



Faculdade de Economia,
Administração e Contabilidade
de Ribeirão Preto
Universidade de São Paulo

Texto para Discussão

Série Economia

TD-E 11 / 2011

**O diferencial de notas entre as
escolas públicas e privadas no
Brasil: uma nova abordagem
quantílica**

André Guerra Esteves de Moraes



Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto
Universidade de São Paulo

Universidade de São Paulo
Faculdade de Economia, Administração e Contabilidade
de Ribeirão Preto

Reitor da Universidade de São Paulo
João Grandino Rodas

Diretor da FEA-RP/USP
Sigismundo Bialoskorski Neto

Chefe do Departamento de Administração
Marcos Fava Neves

Chefe do Departamento de Contabilidade
Adriana Maria Procópio de Araújo

Chefe do Departamento de Economia
Walter Belluzzo Junior

CONSELHO EDITORIAL

Comissão de Pesquisa da FEA-RP/USP

Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto
Avenida dos Bandeirantes, 3900
14049-905 Ribeirão Preto - SP

A série TEXTO PARA DISCUSSÃO tem como objetivo divulgar: i) resultados de trabalhos em desenvolvimento na FEA-RP/USP; ii) trabalhos de pesquisadores de outras instituições considerados de relevância dadas as linhas de pesquisa da instituição. Veja o site da Comissão de Pesquisa em www.cpq.fearp.usp.br.
Informações: e-mail: cpq@fearp.usp.br

O diferencial de notas entre as escolas públicas e privadas no Brasil: uma nova abordagem quantílica

André Guerra Esteves de Moraes

Walter Belluzzo Júnior

Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto – FEARP

Universidade de São Paulo – USP

Resumo: Este trabalho busca gerar robustez aos resultados vistos em estudos que indicam uma maior capacidade da rede particular de ensino em gerar qualidade educacional no Brasil. Para isso, uma metodologia quantílica, baseada na seleção em observadas, é realizada, que ao contrário de outras utilizadas, tem inferência assintótica.

A base de dados utilizada foi a do SAEB de 2005 para as provas de matemática de oitava série. Novamente foi evidenciada uma superioridade das escolas privadas, mesmo controlando para diversas covariadas de alunos, professores e escolas. Isso fortalece a possibilidade de política de cotas para escolas particulares.

Em relação às covariadas que reduziriam a distância entre as distribuições de notas de alunos de escolas públicas e privadas, constatou-se que fatores determinantes do grupo de alunos na escola e sala (*peer group effects*) seriam os mais importantes. Isso também favorece tal política, dado serem estes os fatores considerados importantes na literatura em gerar resultados no curto prazo.

Palavras-chave: Regressão quantílica, decomposição contrafactual, performance da educação, sistema público de ensino

Abstract: This paper aims at bringing strength to the results observed in other studies that point out a larger ability of the private school network to generate quality education in Brazil. To achieve that result, this study applies a quantilic methodology based on the selection on observable variables that, contrary to others utilized, presents asymptotic inference.

The utilized data base was that of SAEB 2005 for math tests in the 8th grade. As in other studies, here again the superiority of the private schools was made evident, even though various students', teachers' and schools' covariates are controlled. This result strengthens the possibility of a policy of quotas for private schools.

In relation to the variables that would reduced the distance between the grades distributions of students on public and private schools, peer group effects were observed to be the more important ones. This results also favors said policy because the literature has found these factors relevant to generate short term results.

Keywords: Quantile regression, counterfactual decomposition, educational performance, public education system

Área ANPEC: Área 11 - Economia Social e Demografia Econômica

Classificação no JEL: I21, C21

1. INTRODUÇÃO

Desde a realização do primeiro exame de proficiência do SAEB (Sistema de Avaliação da Educação Básica) em 1995, vem surgindo um grande número de estudos sobre os determinantes das notas médias de alunos¹. O principal objetivo desses estudos é identificar a contribuição de características observáveis de alunos, professores e escolas, para orientar a formulação de políticas públicas para a melhoria da qualidade educacional. Em geral esses estudos utilizam o conceito de função de produção escolar, através da qual se analisa as notas dos alunos em testes padronizados – *proxy* mais comum na literatura para indicar qualidade educacional, dada a sua grande correlação com a capacidade cognitiva de alunos – em relação a insumos escolares.

Porém, em sua grande maioria, não se avaliou o diferencial de notas entre alunos de escolas públicas e privadas. É uma análise interessante dado que se realmente há uma superioridade no setor privado, fatores como práticas educacionais diferentes tenderiam a elevar as notas dos alunos de escolas públicas, podendo gerar resultados com até menores custos.

Friedman (1955) argumenta que as famílias deveriam receber cupons² do governo que permitissem mandar alguns de seus filhos para escolas particulares ao invés das públicas, devido a três fatores, segundo a literatura. Primeiramente, escolas particulares tendem a ser mais eficientes em gerar qualidade educacional do que as públicas, mesmo controlando para diferenças de nível socioeconômico e outros fatores importantes para definir as notas, ou seja, a diferença entre os dois setores após o devido controle se deve a melhores práticas administrativas e educativas do setor privado. Segundo McEwan (2000), um segundo fator seria que, dependendo da nova aleatorização feita entre alunos de escolas públicas e privadas, no curto prazo, o efeito conjunto da escola³ privada e sua melhor eficiência elevariam as notas de alunos menos favorecidos. Por fim, no longo prazo, o crescimento do setor privado educacional, devido à nova política, faria com que as taxas de inscrições no setor público caíssem, teoricamente elevando a competição e forçando a sua melhoria⁴.

Antes de perseguir esta política no Brasil, é necessário responder a duas questões. O setor privado é superior ao público em geração de qualidade educacional? Uma maior competição entre escolas elevaria a qualidade educacional? Neste trabalho, o foco é responder à primeira questão.

O primeiro grande trabalho que avalia o diferencial de notas entre alunos de escolas públicas e privadas foi efetuado por Coleman *et al.* (1981) para dados *cross-section* americanos. Em seus resultados para uma regressão simples via MQO com uma *dummy* indicadora de tratamento (ser de escola privada) e outros insumos educacionais, as escolas privadas geram maior qualidade educacional do que as públicas. Para eles, isto indica uma maior eficiência das escolas particulares, dado que houve controle pelos outros insumos.

Esse trabalho foi alvo de muitas críticas. Segundo Alexander e Pallas (1985) e Hanushek (1986), as duas principais são: o uso inadequado de controles individuais de alunos e características familiares e o não uso de variáveis de professores e escolas, ou seja, má composição do controle, viés de especificação (considerar que o efeito de covariadas é análogo para os dois setores) e viés de variável omitida; além de não levarem em conta que o diferencial de notas pode simplesmente refletir o fato de que escolas particulares tendem a ter

¹ Dentre eles, destaco alguns: Barbosa e Fernandez (2001), Menezes-Filho (2007) e Menezes-Filho e Ribeiro (2009).

² Para melhor análise dos efeitos de uma política em grande escala de cupons para escolas particulares, ver McEwan (2000).

³ Conhecido como *peer-group effects*.

⁴ Neste quesito, segundo West e Woessmann (2010), a competição forçaria as instituições públicas a se inovarem (até copiando certas práticas educacionais do setor privado) e a conterem custos, supondo que as escolas particulares tenham melhor eficiência em gerar qualidade educacional, ou seja, que gerem melhores resultados e que tenham menores custos em relação às escolas públicas.

alunos já melhores inicialmente (com melhores aspectos familiares e socioeconômicos), isto é, problema de viés de seleção.

Tentando resolver tais problemas, novas metodologias e controles foram testados para avaliar o real diferencial entre os dois setores.

Com relação aos controles, uma grande evolução foi a utilização de dados de painel, que além de conter variáveis passadas, captando a qualidade do estudante antes da realização do teste a ser avaliado, retira efeitos fixos da análise (fatores não observados que influenciam as notas dos alunos e são fixos no tempo). Outra evolução foi incluir na análise dados de grupo (fatores como o nível socioeconômico médio da escola), pois é de se esperar melhores covariadas para o grupo das escolas particulares.

