

DECOMPOSIÇÃO QUANTÍLICA INCONDICIONAL DOS DIFERENCIAIS DE DESEMPENHO ENTRE ALUNOS DE ESCOLAS PRIVADAS E PÚBLICAS PROFISSIONALIZANTES¹

Ronaldo A. Arraes²

Francisca Zilania Mariano³

Este artigo contribui com a literatura, ao trazer novas evidências sobre os diferenciais de desempenho entre os alunos de nível médio de escolas privadas regulares e escolas públicas; estas representadas pelas profissionalizantes que detêm os alunos com melhores resultados escolares. A aplicação empírica diferencia-se pela utilização de regressão quantílica incondicional (RQI) com aplicação da decomposição de Oaxaca-Blinder. Constatou-se haver superioridade dos alunos das escolas privadas em todas as áreas do Exame Nacional do Ensino Médio (Enem), principalmente em redação, cujos diferenciais de *performance* oscilaram entre 12% e 16% ao longo da distribuição. A decomposição detalhada dos efeitos indicou que escolaridade e renda dos pais se destacam para explicar tais diferenciais.

Palavras-chave: desempenho escolar; escolas públicas profissionalizantes; escolas privadas; decomposição quantílica incondicional.

UNCONDITIONAL QUANTILE DECOMPOSITION OF PERFORMANCE DIFFERENTIALS BETWEEN STUDENTS FROM PRIVATE AND VOCATIONAL PUBLIC SCHOOLS

The article contributes to the literature by bringing new evidences on the performance differentials between high school students of vocational public schools and regular private schools. The empirical application is distinguished by applying unconditional quantile regression with the Oaxaca-Blinder decomposition. Private schools' students were found to possess superior performance in all areas of the Enem examination, especially in writing, whose performance differentials ranged from 12% to 16% throughout the distribution. The decomposition of the effects indicated that schooling and parents' earnings are very relevant to explain such differentials.

Keywords: school performance; vocational public schools; private schools; unconditional quantile decomposition.

JEL: C31; I21; I25.

1. Os autores agradecem à Secretaria de Educação do Estado do Ceará e ao Ministério do Trabalho (MTb) por disponibilizarem os dados utilizados na verificação empírica deste trabalho, e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio.

2. Coordenador do Programa de Economia Profissional (PEP) e professor do Programa de Pós-Graduação em Economia e do Departamento de Economia Aplicada da Universidade Federal do Ceará (Caen/UFC). *E-mail:* <ronald@ufc.br>.

3. Professora do Departamento de Economia (*campus* de Sobral) e do Programa de Economia Profissional da UFC, e pesquisadora do Programa de Pós-Graduação em Economia dessa universidade. *E-mail:* <zilania@ufc.br>.

1 INTRODUÇÃO

Embora seja consenso que possuir uma boa educação básica seja uma das eficazes vertentes se alcançar o desenvolvimento de forma sustentável, essa ainda se apresenta de baixa qualidade e com alguns entraves, principalmente nos países em desenvolvimento como o Brasil. Muito se tem discutido sobre formas de melhorá-la nas diferentes fases que a compõe, quais sejam, ensino fundamental I e II e ensino médio. Esta última etapa é considerada uma das mais problemáticas, apresentando altos índices de evasão, repetência e baixo desempenho em testes padronizados, principalmente comparada ao ensino privado. Além disso, essa fase é marcada pela transição para o ensino superior e/ou para o mercado de trabalho, o que requer bastante atenção por partes dos gestores e políticos para que se possam desenvolver um ensino médio de qualidade e atrativo aos jovens, e que os capacitem para a escolha que desejarem tomar. Porém, muitos jovens não se sentem preparados a entrarem em alguma universidade e nem de adquirem um emprego de qualidade, levando, muitas vezes, à condição de nem estudar nem trabalhar. Isso eleva os índices de jovens ociosos em situação de vulnerabilidade social, contribuindo, assim, para o ciclo da pobreza e da desigualdade socioeconômica.

De acordo com informações do Ministério da Educação (MEC), o Brasil registrou melhorias no Índice de Desenvolvimento da Educação Básica (Ideb) para o ensino fundamental, ao contrário do índice do nível médio que permaneceu estagnado, considerando todas as redes de ensino em 2015.⁴ Apesar dessa evolução, apenas o ensino fundamental I conseguiu atingir a meta, superando-a em 0,3 pontos.

Estudos que tratam do diferencial de desempenho entre escolas públicas e privadas evidenciam uma superioridade do ensino particular, mesmo após controlar um conjunto de características pessoais e socioeconômicas, bem como a escolaridade dos pais de alunos (França e Gonçalves, 2010; Moraes e Belluzzo, 2014; Costa e Arraes, 2014).

É comum na literatura trabalhos que comparam os diferenciais de notas de testes padronizados entre alunos provenientes das escolas públicas e alunos advindos de escolas privadas. Buscando mensurar as diferenças de desempenho entre esses tipos de escolas na fase inicial do ensino fundamental, França e Gonçalves (2010) estimam o efeito do ensino privado sobre os resultados de alunos da 4ª série, a partir da aplicação do método dos mínimos quadrados ordinários (MQO) ponderados por *propensity score* aos dados do Sistema Nacional de Avaliação da Educação Básica (Saeb) de 2003. De acordo com os resultados encontrados pelos autores, há uma grande diferença de notas entre os dois setores de ensino, sendo que o desempenho acadêmico de alunos da rede privada é superior em 0,9 desvio-padrão, aproximadamente, mesmo após controlar um conjunto de características socioeconômicas, escolares e regionais.

4. Informações encontram-se no *link* disponível em: <<https://bit.ly/2ErJraR>>. Acesso em: 24 set. 2016.

Costa e Arraes (2014) questionam os resultados de estudos que consideram as suposições de identificação da metodologia de pareamento no escore de propensão e regressões lineares na investigação da diferença de resultados entre as escolas públicas e privadas. Para tanto, utilizam o método de identificação parcial de Manski (1989), o qual permite estimar limites com suposições menos restritivas e verificar se as estimações pontuais das metodologias usuais não ultrapassam os limites do intervalo. A partir de informações do Saeb de 2005 para estudantes do 5º e 9º anos do ensino fundamental, os autores comprovaram que essas metodologias usuais sobre-estimam o efeito das escolas privadas, apesar de esse efeito permanecer positivo e significativo.

