0,638	0,158	0,574	0,155
São Paulo	-0,179	0,104	-0,188	0,104	0,255	0,158	0,269	0,155
Paraná	-0,455	0,100	-0,338	0,101	0,502	0,154	0,498	0,150
Santa Catarina	-0,742	0,113	-0,570	0,114	-0,127	0,164	-0,319	0,162
Rio Grande do Sul	0,177	0,100	0,313	0,100	0,317	0,156	0,260	0,152
Mato Grosso do Sul	0,330	0,104	0,426	0,103	0,655	0,155	0,598	0,151
Mato Grosso	-0,645	0,117	-0,435	0,116	0,568	0,167	0,487	0,163
Goiás	0,755	0,098	0,860	0,098	1,436	0,153	1,341	0,150
Distrito Federal	0,361	0,128	0,311	0,130	1,411	0,188	1,389	0,184

Elaboração dos autores.

Obs.: 1. O estado de Rondônia foi excluído intencionalmente para servir como parâmetro.

2. EP significa erro-padrão