Com relação às metodologias, visando à análise das notas médias dos alunos, três podem ser destacadas.

Primeiramente, a baseada na decomposição de Oaxaca (1973), em que se estimam funções de produção escolares distintas para os dois setores, diminuindo consideravelmente a possibilidade de viés de má especificação e estimando o diferencial de notas entre alunos com características parecidas. Vale ressaltar que, neste caso, problemas de viés de seleção ainda são possíveis, caso fatores não observados determinem se o aluno irá para escola pública ou particular.

Outra metodologia muito utilizada foi o procedimento em dois estágios de Heckman (1979) de correção de viés de seleção. Neste caso, o primeiro estágio é um modelo *probit* para a probabilidade de o aluno ir para uma escola particular; já no segundo estágio, explica-se o diferencial de notas dos alunos com o possível viés retirado. O grande problema desta metodologia é a necessidade de um instrumento que seja correlacionado à escolha do tipo de escola e que não o seja com respeito ao erro da equação das notas.

Em um âmbito mais atual, uma metodologia utilizada foi o modelo hierárquico linear (MHL). Neste procedimento, o primeiro estágio estima as notas médias para cada escola, controlando para características individuais de seus alunos (efeito intraescolar); no segundo estágio, características escolares (entre elas uma *dummy* indicadora de setor da escola que capta o efeito privado) são colocadas em uma regressão, a fim de explicar as notas médias estimadas no primeiro estágio (efeitos entre escolas). Segundo Albernaz *et al.* (2002), levar esta estrutura de estimação em conta pode induzir a resultados diferentes do usual, dado que, ao considerar todas variáveis (agregadas e individuais) em um só nível, há subestimação do efeito das variáveis agregadas.

A grande maioria dos estudos empíricos se foca em países desenvolvidos. Para os EUA: Stevans e Sessions (2000), utilizando a metodologia para correção de viés de seleção citada acima, com a presença de resultados e covariadas de provas anteriores como controles, identificaram um efeito positivo das escolas privadas somente para alunos brancos, não para alunos de classes menos favorecidas; mais recentemente, por meio do modelo hierárquico linear, Braun *et al.* (2006) identificam que, quando não se tem controles para variáveis intra e entre escolas, os resultados demonstram grande eficiência para o setor privado, todavia, com tais covariadas em contexto, o efeito privado se torna nulo para matemática e leitura. Para a Austrália: Willian e Carpenter (1991), utilizando-se de metodologia parecida com a vista em Coleman *et al.* (1981), identificam a superioridade do setor privado; Vella (1999), por meio de um *probit* bivariado, a fim de explicar a probabilidade de o aluno se formar no ensino médio, entrar no mercado de trabalho e receber salários, identifica um efeito de escola particular positivo em relação aos dois primeiros fatores.

Com relação aos países em desenvolvimento, o número de trabalhos empíricos diminuiu consideravelmente⁵. Para o Brasil, o primeiro trabalho neste sentido parece ter sido realizado por Lockheed e Bruns (1990) para as cidades de Curitiba, Fortaleza, Salvador e São Paulo.

⁵ Aqui é citado apenas o caso brasileiro, mas existem trabalhos que realizam uma revisão empírica para países em desenvolvimento, ver Jimenez *et al.* (1991) e Somers *et al.* (2004). Eles têm resultados parecidos com o do caso brasileiro.

Nele, por meio do MHL, constatou-se resultados parecidos aos vistos em Braun *et al.* (2006) e que o diferencial de notas entre os setores caiu quando se utilizou variáveis agregadas das escolas na análise. Albernaz *et al.* (2002), com a mesma metodologia, mas para a base de dados do SAEB de 1999, demonstra que há melhor qualidade educacional para as escolas privadas, mesmo após incluir todas as covariadas de alunos e escolas na análise. Em comum aos dois trabalhos, a constatação de que as variáveis intraescolares respondem pela maior da variação do desempenho.

É importante citar que, nas análises descritas acima, o foco é na nota média, supondo que todos os alunos têm características análogas ao aluno médio. Caso o interesse seja a análise de fatores que determinam toda a distribuição de notas (levando a políticas de maior equidade educacional) e o diferencial público/privado em mesmo plano, a ênfase deve ser na visão quantílica, ou seja, neste caso os maiores objetivos são: coeficientes da regressão quantílica; o efeito tratamento quantílico (capta o diferencial público privado) e a análise contrafactual (distribuição de notas de alunos de escolas públicas, se tivessem respostas aos controles análogos aos vistos em escolas particulares).

Poucos trabalhos foram realizados para este quesito. Para o Brasil, Oliveira *et al.* (2009), por meio da metodologia de Machado e Matta (2005)⁶, avaliam os diferenciais quantílicos entre alunos de escolas públicas e privadas no SAEB 2005. Mesmo com controle de covariadas dos alunos, da sala de aula e da escola, há melhores resultados em todos os quantis para as escolas particulares. Com relação às covariadas, as escolas explicam pouco as notas, enquanto que a renda e a escolaridade média das mães dos alunos na sala de aula mostram-se importantes. Outra interessante descoberta foi que os piores alunos das escolas públicas estariam até piores, caso tivessem coeficientes análogos aos piores alunos das escolas particulares, com tal resultado se invertendo para os quantis maiores, indicando que esses retornos com relação as covariadas sejam decrescentes perante as notas.

Com esses resultados e a maior gama de possibilidades de uma avaliação por meio da análise quantílica, este trabalho tem por fim continuar a investigação de Oliveira *et al.* (2009), todavia, com a metodologia de Melly (2006) para a mesma base de dados. Este estimador paramétrico é também baseado na decomposição de Oaxaca (1973) e na seleção em observáveis, porém, com resultados assintóticos que geram estimador analítico para a variância do parâmetro, retirando a necessidade de erros padrões *bootstrap* que, em caso de uma base de dados grande são muito difíceis de serem calculados por causa do tempo de computação. Melly (2006) realiza simulações de monte carlo e compara o seu estimador com Machado e Matta (2005); ele conclui que seu estimador tem ajuste melhor com bases de dados menores, além de usar de maneira mais eficiente a informação contida nos dados. Assintoticamente, são estimadores comparáveis.

Os diferenciais deste trabalho em relação ao de Oliveira *et al.* (2009), além da utilização de nova metodologia, são: uso de um nível socioeconômico⁷ ao invés da renda per capita estimada via dados do PNAD como controle e a utilização de covariadas relativas ao grupo de alunos da escola (*peer-group effect*) presentes em Albernaz *et al.* (2002).

2. Metodologia

Utilizando a notação de Melly (2006), o principal interesse é o efeito do tratamento binário T sobre o resultado Y . Considerando que há uma amostra de n unidades indexadas por i , com n_1 unidades no tratamento e n_0 no controle. $T_i = 0$, se a unidade i não recebe o tratamento (aluno de escola particular) e $T_i = 1$, se a unidade i recebe o tratamento (aluno de

⁶ Metodologia da literatura de avaliação de políticas públicas que utiliza a decomposição de Oaxaca (1973). É um estimador paramétrico de simulação, baseado na seleção em observadas.

⁷ Modo de cálculo do nível socioeconômico e outras covariadas utilizadas no trabalho serão devidamente apresentados na parte da base de dados e análise descritiva dos dados.

escola pública). Para os resultados (notas dos alunos no SAEB), $Y_i(0)$ é o resultado do controle e $Y_i(1)$ é o resultado do tratado. Adicionalmente, cada unidade i tem um vetor de covariadas X_i K -dimensional. A notação é: $F_Y(q)$ representa a distribuição cumulativa da v.a. (variável aleatória) Y em q , $f_Y(q)$ representa a densidade da v.a. Y no mesmo ponto; $F_Y^{-1}(\theta)$ representa a inversa da distribuição cumulativa, conhecida com função quantílica, avaliada em $0 < \theta < 1$; $F_Y(q/X_i)$ é a função distribuição cumulativa avaliada em q dado X_i . Considere θ o quantil não condicional e τ o quantil condicional.