Moraes e Belluzzo (2014) utilizaram a decomposição por quantis de Melly (2005) para a distribuição condicional de notas, a partir da base de dados do Saeb de 2005 para os alunos do 9º ano do ensino fundamental. Os autores repararam que o diferencial é favorável às escolas privadas, em todos os quantis da distribuição, chegando a 1,5 erro-padrão em torno do percentil 600. Ademais, a análise do contrafactual indica que o desempenho das escolas particulares é relativamente menor na cauda inferior da distribuição.

No Brasil, entre os tipos de políticas públicas que buscam melhorar a qualidade de ensino público e reduzir a quantidade de crianças e adolescentes ociosos, encontra-se o ensino integral, o qual aumenta a carga horária dos alunos nas escolas e fornece outras atividades que sejam atrativas a permanência deles no ambiente escolar. Além desse tipo de ensino, existe o ensino profissionalizante, que busca preparar os jovens para a inserção no mercado de trabalho, por meio de cursos técnicos voltados para diversas áreas.

As escolas de educação de ensino profissional (EEEPs) levam em consideração esses dois tipos de ensino, integral e profissionalizante, e proporcionam ao jovem, durante o ensino médio, cursar as disciplinas básicas dessa etapa e participar de cursos técnicos e profissionalizantes voltados para diversas áreas (saúde, agricultura, computação, finanças etc.). Essa ação possibilita ao aluno concluir o ensino médio e se profissionalizar para o mercado de trabalho, além da alternativa de capacitá-lo para ingressar em universidades.

É, pois, centrado nessas políticas de ensino que este trabalho se desenvolve, com ênfase no ensino profissionalizante. Entretanto, cabe destacar que se optou por conduzir as análises desta pesquisa tendo o estado do Ceará como objeto de estudo, decorrente de sua representatividade e pertinência na geração das amostras para verificações empíricas. Ao considerar esses dois tipos de políticas, esse estado ganhou destaque em âmbito nacional, ao combinar o ensino integral com o profissionalizante e aumentar a participação de jovens no Exame Nacional de Ensino Médio (Enem).

Segundo informações disponibilizadas pelo Ipece (2013), no período 2009-2011, o Ceará apresentou a quarta maior participação de estudantes de escolas públicas estaduais no Enem em 2011. A taxa de participação desses estudantes passou de 2,78%, em 2009, para 8,89%, em 2011, perfazendo a terceira maior variação (219,78%) entre os estados brasileiros. Além disso, em 2011, a média das notas de escolas profissionalizantes cearenses no Enem (480.8) foi superior à média nacional (476.6), enquanto a média das escolas estaduais não profissionalizantes do estado foi de 442.5. Nesse sentido, se considerasse apenas a média dessas escolas para representar o Ceará, o estado ocuparia a oitava posição na classificação geral.

Depreende-se dos trabalhos supracitados que os alunos são influenciados por diversos fatores não controlados, tornando-os bastante heterogêneos, razão da forte explicação de fatores não observáveis. Ao contrário, este estudo busca reduzir a influência desses fatores, ao se construir grupos de alunos mais homogêneos para se aferir o diferencial de rendimento escolar líquido, notadamente entre os alunos provenientes de escolas públicas profissionalizantes e privadas no Ceará, a partir dos resultados destes no Enem de 2014.

Vale ressaltar que esses autores não encontraram trabalhos na literatura que analisem o diferencial de notas entre alunos do ensino médio das escolas públicas e privadas, impedindo que seja feita uma comparação. Cabe também destacar que, embora a análise aqui conduzida lide com apenas um estado da Federação, se observa, por meio do gráfico A.1, no apêndice A, que o desempenho dos alunos das escolas particulares do Ceará no Enem de 2014 possui distribuição semelhante aos de alunos de outros estados de diferentes regiões; por exemplo, Rio de Janeiro, Rio Grande do Norte, Pará, Santa Catarina e Goiás.⁵ Assim, há indícios que os resultados aqui encontrados podem ser generalizados para os demais estados.

A pertinência para tal escolha prende-se ao fato de as escolas de ensino profissional apresentarem desempenho superior às demais escolas públicas, uma vez que os estudantes ingressantes passam por um processo seletivo mais rigoroso; conseqüentemente se acredita que os melhores alunos provenientes do ensino fundamental são selecionados, além de possuírem estruturas mais adequadas e especializadas para proporcionar um ensino de melhor qualidade.⁶ Nesse sentido, pretende-se responder aos seguintes questionamentos: esse novo modelo de escola pública provê uma estrutura de ensino equiparável ao da escola privada? Como os alunos dos dois tipos de escola se comparam ao longo da distribuição de rendimento escolar?

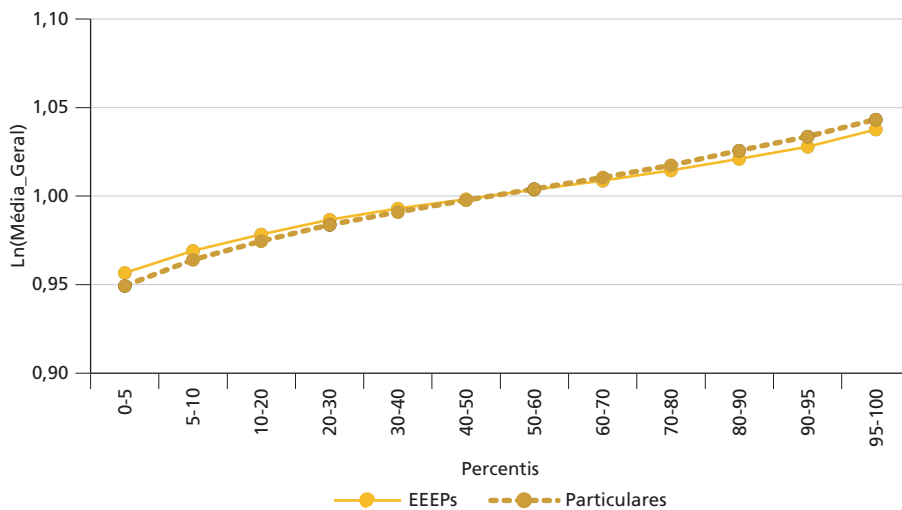
5. Optou-se por comparar o Ceará com um estado de cada região para não expor o gráfico com excesso de informações e ilegível.

6. Para comprovar essa afirmação, optou-se por encontrar o diferencial no desempenho no Enem dos alunos das escolas profissionalizantes e dos estudantes das escolas estaduais regulares (não profissionalizantes), para posteriormente fazer a comparação apenas das escolas particulares com as de ensino profissional.