Conforme dito anteriormente, é mais comum na literatura a estimativa do diferencial médio de notas, isto é, ATE (efeito tratamento médio) e do ATT (efeito tratamento para os tratados):

$$ATE = E[Y(1)] - E[Y(0)] \quad \text{e} \quad (1)$$

$$ATT = E[Y(1)/T = 1] - E[Y(0)/T = 1]. \quad (2)$$

Para definir o efeito tratamento quantílico (QTE) e o efeito tratamento quantílico para os tratados (QTT), é necessária a definição de função quantílica. Qualquer variável aleatória a valores reais Y pode ser caracterizada por sua função distribuição cumulativa,

$$F_Y(q) = Prob(Y \leq q) \quad (3)$$

enquanto para qualquer $0 < \theta < 1$,

$$q(\theta) = \inf\{q: F_Y(q) \geq \theta\} = F_Y^{-1}(\theta) \quad (4)$$

é chamado de θ -ésimo quantil de y , ou, função quantílica. Desta forma, tem-se para o θ -ésimo quantil:

$$QTE(\theta) = F_{Y(1)}^{-1}(\theta) - F_{Y(0)}^{-1}(\theta) \quad \text{e} \quad (5)$$

$$QTT(\theta) = F_{Y(1)}^{-1}(\theta/T = 1) - F_{Y(0)}^{-1}(\theta/T = 1). \quad (6)$$

Dois pontos importantes a serem ressaltados. Primeiro, neste trabalho é identificada e estimada a diferença entre quantis e não o quantil da diferença. Segundo, o conhecimento de todos QTEs é mais informativo do que o ATE, pois a média pode ser sempre conseguida ao se integrar os quantis, outro ponto a favor da análise distribucional.

O grande problema da literatura de avaliações de políticas públicas é que os resultados potenciais são observados apenas parcialmente, dado que é observado $Y_i = (1 - T_i)Y_i(0) + T_iY_i(1)$. Com isso, são necessárias algumas suposições para identificar os parâmetros de interesse. Focando-se na seleção em observadas, pela suposição CIA (condicionando em observadas) abaixo:

$$Y(0), Y(1) \perp\!\!\!\perp T/X,$$

Tem-se que:

$$E[Y(0)/T = 1, X] = E[Y(0)/T = 0, X] = E[Y(0)/X] \quad \text{e} \quad (7)$$

$$F_{Y(0)}^{-1}(\theta/T = 1, X) = F_{Y(0)}^{-1}(\theta/T = 0, X) = F_{Y(0)}^{-1}(\theta/X). \quad (8)$$

A ideia é condicionar para fatores que determinam o aluno receber ou não o tratamento, assim, o restante da variabilidade seria aleatório, deixando o estimador do parâmetro de interesse não viesado. No Brasil, é plausível imaginar que algumas variáveis específicas que melhoram as notas dos alunos (condição da escola) e outras características, como nível

socioeconômico, determinam o aluno ir a escola particular ou pública, dado que também refletem a vontade dos pais. Mesmo assim, é possível imaginar um viés de seleção, se fatores não observados determinem o aluno ir ou não a escola particular (ex.: vontade dos pais de que os filhos obtenham os benefícios das cotas para escolas públicas nos vestibulares), porém a grande dificuldade seria encontrar instrumentos correlacionados a T e não a Y para utilização de modelos que tratam da endogeneidade de T, como em Imbens e Angrist (1994) para a média e Chernozhukov e Hansen (2001) para o caso quantílico.

Com suposições paramétricas adicionais, a decomposição de Oaxaca (1973) permite a identificação do ATE e QTE. Supondo que a esperança de Y condicional a X é uma função linear de X, então média contrafactual pode ser consistentemente estimada por $\bar{X}^1 \hat{\beta}_{OLS}^0$ em que $\bar{X}^1 = n_1^{-1} \sum_{i:T_i=1} X_i$ e $\hat{\beta}_{OLS}^0$ é o vetor de coeficiente da regressão entre Y e X, usando só unidades do controle. Pode-se decompor a diferença entre as médias dos resultados entre os tratados e controle como:

$$ATE = \bar{Y}^1 - \bar{Y}^0 = [\bar{X}^1 \hat{\beta}_{OLS}^1 - \bar{X}^1 \hat{\beta}_{OLS}^0] + [\bar{X}^1 \hat{\beta}_{OLS}^0 - \bar{X}^0 \hat{\beta}_{OLS}^0]. \quad (9)$$

Sob tais hipóteses, o primeiro termo do lado direito da equação (9) é o efeito dos coeficientes, enquanto o segundo termo é o efeito das características observadas (diferencial entre os tratados e controle, se não tivessem recebido o tratamento). No caso do QTE, algumas suposições adicionais são necessárias: todos os quantis de Y devem ser funções lineares de X; além de outras suposições típicas de regressão quantílica de Koenker e Basset (1978).

A metodologia de Melly (2006) também é baseada em tais hipóteses e tem como primeiro estágio, regressões quantílicas distintas para o tratamento (ser aluno de escola pública) e controle (ser de escola particular), gerando-se assim os coeficientes das covariadas dos dois grupos em questão para cada quantil τ^8 , conforme pode ser visto abaixo:

$$\hat{\beta}_t(\tau, x) = \operatorname{argmin}_b n_t^{-1} \sum_{i:T_i=t} \rho_\tau(Y_i - X_i b), \quad (10)$$

onde $\rho_\tau(w) = (w - 1(w \leq 0))$ é a função *check*, com $1(\cdot)$ sendo uma função indicadora.

O τ -ésimo quantil de $Y(t)$ dado X_i é estimado de modo consistente por $X_i \hat{\beta}_t(\tau, x)$. Como os quantis não condicionais não são iguais à integral dos quantis condicionais em X (somente vale isso para a média) e tal propriedade valer para as distribuições cumulativas, logo, poderia ser feita inicialmente a inversão da função quantílica condicional, a fim de se obter a distribuição cumulativa condicional e, depois, a integração em X para se obter a distribuição cumulativa não condicional. O grande problema é que não necessariamente a função quantílica condicional é monotônica, o que inviabilizaria a inversão. Para contornar esta dificuldade, aplica-se a propriedade da função distribuição condicional abaixo:

$$F_{Y(t)}(q/X_i) = \int_0^1 1(F_{Y(t)}^{-1}(\tau/X_i) \leq q) d\tau = \int_0^1 1(X_i \hat{\beta}_t(\tau, x) \leq q) d\tau. \quad (11)$$

Desta forma, um estimador natural seria, para $(\tau_0 = 0, \tau_1, \tau_2, \dots, \tau_J = 1)$:

$$\hat{F}_{Y(t)}(q/X_i) = \int_0^1 1(F_{Y(t)}^{-1}(\tau/X_i) \leq q) d\tau = \sum_{j=1}^J (\tau_j - \tau_{j-1}) 1(X_i \hat{\beta}_t(\tau, x) \leq q). \quad (12)$$

Integrando sobre o domínio das covariadas, encontra-se a distribuição cumulativa para o tratamento $T=t$, ou seja:

⁸ Duzentos quantis foram estimados para cada tipo de escola. O uso de mais quantis, além de gerar pouca diferença nos resultados, eleva consideravelmente o tempo de computação.

$$\hat{F}_{Y(t)}(q/T = t) = \int \hat{F}_{Y(t)}(q/x) dF_X(x/T = t) = n_t^{-1} \sum_{i:T_i=t} \hat{F}_{Y(t)}(q/X_i). \quad (13)$$

Adotando o conceito de ínfimo na função quantílica vista em (4), tem-se que:

$$\hat{q}_t(\theta) = \inf\{q: n_t^{-1} \sum_{i:T_i=t} \hat{F}_{Y(t)}(q/X_i) \geq \theta\}. \quad (14)$$