Quais fatores observáveis mais influenciam os desempenhos dos alunos das duas escolas? Como os fatores não observáveis se comparam quanto ao diferencial de rendimentos entre os alunos?

Os gráficos 1 e 2 destacam o comportamento das notas na média geral (MG) no Enem e os diferenciais segundo decis da distribuição. No gráfico 1, é possível ter uma ideia da evolução das notas ao longo da curva. Uma visualização mais clara dos diferenciais de notas entre as escolas profissionalizantes e particulares é exposta no gráfico 2, em que se verifica uma diferença negativa e positiva nos decis inferiores e superiores, respectivamente. Dessa forma, a aplicação de um modelo quantílico se mostra mais adequada para analisar o diferencial no desempenho entre essas escolas.

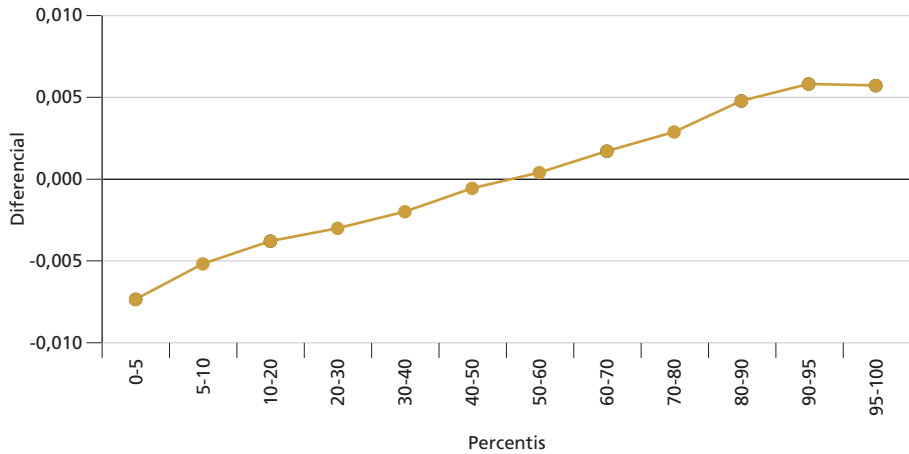
GRÁFICO 1
Logaritmo das notas da média geral do Enem de 2014



Fonte: Microdados do Enem 2014. Disponível em: <<https://bit.ly/2QGmubm>>.
Elaboração dos autores.

Tendo em vista que a aplicação de distribuição quantílica se torna apropriada, este trabalho buscou inovar ao utilizar o método de regressão quantílica incondicional (RQI) proposto por Firpo, Fortin e Lemieux (2007), o qual é considerada uma generalização do método de Oaxaca-Blinder, com o emprego de regressões baseadas na função de influência recentrada (FIR) combinadas com um procedimento de reponderação inspirado em DiNardo, Fortin e Lemieux (1996). Essa foi escolhida em detrimento do método tradicional de regressão quantílica (RQ) condicional proposto por Koenker e Bassett (1978), pois este não permite interpretar as estimativas como o impacto marginal da variável explicativa sobre a variável dependente (Firpo, Fortin e Lemieux, 2007; 2009).

GRÁFICO 2
Diferencial de notas na média geral no Enem de 2014



Fonte: Microdados do Enem 2014. Disponível em: <<https://bit.ly/2QGmubm>>. Elaboração dos autores.

Em sequência, o texto está organizado em mais três seções, além desta introdução. A seção 2 aborda o método de decomposição de Firpo, Fortin e Lemieux (2007), já a seção 3 mostra os resultados do modelo. Por fim, na seção 4, apresentam-se as conclusões deste estudo.

2 MÉTODO DE DECOMPOSIÇÃO

Quando se pretende realizar uma decomposição do tipo Oaxaca-Blinder em diferentes quantis ao longo da distribuição, não se pode simplesmente executar RQ nos quantis de interesse e usar os coeficientes estimados no método Oaxaca-Blinder, pois, diferentemente dos obtidos via MQO, que possuem interpretação condicional e incondicional, os coeficientes estimados por regressões quantílicas têm apenas a interpretação condicional.

Sabe-se que o método MQO estima os coeficientes (β) de uma média condicional, $E[Y|X] = X\beta$, da qual se interpreta β como $\partial E[Y|X]/\partial X$. Aplicando a lei das expectativas iteradas, $E_X[E[Y|X]] = E[Y] = E[X]\beta$, β também pode ser explicado como o efeito das alterações na média de X na média incondicional de Y , embora essa lei não seja válida para quantis; ou seja, se aplicar a esperança condicional em quantil τ específico, $Q_\tau(X)$, não se encontra o quantil incondicional, $E_X[Q_\tau(X)] \neq Q_\tau$. Consequentemente, os coeficientes da regressão linear quantílica $Q_\tau(X) = X\beta$ têm somente interpretação condicional, de modo que não se pode interpretá-los como o efeito das alterações na média de X no quantil τ th marginal (incondicional).

Dessa forma, RQ fornece uma abordagem dos impactos diferenciados de covariáveis ao longo da distribuição de um resultado; assim, esta é utilizada para avaliar o efeito de uma variável em um quantil do resultado condicional a valores específicos de outras variáveis. Entretanto, Borah e Basu (2013) afirmam que, na maioria dos casos, RQ pode gerar resultados que muitas vezes não são generalizáveis ou interpretáveis.

Na média, é suficiente estimar somente o valor médio condicional nos quantis de interesse, embora todo o contrafactual da distribuição acumulada incondicional $F_{Y_B^A}(Y)$ deve ser estimado, e um contrafactual quantil incondicional de interesse pode ser recuperado por inversão, $Q_{\tau,B}^A = F_{Y_B^A}^{-1}(\tau)$, no qual os subscritos A e B representam os dois grupos em comparação.

Vários procedimentos para a obtenção de distribuição contrafactual $F_{Y_B^A}(Y)$ têm sido sugeridos na literatura. Juhn, Murphy e Pierce (1993), Machado e Mata (2005) e Melly (2005) visam substituir cada resultado em B com um contrafactual Y_B^A . Juhn, Murphy e Pierce (1993) utilizam uma abordagem de imputação residual, enquanto Machado e Mata (2005) se baseiam em simulações e regressões de quantil condicional. DiNardo, Fortin e Lemieux (1996) propuseram uma abordagem semiparamétrica reponderada, que usa um fator de reponderação estimado de impor as características dos indivíduos do grupo B sobre os indivíduos do grupo A.