Assim, os estimadores dos parâmetros de interesse são dados por:

$$\hat{F}_{Y(0)}(q/T = 1) = n_1^{-1} \sum_{i:T_i=1} \hat{F}_{Y(0)}(q/X_i) \text{ Distribuição Contrafactual}^9 \text{ e} \quad (15)$$

$$\widehat{QTE}(\theta) = \inf\{q: n^{-1} \sum_{i=1}^n \hat{F}_{Y(1)}(q/X_i) \geq \theta\} - \inf\{q: n^{-1} \sum_{i=1}^n \hat{F}_{Y(0)}(q/X_i) \geq \theta\}. \quad (16)$$

A decomposição de Oaxaca (1973), neste caso, é:

$$\widehat{QTE}(\theta) = [\hat{F}_{Y(1)}^{-1}(\theta/T = 1) - \hat{F}_{Y(0)}^{-1}(\theta/T = 1)] + [\hat{F}_{Y(0)}^{-1}(\theta/T = 1) - \hat{F}_{Y(0)}^{-1}(\theta/T = 0)]. \quad (17)$$

Outro contrafactual de interesse é o efeito de cada covariada separadamente para explicar o diferencial de notas entre alunos de cada setor. Para tal, seguindo procedimento análogo ao de Oliveira *et al.* (2009), adaptado a esta metodologia, deve-se montar uma distribuição marginal seguindo procedimento análogo ao visto acima, com os coeficientes das regressões quantílicas para as escolas públicas mantidos, mas apenas uma covariada em X_1 tendo distribuição igual à vista para escolas particulares. Tomando como base o diferencial médio entre tratados e controle para a explicação, seria uma distribuição assim: $\bar{X}_{10} \hat{\beta}_1$, onde $X_{10} = [x_1^1, x_1^2, \dots, x_0^k]$. O efeito diferencial desejado seria $\bar{X}_1 \hat{\beta}_1 - \bar{X}_{10} \hat{\beta}_1$, sendo que um valor negativo indica que, se os alunos de escolas públicas tivessem distribuição para tal covariada em questão análoga ao dos alunos da rede particular, suas notas seriam melhores.

Para a realização deste contrafactual, três passos são necessários. Primeiramente, particiona-se a variável em questão (no caso, x_0^k) em m classes (o número de diferentes picos no histograma para variáveis contínuas¹⁰ ou 2 classes para o caso de uma covariada *dummy*). No segundo passo, analisa-se, para cada classe, o número de alunos presentes (suponha 100 em cada uma) e escolhe-se, de forma aleatória (com reposição) 100 alunos de escolas públicas com valores dentro de cada classe, assim teremos uma distribuição com $m \cdot 100$ alunos, onde somente uma covariada é similar a do grupo de alunos de escolas particulares. Por fim, realiza-se o procedimento descrito acima para descobrir a densidade marginal de interesse ($\bar{X}_{10} \hat{\beta}_1$).

3. Base de dados e análise descritiva

Neste trabalho, utilizamos a base de dados do SAEB de 2005. Segundo o INEP e o MEC, ela é uma prova de matemática e língua portuguesa baseada na técnica TRI (Teoria de Resposta ao Item), não obrigatória e feita para a quarta e a oitava séries do ensino fundamental, além da terceira série do ensino médio. A amostra é feita por sorteio para escolas públicas (redes municipais e estaduais) e privadas. As covariadas foram retiradas dos questionários respondidos por alunos, professores e diretores. Aqui, serão focadas as notas de matemática para a oitava série do ensino fundamental para o Brasil como um todo. A base final tem 13.394 alunos, sendo 6.674 para escolas privadas e 6.720 para escolas públicas.

⁹ A ideia desta distribuição é de que as notas e o conjunto das covariadas são das escolas públicas, mas ao invés de utilizar o coeficiente delas para construir a distribuição incondicional, utiliza-se o coeficiente das escolas particulares.

¹⁰ Número de diferentes picos no histograma = $\min\left(\sqrt{n_0}, 10 * \left(\frac{\ln(n_0)}{\ln(10)}\right)\right)$.

Essa base de dados foi escolhida por três fatores: primeiro, foi realizada no país inteiro, com um número maior de escolas particulares em relação a outros exames¹¹; segundo, foi o último ano que ela foi realizada, logo são os dados mais atuais possíveis; por fim, como já informado na introdução, por ter sido a mesma base de dados utilizada por Oliveira *et al.* (2009), os resultados gerarão robustez às constatações sobre o diferencial público/privado na educação brasileira.

Para a escolha das covariadas, primeiramente se definiu um conjunto visto na literatura como significativo a fim de explicar a variabilidade das notas nos dois setores, assim, utilizou-se o conjunto de variáveis explicativas presentes em Albernaz *et al.* (2002) e Oliveira *et al.* (2009). Para limitar este conjunto inicial, visando as mais importantes, regressões lineares separadas para escolas públicas e privadas foram realizadas e se mantiveram as variáveis que apresentaram maior significância estatística. O nível socioeconômico individual, calculado de forma similar ao Critério Brasil¹², indicador de posição social, utilizado por empresas no mercado a fim de dividir famílias em classes, segundo o poder de compra, não se mostrou significativo para as médias, porém foi mantido por ser considerado um fator importante para definir se o aluno irá ou não para escolas nos dois setores. O nível socioeconômico médio por escola foi calculado com base em todas as séries e provas e, posteriormente, introduzido como variável do aluno. Vale ressaltar que, ao contrário da variável individual, a relativa ao *peer group effect* se mostrou bastante significativa.

As definições e a análise descritiva das covariadas podem ser vistas na tabela 1 abaixo. Percebe-se que os alunos das escolas particulares possuem mais características tidas como importantes para definir a qualidade educacional no Brasil que os alunos de escolas públicas, com exceção à proporção de professores com mais de 5 anos de experiência e com pós-graduação.

¹¹ A sua substituta, a Prova Brasil, não coleta informações sobre as escolas particulares.

¹² Disponível no site: <http://www.abep.org/novo/Content.aspx?SectionID=84>.

Tabela 1 - Descrição e análise descritiva das covariadas

Variáveis	Descrição	Rede	Média	Desvio-padrão	Mínimo	Máximo
Anse	Nível Socioeconômico do Aluno Valores extremos: 1=classe E e 8=classe A1	Pública	3,72	1,49	1,00	8,00
		Particular	5,36	1,18	1,00	8,00
Asexo	1=Aluno do sexo Masculino e 0=caso contrário(c.c)	Pública	0,47	0,50	0,00	1,00
		Particular	0,49	0,50	0,00	1,00
Areprovou	1=Aluno reprovou 1 ou 2 anos e 0=caso contrário	Pública	0,35	0,48	0,00	1,00
		Particular	0,15	0,36	0,00	1,00
AbrancoAmarelo	1=Aluno da cor Branca ou Amarela e 0=caso contrário	Pública	0,41	0,49	0,00	1,00
		Particular	0,59	0,49	0,00	1,00
Aescolamae4a8	1=Mãe do aluno estudou entre 4 a 8 anos e 0=caso contrário	Pública	0,41	0,49	0,00	1,00
		Particular	0,10	0,30	0,00	1,00
Aescolamae8a11	1=Mãe do aluno estudou entre 8 a 11 anos e 0=caso contrário	Pública	0,26	0,44	0,00	1,00
		Particular	0,30	0,46	0,00	1,00
Aescolamae11a15	1=Mãe do aluno estudou entre 11 a 15 anos e 0=caso contrário	Pública	0,11	0,31	0,00	1,00
		Particular	0,53	0,50	0,00	1,00
Aescolapai4a8	1=Pai do aluno estudou entre 4 a 8 anos e 0=caso contrário	Pública	0,27	0,45	0,00	1,00
		Particular	0,11	0,31	0,00	1,00
Aescolapai8a11	1=Pai do aluno estudou entre 8 a 11 anos e 0=caso contrário	Pública	0,22	0,41	0,00	1,00
		Particular	0,28	0,45	0,00	1,00
Aescolapai11a15	1=Mãe do aluno estudou entre 11 a 15 anos e 0=caso contrário	Pública	0,09	0,28	0,00	1,00
		Particular	0,49	0,50	0,00	1,00
Amediaeducmae	Escolaridade média da mãe dos alunos na sala	Pública	6,78	2,07	0,00	15,00
		Particular	10,95	1,53	1,50	13,80
Ppos	1=Professor fez pós-graduação e 0=caso contrário	Pública	0,54	0,50	0,00	1,00
		Particular	0,50	0,50	0,00	1,00
PexpeMaisde5	1=Professor tem mais de 5 anos de experiência como professor e 0=caso contrário	Pública	0,83	0,37	0,00	1,00
		Particular	0,81	0,39	0,00	1,00
Ensem	Nível socioeconômico médio da escola do aluno Valores extremos: 1=classe E e 8=classe A1	Pública	3,74	0,74	1,48	5,60
		Particular	5,37	0,60	2,29	6,57
Earejada	1=Escola é arejada e 0=caso contrário	Pública	0,82	0,39	0,00	1,00
		Particular	0,94	0,24	0,00	1,00
Ecomput	1=Escola tem mais de 10 computadores para uso dos alunos e 0=caso contrário	Pública	0,49	0,50	0,00	1,00
		Particular	0,74	0,44	0,00	1,00
Exerox	1=Escola tem uma máquina fotocopadora e 0=caso contrário	Pública	0,58	0,49	0,00	1,00
		Particular	0,93	0,26	0,00	1,00
Erecursos	1=Insuficiência de recursos financeiros e 0=caso contrário	Pública	0,75	0,43	0,00	1,00
		Particular	0,46	0,50	0,00	1,00

Fonte: Elaboração própria com base em microdados do SAEB 2005.

Ao analisar as notas para os dois setores educacionais, como pode ser visto na tabela 2, percebe-se que a distribuição de notas das escolas privadas está deslocada à direita das públicas, evidenciando as características observadas na tabela 1.

Tabela 2 - Análise descritiva das notas

	Pública	Privada
Média	240,20	294,30
Desvio-padrão	45,95	51,04
25º percentil	208,30	260,30
Mediana	238,00	297,90
75º percentil	270,50	330,80

Fonte: Elaboração própria com microdados do SAEB 2005.

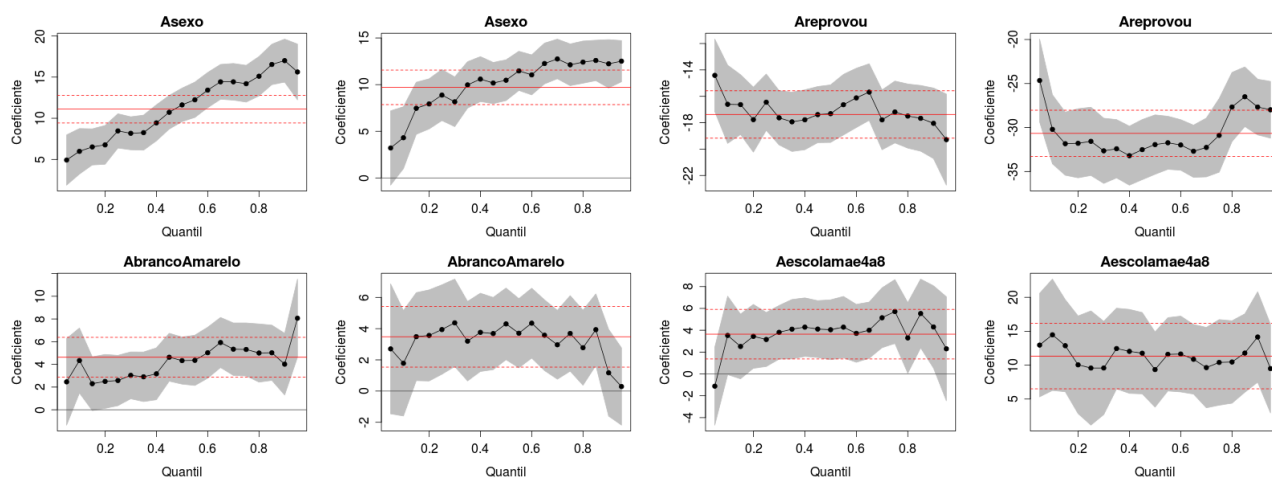
Em suma, as notas se mostram a favor das escolas particulares, mas ainda falta um condicionamento em relação às covariadas, retirando o papel delas nesse diferencial.

4. Resultados

Nesta seção, serão avaliados, inicialmente, os coeficientes das regressões quantílicas para cada setor, visando definir como as escolas públicas e particulares se diferem em relação a tais coeficientes e possíveis políticas que fariam aumentar a qualidade em cada rede de ensino. Na sequência, será analisada a decomposição de Oaxaca (1973) e a contrafactual. Por fim, a análise contrafactual do efeito de cada covariada separadamente poderá indicar o papel de cada uma para explicar o diferencial de notas antes da escolha do setor.

4.1. Coeficientes das regressões quantílicas

Aqui, serão apresentados os coeficientes que, na maioria dos quantis, foram considerados significativos. Os resultados podem ser vistos na figura 1 abaixo¹³, na qual, dois a dois, são apresentados primeiramente os das escolas públicas e, depois, os das particulares. Nota-se que, em sua maioria, eles não podem ser considerados diferentes das estimativas da média, assim, a análise para a média é uma boa aproximação.



¹³ Na área sombreada são encontrados os intervalos de confiança a 5% para os coeficientes (linhas pretas com pontos). A linha vermelha indica as estimativas para a média, com os intervalos de confiança a 5% na linha vermelha tracejada.