Existem também métodos que primeiro estimam a distribuição condicional $F_{Y_{A|X_A}}(Y|X)$ e, em seguida, integram-na sobre a distribuição dos indivíduos do grupo B, $F_{X_B}(X)$, para obter a distribuição contrafactual $F_{Y_B^A}(Y)$. Neles, são incluídos os métodos paramétricos de Donald, Green e Paarsch (2000) e Fortin e Lemieux (1998), bem como uma abordagem menos restritiva de Chernozhukov, Fernandez-Val e Melly (2013).

Se o objetivo for apenas realizar uma decomposição agregada do quantil – ou seja, estimar os efeitos explicáveis e não explicáveis da variável de resultado, sem o interesse de se fazer uma decomposição mais detalhada –, qualquer um desses métodos poderia ser utilizado; caso contrário, esses métodos seriam limitados e inapropriados. Apesar de alguns permitirem, os procedimentos variam em grau de complexidade, e a decomposição deve ser feita sequencialmente; assim, os resultados irão depender da ordem em que as características forem introduzidas. Firpo, Fortin e Lemieux (2007) propuseram uma técnica simples que permite fazer uma decomposição detalhada no quantil de interesse, cujos resultados independem da sequência. O método baseia-se na função de regressão de influência (função de influência – FI) ou RQI de Firpo, Fortin e Lemieux (2009). O método RQI tem sido aplicado em diversos trabalhos na literatura, tais como: Fortin (2008); Edoka (2012); Le e Booth (2014); Borah e Basu (2013); Rubil (2013); Fisher e

Marchand (2014); Hirsch e Winters (2014); Kassenboehmer e Sinning (2014); Morin (2015); e Nguyen (2015). Segundo Firpo, Fortin e Lemieux (2007), existem duas formas de se obter o efeito de uma covariável no quantil incondicional. A primeira delas seria usar as estimativas dos coeficientes da RQ para recuperar o impacto incondicional, que, apesar de intuitivamente simples, muitas vezes se torna intratável. Especificamente, mostra-se que o efeito parcial de uma covariável em um quantil incondicional de Y pode ser escrito como uma média ponderada – por meio da distribuição de X – do efeito parcial sobre um quantil condicional específico de Y , que corresponde ao quantil incondicional de interesse. Portanto, se conseguir mapear todos os quantis incondicionais de Y para os quantis condicionais correspondentes sob diferentes argumentos condicionado, então tal abordagem ponderada pode ser facilmente implementada. No entanto, é evidente que isso é uma tarefa árdua, pois requer técnicas não paramétricas, tornando-a intratável.

Uma alternativa para a abordagem mencionada anteriormente foi proposta por Machado e Mata (2005), que assumiram uma mudança na distribuição incondicional ao longo do tempo e a decompuseram em componentes que são atribuíveis a mudanças na distribuição marginal de diferentes X s. Contudo, esse método captura o efeito total de uma alteração na distribuição marginal de X sobre todos os quantis incondicionais de Y , mas não apenas em um quantil específico.

A segunda forma de se obter o efeito de uma covariável no quantil incondicional é solucionar o problema de intratabilidade e superar a limitação do modelo condicional. Firpo, Fortin e Lemieux (2007) sugerem um modelo RQI baseado nos conceitos de FI e na FIR, tal como utilizado na robusta literatura estatística (Hampel *et al.*, 1986). Uma FI é uma ferramenta analítica que pode ser usada para avaliar o efeito (influência) de uma remoção/adição de uma observação sobre o valor de uma estatística, $v(F)$, sem ter de recalculá-la, e é definida por:

$$(y, v(F)) = \lim_{\varepsilon \rightarrow 0} \left\{ \frac{v[(1 - \varepsilon) \cdot F + \varepsilon \cdot \delta_y] - v(F)}{\varepsilon} \right\}, 0 \leq \varepsilon \leq 1, \quad (1)$$

em que F representa a função de distribuição acumulada de Y e δ_y é uma distribuição que é agregada no valor y .

Uma FIR é obtida adicionando a estatística $v(F)$ a FI:

$$FIR(y, v) = v(F) + FI(y, v). \quad (2)$$

Uma característica dessa abordagem é que a expectativa é igual a $v(F)$ – ou seja, se a estatística de interesse for a média, então a FI será o residual avaliado em y e a FIR será o próprio valor de y :

$$FI(y, \mu) = \lim_{\varepsilon \rightarrow 0} \left\{ \frac{(1 - \varepsilon) \cdot \mu + \varepsilon \cdot y - \mu}{\varepsilon} \right\} = y - \mu, \text{ e } FIR(y, \mu) = \mu + y - \mu = y. \quad (3)$$

Por conseguinte, a regressão da FIR para a média, em X , produz os mesmos coeficientes da regressão por MQO.

Quando a estatística de interesse é um quantil τ específico da distribuição, tem-se:

$$\begin{aligned} &FI(y, q_\tau) \\ &= \frac{\tau - I\{Y \leq q_\tau\}}{f_Y(q_\tau)}, \end{aligned} \quad (4)$$

em que q_τ se refere ao quantil τ th da distribuição incondicional de Y , $f_Y(q_\tau)$ é a função densidade de probabilidade de Y avaliada nesse quantil (q_τ) e $I\{Y \leq q_\tau\}$ é uma variável indicadora que delimita a ocorrência de Y até o limite do quantil. Por definição,

$$\begin{aligned} FIR(y, q_\tau) &= q_\tau \\ &+ FI(y, q_\tau). \end{aligned} \quad (5)$$

Quando a esperança condicional da $FIR(y, q_\tau)$ é modelada como uma função das variáveis explicativas, $X, E[FIR(y, q_\tau)|X = x] = m_\tau(x)$, uma regressão FIR pode ser vista como uma RQI (Firpo, Fortin e Lemieux, 2009). Por definição da FIR, $E_X E[FIR(y, q_\tau)|X] = q_\tau$; então, $E_X(dm_\tau(x)/dX)$ pode ser interpretado como efeito marginal de uma variação na distribuição das covariadas no τ quantil incondicional de Y , mantendo tudo o mais constante.