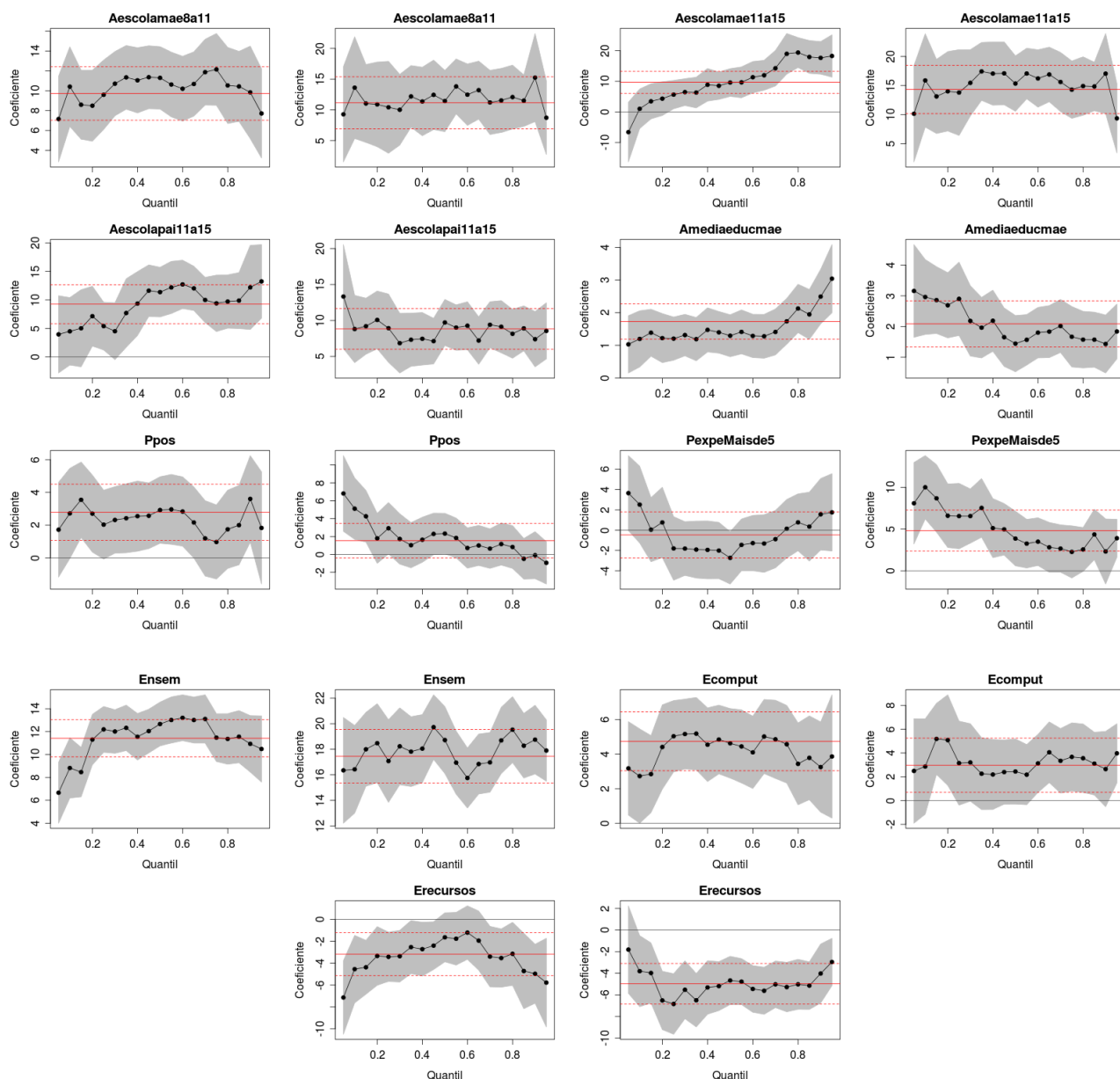


Figura 1 - Coeficientes das escolas públicas e privadas

Fonte: Elaboração própria com base em microdados do SAEB 2005.

Os alunos homens têm maiores notas em matemática do que as mulheres em todos os quantis, um fato estilizado na literatura da economia da educação; para as escolas públicas, o diferencial nos quantis superiores é maior. O fato de o aluno ter sido reprovado (*proxy* para capacidade cognitiva antes da realização dos testes) gera resultados negativos em todos os quantis da distribuição de notas, com um efeito maior para escolas particulares. Alunos brancos ou amarelos tendem a ter maiores notas que outros, com uma ligeira superioridade para as escola públicas, especialmente nos quantis maiores de notas, evidenciando uma possível maior discriminação no setor público. A escolaridade dos pais tende a afetar positivamente a nota dos alunos nas duas redes, mas há maior efeito para escola privadas. É interessante notar que, para as escolas públicas, quanto maior o quantil, maior o efeito de a mãe ter mais de 11 anos de estudos, já para o setor privado, o efeito é relativamente constante. Com relação à escolaridade média da mãe dos alunos na sala, ela tende a gerar qualidade educacional, com uma ligeira superioridade para o setor privado, tal como observado por Oliveira *et al.* (2009) e que tende a ser um efeito indireto da escolaridade da mãe.

Com relação às variáveis escolares, o fato de o professor ter pós-graduação é mais efetivo para escolas públicas e a experiência do professor é importante somente para o setor privado, assim, uma política de maior escolaridade dos professores da rede pública tenderia a aumentar as notas dos alunos em sua rede, mas análises de custo-benefício deveriam ser realizadas sobre tal quesito. O nível socioeconômico médio da escola tende a gerar mais resultados para escolas particulares portanto, nota-se que realmente é um fator importante para as notas dos alunos; assim, isto teoricamente favoreceria a política de cotas para escola particulares devido ao seu papel no *peer group effects* (a escolaridade média da mãe dos alunos na sala tende a ter o mesmo efeito). O número de computadores é importante para gerar maior qualidade de educação para alunos de escolas públicas. A falta de recursos financeiros afeta mais as notas das escolas privadas.

4.2. Decomposição de Oaxaca (1973) e análise contrafactual

É evidenciado na tabela 2 que a distribuição de notas dos alunos de escolas particulares é superior ao das escolas públicas, como se houvesse um deslocamento à direita na distribuição de notas de alunos de escolas públicas. Como já descrito na introdução, somente isso não é evidência para definir se o setor privado é superior em gerar qualidade educacional, assim, a análise do QTE e da decomposição de Oaxaca (1973) são necessários. Na figura 2¹⁴, além de tais parâmetros, há uma comparação entre as distribuições cumulativas dos alunos de escolas públicas e seu contrafactual.

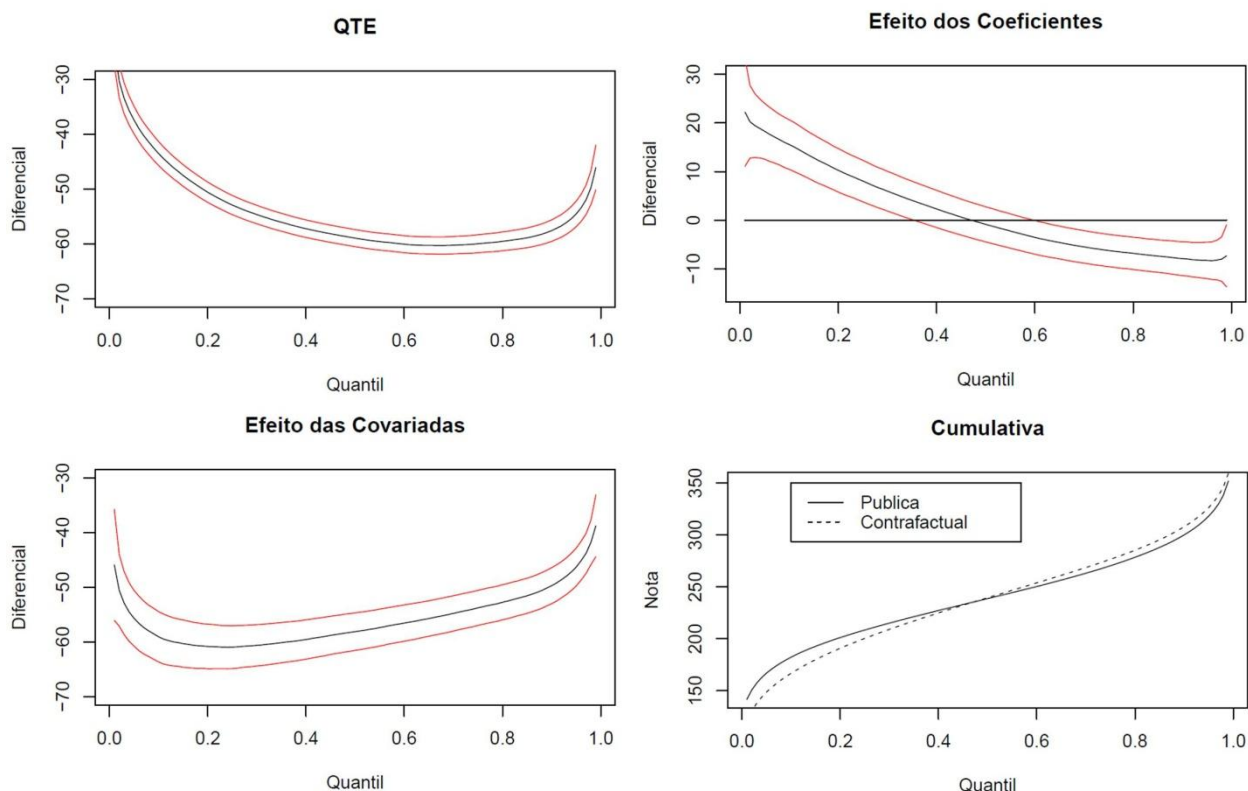


Figura 2 - Decomposição de Oaxaca (1973) e contrafactual

Fonte: Elaboração própria com base em microdados do SAEB 2005.

Os resultados são bastante similares aos vistos em Oliveira *et al.* (2009), mas agora comprovados pelos intervalos de confiança, algo não evidenciado no estudo anterior. Em todos os quantis, o QTE é negativo, demonstrando a superioridade do setor privado em geração de

¹⁴ As linhas em vermelho representam o intervalo de confiança a 5%.

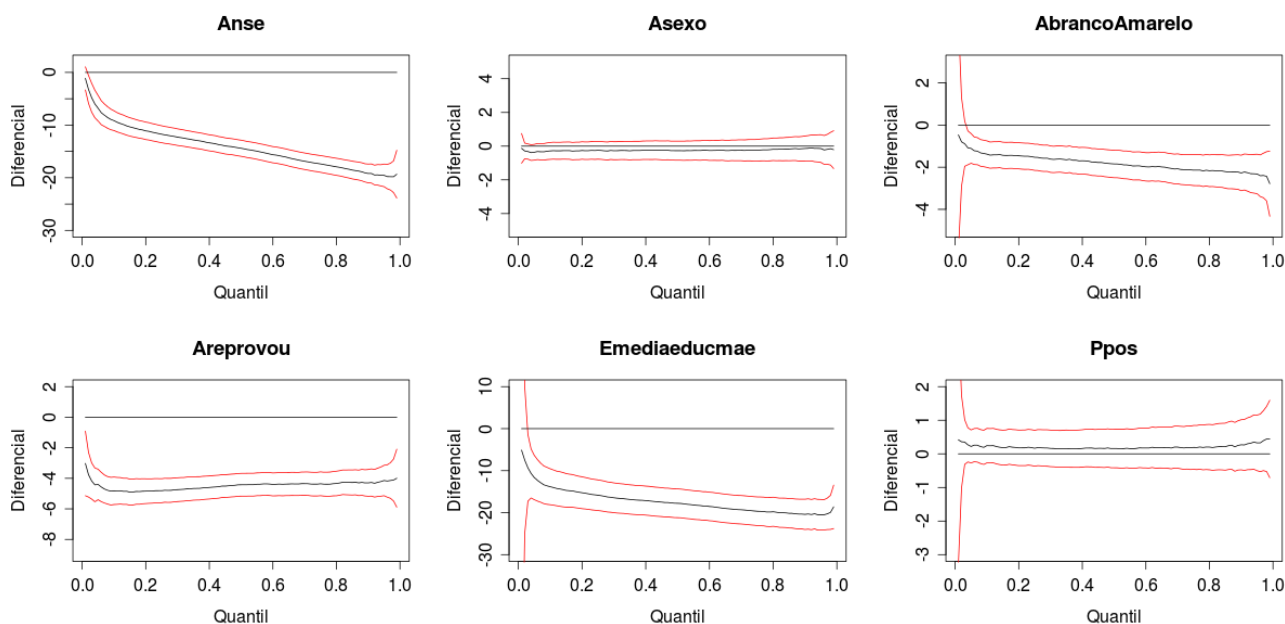
qualidade educacional. É interessante notar que o diferencial é maior nos quantis mais centrais da distribuição.

O efeito dos coeficientes¹⁵ (diferencial entre as funções quantílicas de alunos de escolas públicas e os do contrafactual), assim como a comparação entre as distribuições cumulativas, demonstram que os piores alunos das escolas públicas estariam até piores que antes, caso tivessem os coeficientes dos piores alunos das escolas particulares; um efeito contrário é visto para os melhores alunos. Isso implica duas considerações: primeiro, a variabilidade das notas seria maior, ou seja, elas seriam mais desiguais nesse caso; segundo, os retornos às características dos alunos da rede pública parecem ser decrescentes em relação à performance, dado que, mesmo com a superioridade dos piores alunos das escolas particulares nas notas, seus retornos com relação às covariadas são menores, porém, como há, nos maiores quantis, uma superioridade dos alunos de escolas particulares nos coeficientes, tal padrão também pode ser explicado por uma relação crescente desses com as suas notas, mesmo que seja em menor magnitude.

Quanto aos efeitos das covariadas, assim como ocorre para o QTE, são todos negativos e, nos quantis extremos das distribuições, há um menor efeito. Isso explica o fato de que mesmo os piores estudantes nas escolas públicas tendo uma maior resposta às características com relação aos piores alunos das escolas particulares, eles têm notas piores que os da outra rede.

4.3. Efeito individual das covariadas

Nota-se, na seção anterior, que grande parte do diferencial entre os alunos das duas redes se deve às suas características. Para identificar qual dos fatores possibilitaria uma maior diminuição da distância entre as duas distribuições, foi realizado o procedimento definido na metodologia¹⁶ para captar os efeitos das covariadas individuais, sendo gerada a figura 3 abaixo (interpretar de mesma forma que o efeito das covariadas visto na figura 2):



¹⁵ Os coeficientes também podem ser definidos como retorno das covariadas.

¹⁶ Na análise descritiva constata-se que, para as covariadas Amediaeducmae e Ensem, os extremos são diferentes para as duas redes, assim, para a construção de seus contrafactuals individuais, foram utilizados os menores limites em módulo entre os dois setores para a divisão em diferentes classes.

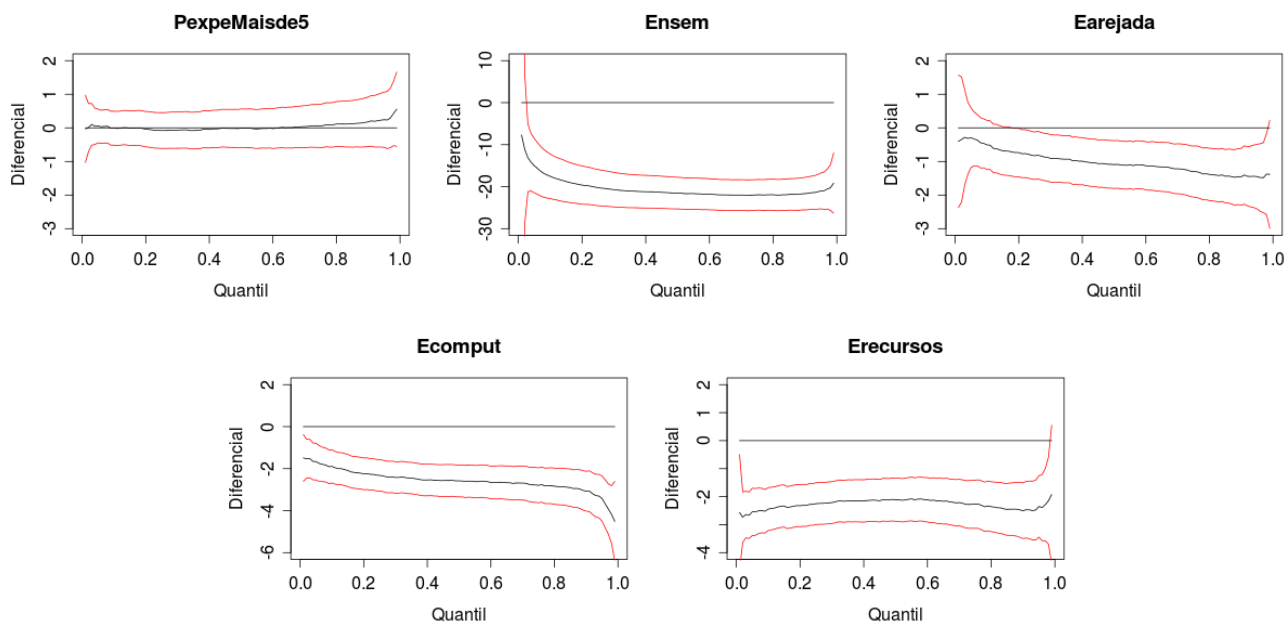


Figura 3 - Efeito das covariadas individuais
 Fonte: Elaboração própria com base em microdados do SAEB 2005.