Firpo, Fortin e Lemieux (2007) afirmam que os procedimentos para estimação da RQI são simples e semelhantes à regressão MQO e se dão da seguinte forma: primeiro, para um quantil τ específico, estima-se a FIR do quantil τ de Y seguindo (3.2.4) e (3.2.5). q_τ é estimado usando a estimativa da amostra do quantil τ incondicional, assim como a densidade $f_Y(q_\tau)$ no ponto q_τ é estimada usando o método de Kernel. Segundo, aplica-se MQO na $FIR(y, q_\tau)$ sobre as covariadas observadas, X . Os autores também destacam as medidas para calcular o efeito quantílico incondicional, que mede o impacto de uma covariável sobre Y em um quantil específico, no qual, na regressão FIR-MQO, se pressupõe que os efeitos parciais quantílicos incondicionais são os coeficientes estimados.⁷

Dessa forma, enquanto a regressão quantílica condicional provê estimativas do retorno de características individuais, *coeteris paribus*, em que este retorno é variável entre os indivíduos de acordo com o quantil incondicional ao qual ele pertence, a regressão quantílica incondicional estima, *coeteris paribus*, o efeito

7. Podem ser implementados pelo *software* Stata a partir do comando *rifreg*. Disponível em: <<https://bit.ly/34DnPU5>>.

de pequenas mudanças em uma característica dos indivíduos em cada quantil da distribuição, permitindo, assim, avaliar o efeito sobre diversas estatísticas da distribuição (Fournier e Koske, 2012).

Uma característica do método RQI é que os resultados podem ser diretamente aplicados ao modelo de decomposição de Oaxaca-Blinder,⁸ o qual será utilizado por este trabalho para verificar os fatores que contribuem para os diferenciais de notas no Enem entre os alunos das escolas públicas profissionalizantes e os estudantes das escolas privadas do Ceará ao longo de toda a distribuição, definidos genericamente como A e B, aplicando-se um método de decomposição proposto por Firpo, Fortin e Lemieux (2007). Generalizando-se a decomposição de Oaxaca-Blinder, pode-se representar a diferença nas distribuições de notas entre os dois tipos de escolas por meio de alguma estatística dessas distribuições. Assim, o diferencial total entre as distribuições é escrito como:

$$\Delta^v = v(F_{yA}) - v(F_{yB}), \quad (6)$$

em que $v(F_{yr})$ representa uma estatística da distribuição de notas para a escola $r = A, B$. A decomposição irá dividir Δ^v em um componente associado a diferenças de características observáveis dos indivíduos e um componente associado a diferenças na estrutura de notas entre os dois tipos de escolas. Para tanto, faz-se necessária a construção de um contrafactual (F_{yC}), para que se possa simular uma distribuição de notas, compondo a estrutura de notas da escola A, e de características – observáveis e não observáveis – dos alunos da escola B. Construindo-se esse contrafactual e algumas manipulações algébricas, pode-se obter:

$$\Delta^v = [v(F_{yB}) - v(F_{yC}) + [v(F_{yC}) - v(F_{yA})]]_e \quad (7)$$

$$\Delta^v = \Delta_S^v + \Delta_X^v. \quad (8)$$

O primeiro termo constitui a *parte não explicada* da decomposição (efeito coeficiente (retorno) ou discriminação quando se refere a mercado de trabalho, definido como uma medida da diferença entre os coeficientes estimados das covariadas para o grupo B em relação ao A). O último termo deste diferencial total é o efeito quantidade, considerado a *parte explicada* da decomposição (efeito dotação ou característico, o qual se refere à diferença nas dotações dos dois grupos, em que a diferença de nota entre os alunos das escolas públicas profissionalizantes perante os das privadas pode ser parcialmente atribuída ao fato de possuírem características mais favoráveis para atingir notas mais elevadas).

8. A regressão FIR quantílica pode ser usada na decomposição de Oaxaca através do comando `oaxaca8` no *software* Stata.

Aplicando-se o valor esperado das funções de influência recentrada e assumindo-se uma especificação linear, têm-se as equações (9) e (10) para as distribuições observadas e para a distribuição contrafactual, respectivamente.

$$v(F_{y_B}) = E[FIR(y_r; v_r)|X, R = r] = X_r \beta_r \text{ para } r = A, B \quad (9)$$

$$\begin{aligned} v(F_{y_C}) &= E[FIR(y_A; v_C)|X, R = B] \\ &= X_C \beta_C \end{aligned} \quad (10)$$

Ao aplicar a decomposição de Oaxaca-Blinder sob a suposição de linearidade, a expectativa da estatística contrafactual em que se adota a distribuição de características da escola B e a estrutura de notas da escola A é estimada de forma que $\bar{X}_C \beta_C = \bar{X}_B \beta_A$. Entretanto, quando a linearidade não se mantém, essa igualdade passa a não ser válida.

A partir de Barsky *et al.* (2002), Firpo, Fortin e Lemieux (2007) propõem uma abordagem que combina o método de regressões FIR com reponderações inspiradas na técnica de DiNardo, Fortin e Lemieux (1996). Esse método estabelece o uso de um fator de reponderação, o qual pode ser definido por:

$$\varphi(X) = \frac{\Pr[R = B|X]}{\Pr[R = A|X]} \cdot \frac{\Pr[R = A]}{\Pr[R = B]}, \quad (11)$$

em que $\Pr[R = r|X]$ é a probabilidade de um aluno pertencer à escola r ; dadas suas características, X e $\Pr[R = r]$ denotam a proporção de alunos na escola r . Vale ressaltar que esse fator de ponderação pode ser estimado através de um modelo de probabilidade para $\Pr[R = B|X]$, usando as probabilidades previstas para calcular o valor $\widehat{\varphi}(X)$ para cada observação:

$$\widehat{\beta}_r = (\sum_{i \in r} \widehat{w}_r \cdot X_i \cdot X_i')^{-1} \cdot \sum_{i \in r} \widehat{w}_r \cdot FIR(\widehat{y}_{ri}; v_r) \cdot X_i, \text{ para } r = A, B, \quad (12)$$

em que \widehat{w}_r é um fator de ponderação baseado na própria composição da amostra total. Para o contrafactual, estima-se:

$$\widehat{\beta}_C = (\sum_{i \in A} \widehat{\varphi}(X_i) \cdot X_i \cdot X_i')^{-1} \cdot \sum_{i \in A} \widehat{\varphi}(X_i) \cdot FIR(\widehat{y}_{Ai}; v_C) \cdot X_i \quad (13)$$

$$\bar{X}_C = \sum_{i \in A} \widehat{\varphi}(X_i) \cdot X_i \quad (14)$$

A decomposição é dada por:

$$\widehat{\Delta}^v = [\bar{X}_B \widehat{\beta}_B - \bar{X}_C \widehat{\beta}_C] + [\bar{X}_C \widehat{\beta}_C - \bar{X}_A \widehat{\beta}_A] \quad (15)$$

$$\hat{\Delta}^v = \hat{\Delta}_S^v + \hat{\Delta}_X^v. \quad (16)$$

A estimativa do efeito composição $\hat{\Delta}_S^v$ pode ser dividida em dois componentes em que um termo padrão $(\bar{X}_C - \bar{X}_A)\hat{\beta}_A$ é somado a um erro de especificação $\bar{X}_C(\hat{\beta}_C - \hat{\beta}_A)$, de forma que:

$$\hat{\Delta}_X^v = (\bar{X}_C - \bar{X}_A)\hat{\beta}_A + \bar{X}_C(\hat{\beta}_C - \hat{\beta}_A). \quad (17)$$

O erro de especificação tende a ser nulo, se a especificação linear do modelo for correta, uma vez que $plim(\hat{\beta}_C) = plim(\hat{\beta}_A)$. Na decomposição detalhada, em que é possível estimar a contribuição de cada variável explicativa, o efeito composição pode ser escrito como:

$$\hat{\Delta}_X^v = \sum_{k=1}^K (\bar{X}_{Ck} - \bar{X}_{Ak})\hat{\beta}_A. \quad (18)$$

O efeito estrutural das notas passa a ser definido por:

$$\hat{\Delta}_S^v = \bar{X}_B(\hat{\beta}_B - \hat{\beta}_C) + (\bar{X}_B - \bar{X}_C)\hat{\beta}_C. \quad (19)$$

Esse efeito se reduz ao primeiro termo, visto que, em grandes amostras e com a estimação correta do fator de ponderação, o erro de reponderação tende a ser nulo. Destaca-se que o método de regressões quantílicas incondicionais fornece uma estimativa de equilíbrio parcial da variável de interesse, assumindo que mudanças marginais em X não exercem impacto sobre a distribuição conjunta de X e y . Além disso, o método também reconhece a forte independência entre heterogeneidade não observada e as características observadas. De acordo com Fournier e Koske (2012), embora essas suposições não se sustentem na prática, elas argumentam que uma comparação entre estimativas de diferentes quantis permanece válida, desde que o viés potencial seja igual em toda a população da amostra.

2.1 Base de dados

As informações referentes ao desempenho nas cinco áreas de conhecimento e na média geral do Enem de 2014,⁹ no tocante às condições socioeconômicas dos alunos das EEEPs da rede estadual de educação e das escolas privadas do Ceará, foram retiradas do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). A identificação das escolas profissionalizantes e de suas dotações de infraestrutura de todas as escolas ocorreu por meio de dados da Secretaria da Educação do Estado do Ceará (Seduc) e do Censo Escolar de 2014, respectivamente.

A amostra utilizada neste trabalho é constituída apenas pelos alunos com as seguintes características: estão cursando o ensino médio das escolas de ensino

9. Último ano com informações disponíveis sobre o desempenho individual no Enem.

profissionalizante e privado do estado;¹⁰ concluíram o ensino médio em 2014;¹¹ estiveram presentes nos dias de provas do Enem do referido ano; e suas escolas possuem informações disponibilizadas pelo Censo Escolar de 2014. Desse modo, obteve-se um total de 20.947 alunos, sendo 8.955 da rede estadual de ensino profissional e 11.992 das escolas privadas.

2.2 Variáveis do modelo

As variáveis, descritas no quadro 1, foram escolhidas com o intuito de captar os diferenciais nas notas das cinco áreas e na média geral do Enem, compostas por vetores com as características pessoais dos alunos, o *status* educacional dos pais e as condições de infraestrutura das escolas.

QUADRO 1
Descrições das variáveis

Variáveis	Descrição
Dependentes: notas logaritimizadas do aluno em cada uma das áreas: ciências da natureza (CN); ciências humanas (CH); linguagens e códigos (LC); matemática (MT); redação (R); e média geral (MG).	
Explicativas	
Alunos	
Priv	1, se o aluno pertence à escola privada; 0, se o aluno pertence à EEEP
Idade	Anos de idade
Sexo	1, se masculino; 0, caso contrário (c.c.)
Raça	1, se branca; 0, c.c.
Estado civil	1, se solteiro; 0, c.c.
Anos_Fund	1, se concluiu o ensino fundamental em oito anos; 0, c.c.
Anos_méd	1, se concluiu o ensino médio em três anos; 0, c.c.
Escolaridade do pai	
E_Pai1	1, se o pai não estudou; 0, c.c.
E_Pai2	1, se o pai possui de escolaridade o ensino fundamental incompleto ou completo; 0, c.c.
E_Pai3	1, se o pai possui de escolaridade o ensino médio incompleto ou completo; 0, c.c.
E_Pai4	1, se o pai possui de escolaridade o ensino superior incompleto ou completo; 0, c.c.
Escolaridade da mãe	
E_mae1	1, se a mãe não estudou; 0, c.c.
E_mae2	1, se a mãe possui de escolaridade o ensino fundamental incompleto ou completo; 0, c.c.
E_mae3	1, se a mãe possui de escolaridade o ensino médio incompleto ou completo; 0, c.c.
E_mae4	1, se a mãe possui de escolaridade o ensino superior incompleto ou completo; 0, c.c.

(Continua)

10. Foram excluídos os alunos das escolas estaduais de ensino regular, municipais e federais.

11. Foram excluídos os alunos que já concluíram o ensino médio ou que não irão concluir em 2014.

(Continuação)

Variáveis	Descrição
Condições socioeconômicas	
Renda00	1, se a família não possui renda; 0, c.c.
Renda01	1, se a família possui até 1 salário mínimo (SM); 0, c.c.
Renda02	1, se a família possui de 1 até 2 SMs; 0, c.c.
Renda03	1, se a família possui de 2 até 5 SMs; 0, c.c.
Renda04	1, se a família possui de 5 até 10 SMs; 0, c.c.
Renda05	1, se a família possui de 10 até 15 SMs; 0, c.c.
CSE ¹	Índice de condição socioeconômica
Escolas	
Infra ²	Índice de infraestrutura da escola

Elaboração dos autores.