Dentre os fatores, os mais importantes foram o nível socioeconômico do aluno, a média de educação das mães dos alunos na sala e o nível socioeconômico médio das escolas. Dentre eles, dois refletem *peer group effects*, também favorecendo uma política de cotas em escolas particulares. Com relação às variáveis diretamente passíveis de políticas, sem mudar o formato hierárquico do ensino fundamental, um maior número de computadores, menores taxas de reprovação¹⁷ e mais recursos ao setor público são as que teoricamente ajudariam a diminuir a distância entre as duas instituições em geração de qualidade educacional.

5. Conclusões

Este trabalho teve por objetivo gerar robustez aos resultados observados em outros estudos sobre diferencial entre as notas de escolas públicas e particulares. Para tal, por meio de metodologia baseada na decomposição de Oaxaca (1973) e seleção em observadas de Melly (2006), buscou-se controlar para fatores individuais de alunos, professores e escolas (incluindo os relativos ao *peer group effects*), a fim de observar qual dos dois setores realmente é melhor em gerar qualidade educacional, o que responderia à primeira condição a favor de uma política de cotas para escolas particulares. Vale ressaltar que, diferentemente de outros trabalhos no âmbito quantílico, este utilizou metodologia com inferência assintótica.

Primeiramente, analisou-se como cada covariada afeta de forma individual as distribuições em questão, tentando buscar fatores que melhorariam as notas de ambas as redes de ensino. Assim como em estudos que se focam em escolas públicas, resultados parecidos foram vistos para escolas particulares. Destaca-se que professores com pós-graduação e escolas com mais computadores são mais importantes para escolas públicas; já para escolas particulares, professores com maior experiência e a inexistência de problemas com recursos financeiros são mais expressivos.

Depois, com a análise do QTE, constatou-se o diferencial visto entre as duas redes de ensino para todos os quantis. Vale notar que, na análise contrafactual e no efeito dos coeficientes, assim como em estudos anteriores, os piores alunos das escolas públicas

¹⁷ Isso não implica que há favorecimento na política de educação continuada aplicada em alguns estados brasileiros; uma melhor análise neste quesito deve ser feita. Para maiores detalhes, ver Menezes-Filho et al.(2008).

estariam piores, caso tivessem coeficientes dos piores alunos de escolas privadas, havendo um efeito contrário em relação aos melhores estudantes, ou seja, tal possibilidade faria uma distribuição de notas menos igualitária.

Com relação ao efeito das covariadas, elas explicam grande parte do diferencial dos alunos, em especial as relativas ao *peer group effect*, também gerando mais pontos a favor de políticas de cotas para escolas particulares, dado que alunos menos favorecidos estariam em um ambiente mais favorável, gerando resultados no curto prazo. Com relação às covariadas passíveis de políticas que não mudem o caráter hierárquico do ensino fundamental brasileiro, mais computadores, menores taxas de reprovação e menos problemas de falta de recursos financeiros ajudariam a diminuir a distância entre a distribuição de notas nas duas redes de ensino.

Referências Bibliográficas

- ALBERNAZ, A.; FERREIRA, F. e FRANCO, C.. Qualidade e equidade no ensino fundamental brasileiro. *Pesquisa e Planejamento Econômico*, v. 32, n. 3, p.453-476, 2002.
- ALEXANDER, K. L. e PALLAS, A. M.. School sector and cognitive performance: When is a little a little? *Sociology of Education*, v. 58, n. 2, p. 115-128, 1985.
- BARBOSA, M., FERNANDES, C.. A escola brasileira faz diferença?: uma investigação dos efeitos da escola na proficiência em matemática dos alunos da quarta série. In: FRANCO, C. (ed.).*Promoção, ciclos e avaliação na educação*. Porto Alegre: ArtMed, 2001.
- BRAUN, H.; JENKINS, F. e GRIGG, W.. *Comparing Private Schools and Public Schools Using Hierarchical Linear Modeling*. U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences, Washington, DC. 2006.
- CHERNOZHUKOV, V. e HANSEN, C..An IV model of Quantile Treatment Effects. *Econometrica*, v. 73, p. 245-261, 2005.
- COLEMAN, J. S., HOFFER, T., and KILGORE, S.. *Public and private schools.Report to the National Center for Education Statistics*. National Opinion Research Center, Chicago.1981.
- FRIEDMAN, M. The role of Government in Education. In: SOLO, R. A.(ed.). *Economics and the Public Interest*. New Brunswick, N.J.: Rutgers University Press, 1, 1955, p. 123-144.
- HANUSHEK, E. The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature*, v. 24, p. 1.141-1.177, 1986.
- HECKMAN, J.. Sample Selection Bias as a Specification Error. *Econometrica*, v.47, p. 153-161, 1979.
- IMBENS, G. e ANGRIST, J.. Identification and Estimation of Local Average Treatment Effects. *Econometrica*, v. 62, p.467-476, 2004.
- JIMENEZ, E.; LOCKHEED, M. E. e PAQUEO, V.. The relative efficiency of private and public schools in developing countries. *The World Bank Research Observer*, v. 6, n. 2, p.205-218,1991.
- LOCKHEED, M. E. e BRUNS, B.. *School Effects on Achievement in Secondary Mathematics and Portuguese in Brazil.WPS 525*, World Bank, Washington, D.C., 1990.

- LOCKHEED, M. E. e PAQUEO, V.. The relative efficiency of private and public schools in developing countries. *The World Bank Research Observer*, v. 6, n. 2, p.205-218,1991.
- KOENKER, R. e BASSET, G.. Regression Quantiles. *Econometrica*, v. 46, n. 1, 1978.
- MACHADO, J. A. F. e MATA, J.. Counterfactual Decomposition of Changes in Wage Distribution using Quantile Regression. *Journal of Applied Econometrics*, v. 20, n. 4, p. 445-465, 2005.
- MCEWAN, P. J.. The Potential Impact of Large-Scale Voucher Programs. *Review of Educational Research*, v. 70, n. 2, pp.103-149, 2000.
- MELLY, B..Estimation of Counterfactual Distribution using Quantile Regressions. University of St. Gallen, 2006.
- MENEZES-FILHO, Naercio. "Os Determinantes do Desempenho Escolar do Brasil"; Instituto Futuro Brasil, Ibmec-SP e FEA-USP; 2007.
- MENEZES-FILHO, Naercio; RIBEIRO, Fernanda P.. Os determinantes da Melhoria do Rendimento Escolar. In: GIAMBIAGI, Fabio; HENRIQUES, Ricardo; PESSÔA, Samuel; VELOSO, Fernando (Org.). *Educação Básica no Brasil*. Rio de Janeiro: Elsevier Editora Ltda., 2009, v.1, p.171-188.
- OAXACA, R..Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, v. 14, p. 693-709, 1973.
- OLIVEIRA, P. R.; BELLUZZO, W. e PAZELLO, E. T.. Public-private sector differentials in Brazilian education: a counterfactual decomposition approach. In: 31 Encontro Brasileiro de Econometria, 2009, Foz do Iguaçu. 31 Encontro Brasileiro de Econometria, 2009.
- STEVANS, L. e SESSIONS, D. N.. Private/Public School Choice and Student Performance Revisited. *Education Economics*, v.8, n.2, 2000.
- Vella, F.. Do catholic schools make a difference? Evidence from Australia.*The Journal of Human Resources*, v. 34, n.1 , p. 208-224,1999.
- WEST, M. R. e WOESSMANN, L.. 'Every Catholic Child in a Catholic School': Historical Resistance to State Schooling, Contemporary Private Competition and Student Achievement across Countries. *The economic Journal*, v.120, p. 229-255, 2010.
- WILLIAMS, T. e CARPENTER, P.. Private schooling and public achievement in Australia. *International Journal of Educational Research*, v. 15, n. 5, p. 411-431,1991.