Notas: ¹ Índice criado a partir da análise de componente principal, que engloba a existência das seguintes características dos alunos obtidas pelo Enem de 2013: se possui televisão, DVD, rádio, microcomputador, automóvel, máquina de lavar, geladeira, freezer, telefone fixo, celular, acesso à internet, televisão por assinatura, aspirador de pó, empregada e banheiro. ² Índice gerado pela análise de componentes principais, que engloba a existência das seguintes características das escolas obtidas no Censo Escolar de 2013: abastecimento de água por rede pública; esgoto sanitário por rede pública; coleta periódica de lixo; sala de diretoria; sala de professores; laboratório de informática e de ciências; quadra de esportes; cozinha; biblioteca; parque infantil; sanitário adequado a alunos portadores de necessidades especiais (PNEs); dependências e vias adequadas a PNEs; televisão; videocassete; DVD; parabólica; copiadora; retroprojeto; impressora; computadores; acesso à internet; e alimentação escolar para os alunos.

A variável que reflete a renda familiar está disponível em intervalos de SMs; nesse contexto, optou-se por fazer um corte de dois salários, tendo em vista que são pais de alunos de escolas públicas, pressupondo-se, assim, não possuírem renda familiar muito elevada. Para facilitar a interpretação da medida dos índices CSE e Infra, as variáveis foram normalizadas, de forma a assumir valores entre 0 a 1; quanto maior for o índice, mais favorável será a disposição de equipamentos.

3 RESULTADOS

3.1 Análise descritiva

Estatísticas descritivas das notas nas cinco áreas e na média geral do Enem e das variáveis de controles relacionadas aos alunos de EEEPs (públicas profissionalizante), regulares (públicas não profissionalizantes) e privadas do Ceará em 2014 estão explicitadas na tabela 1. Observa-se que as médias das notas dos alunos das escolas de ensino privado são superiores às dos alunos de escolas profissionalizantes, e estas são maiores que as médias dos alunos das escolares regulares, em todas as provas. Além disso, as demais características dos alunos das EEEPs se aproximam mais das características dos alunos das escolas particulares. Embora seja apenas a média, esses resultados corroboram a hipótese deste trabalho de que os alunos das escolas profissionalizantes apresentam desempenho superior aos alunos das escolares regulares, e, ao utilizar esse grupo para representar as escolas públicas, isso tornará os grupos mais homogêneos. Fazendo a comparação das escolas de

ensino profissional com as privadas, verifica-se que, na área de ciências da natureza, o diferencial de médias é de 52,24 pontos a favor das escolas privadas, embora estas apresentem uma variação relativa à média 21,7% (0,1572/0,1291) superior à EEEP. As áreas de ciências humanas e linguagens e códigos revelam diferenças semelhantes, 37,39 e 32,67 pontos, respectivamente, com dispersão relativa à média de 4% e 3% superior à EEEP, respectivamente.

Apresentando diferenciais mais elevados e menos homogêneos estão as áreas de matemática e redação, com diferença de 71,92 e 83,97 pontos, respectivamente, e variação relativa semelhante nos dois tipos de escolas, e de 23,5% em matemática, superior à EEEP. Na média geral, os alunos das escolas particulares apresentaram desempenho superior em 55,6 pontos, 11% a mais que os alunos das escolares de ensino profissionalizante, apesar de estas deterem menor coeficiente de variação.

TABELA 1
Estatísticas descritivas das variáveis da amostra (2014)

Variáveis	Regulares			EEEPs			Privadas		
	Média	DP ¹	CV ²	Média	DP	CV	Média	DP	CV
NOTA_CN	455.52	57.316	0.1258	493.58	63.749	0.1291	545.82	85.8326	0.1572
NOTA_CH	509.365	67.124	0.1318	556.17	63.8829	0.1148	593.56	70.9712	0.1195
NOTA_LC	479.3558	64.26	0.1341	519.56	58.945	0.1134	552.23	64.4928	0.1167
NOTA_MT	434.7047	76.527	0.1760	486.98	95.5583	0.1962	558.9	135.454	0.2423
NOTA_R	436.2812	144.381	0.3309	549.3	139.2176	0.2534	633.27	158.1479	0.2497
NOTA_MG	463.0465	62.199	0.1343	521.12	64.5028	0.1237	576.76	85.6248	0.1484
Alunos									
Idade	18.86	4.43	0.2349	17.467	0.8507	20.532	17.643	2.6064	6.7691
Sexo	0.445	0.4865	1.0933	0.424	0.4942	1.1655	0.4654	0.4988	1.0716
Raça	0.1677	0.3736	2.2278	0.1934	0.3949	2.0422	0.4237	0.4941	1.1661
Estado civil	0.9545	0.2083	0.2182	0.9946	0.073	0.0734	0.9908	0.0953	0.0962
Anos_Fund	0.3663	0.4818	1.3153	0.3794	0.4852	1.2788	0.459	0.4983	1.0855
Anos_Méd	0.3271	0.4691	1.4341	0.357	0.4791	1.3421	0.527	0.4992	0.9473
Escolaridade do pai									
E_Pai1	0.1206	0.3257	2.7007	0.0692	0.2538	3.6667	0.0149	0.1212	8.124
E_Pai2	0.663	0.4726	0.7128	0.5786	0.4937	0.8533	0.2317	0.4219	1.8208
E_Pai3	0.1844	0.3878	2.1030	0.2944	0.4558	1.5479	0.394	0.4886	1.2399
E_Pai4	0.0277	0.1641	5.9242	0.0501	0.2182	4.3527	0.2655	0.4416	1.6629
Escolaridade da mãe									
E_mae1	0.0685	0.2527	3.6891	0.0279	0.1647	5.9011	0.0068	0.0824	12.0522
E_mae2	0.6507	0.4767	0.7326	0.4943	0.4999	1.0113	0.1571	0.3639	2.3156
E_mae3	0.2204	0.4145	1.8807	0.3563	0.4789	1.344	0.3922	0.4882	1.2447
E_mae4	0.0466	0.2108	4.5236	0.0904	0.2868	3.1712	0.2952	0.4561	1.5449

(Continua)

(Continuação)

Variáveis	Regulares			EEEPs			Privadas		
	Média	DP ¹	CV ²	Média	DP	CV	Média	DP	CV
Condições socioeconômicas									
Renda00	0.0306	0.1723	5.6307	0.0155	0.1236	7.9643	0.0049	0.0699	14.2223
Renda01	0.6899	0.4625	0.6704	0.6391	0.4802	0.7513	0.1323	0.3388	2.5606
Renda02	0.2279	0.4195	1.8407	0.2797	0.4488	1.6047	0.2635	0.4405	1.6718
Renda03	0.0469	0.2115	4.5096	0.0599	0.2374	3.9595	0.299	0.4578	1.5311
Renda04	0.0039	0.0624	16.000	0.0052	0.0722	13.767	0.1752	0.3802	2.1692
Renda05	0.0003	0.0178	59.333	0.0002	0.0149	66.922	0.0617	0.2407	3.8967
CSE	0.3777	0.1513	0.4006	0.4806	0.1461	0.304	0.674	0.1847	0.274
Escolas									
Infra	0.2247	0.0637	0.2835	0.3547	0.0903	0.2546	0.4144	0.149	0.3597

Elaboração dos autores.

Notas: ¹ DP – desvio-padrão.² CV – coeficiente de variação.

Obs.: 1. A proporção de alunos com pai que possui pós-graduação (E_pai5) foi 0.007 para as profissionalizantes e 0.093 para as privadas, perfazendo 100% quando somada às demais.

2. A proporção de alunos com mãe que possui pós-graduação (E_mae5) foi 0.030 para as profissionalizantes e 0.148 para as privadas, perfazendo 100% quando somada às demais.

3. A proporção de alunos com renda familiar maior que 15 SMs (renda_6) foi 0.0001 para as profissionalizantes e 0.0631 para as privadas, perfazendo 100% quando somada às demais.

Os gráficos A.2 a A.7, no apêndice A, mostram a distribuição das proficiências dos alunos nas cinco áreas e na média geral para as escolas particulares, profissionalizantes e regulares. Por meio do gráfico A.7, pode-se notar que a curva para os alunos do ensino privado se encontra mais concentrada para a direita, indicando que a frequência de alunos com notas maiores é superior aos alunos pertencentes ao ensino profissionalizante, e estes estão melhor distribuídos que os alunos das escolas regulares.

Referente às demais variáveis presentes na tabela 1, observa-se que, apesar de apresentar idade média semelhante, 17 anos, o grupo das escolas privadas é menos homogêneo ao das escolas profissionalizantes, pois apresenta variação absoluta de, aproximadamente, dois anos, enquanto, no outro grupo, o desvio-padrão é de meses. Além disso, a proporção de homens e alunos solteiros é semelhante: nas EEEP's e nas escolas privadas, 42% e 46% dos alunos são do sexo masculino, respectivamente, e, em ambos os grupos, 99% são solteiros. Quando se refere à raça, a diferença entre os percentuais aumenta: 19,3% dos estudantes são considerados brancos nas escolas profissionalizantes e 42,3%, nas particulares.

As variáveis *Anos_Fund* e *Anos_Médio* apresentam a proporção de alunos que terminaram o ensino fundamental em oito anos e o médio em três anos, os quais, para as escolas profissionalizantes, foram de 38% e 36%, e, para as particulares, 46% e 53%, respectivamente. Isso mostra indícios que os alunos que fizeram o

3º ano do ensino médio nas escolas profissionalizantes apresentam problemas de repetência maiores do que as escolas particulares.

Ao se analisar as escolaridades do pai e da mãe, percebe-se que os alunos das escolas privadas apresentam pais com níveis mais elevados de estudo. Para as EEEPs, a maior proporção reside em pai e mãe com fundamental incompleto ou completo, 57,86% e 49,43%, respectivamente, enquanto, para as escolas regulares não profissionalizantes, pais com ensino médio incompleto ou completo são maioria, aproximadamente 39%. Ademais, a proporção de pais com ensino superior nas escolas de ensino privado é maior do que no grupo das profissionalizantes, aproximadamente 20 pontos percentuais (p.p.) a mais.

Através das variáveis de condições socioeconômicas, observa-se que a renda familiar dos alunos das escolas profissionalizantes se concentra em até 2 SMs, ao passo que, nas escolas privadas, a concentração se dá a partir de 3 SMs. Além disso, os alunos destas escolas apresentam um índice CSE médio mais elevado do que o observado para as EEEPs, 0.6740 e 0.4806, respectivamente. Apesar de as EEEPs apresentarem boa infraestrutura (Ceará, 2014), as escolas particulares detêm uma melhor oferta de equipamentos aos alunos, conforme atestam os respectivos índices, com valores de 0.3547 e 0.4144.

3.2 Regressão FIR nos quantis incondicionais

Para decompor os diferenciais de notas entre os alunos das escolas regulares e profissionalizantes e entre os alunos destas e das privadas por meio da metodologia de Firpo, Fortin e Lemieux (2007), faz-se necessário estimar as regressões FIR as quais têm o objetivo de estimar efeitos das covariáveis sobre os quantis incondicionais do *log* das notas das provas do Enem, cujos resultados para os quantis 0.1, 0.5 e 0.9 se encontram nas tabelas do apêndice B. Além disso, essas tabelas apresentam as estimativas por MQO. Pode-se observar que os efeitos não são constantes ao longo da distribuição, justificando, assim, a análise por regressões quantílicas,¹² além de diferirem entre os tipos de escolas (regulares, profissionalizantes e particulares).

A análise desses resultados se concentra na comparação das escolas profissionalizantes e particulares. Os coeficientes estimados para a variável *idade* foram negativos e significativos em todos os quantis analisados e áreas de conhecimentos para ambos os tipos de ensino. Isso indica que quanto maior a idade do aluno, menor será seu desempenho no Enem, independentemente da parte da distribuição na qual o aluno se encontra. Tal efeito decresce no sentido da parte superior da distribuição.

12. O teste de Koenker e Bassett (1982) mostrou que os coeficientes são estatisticamente diferentes, apresentando uma estatística *F* igual a 25.58.

Os resultados mostram que o efeito gênero difere entre os tipos de escolas, ao longo da distribuição e entre as áreas de conhecimento. Na média geral, nas áreas de ciências humanas, nas ciências da natureza e em matemática, considerando-se a mediana e o quantil 0.9, pode-se observar que o efeito é positivo e significativo em ambas as escolas, embora com efeito maior nas profissionalizantes, enquanto, no quantil 0.1, o efeito difere entre as áreas e entre as escolas. Na média geral, este foi insignificante para as EEEPs e negativo significativo para as escolas particulares. Na área de ciências humanas, o efeito foi positivo e significativo para as profissionalizantes e insignificante para as privadas. Já nas áreas de ciências da natureza e matemática, ambos foram positivos e significativos. Apresentando resultados negativos e significativos em todos os quantis analisados e nos dois tipos de escolas, encontra-se a área de redação; por sua vez, em linguagens e códigos, esse tipo de efeito se sobressai apenas para as escolas particulares, na mediana e no quantil 0.9.

Referente à raça, verificou-se que, no primeiro quantil, o efeito é predominantemente insignificante, exceto em redação, o qual foi positivo significativo para os tipos de escolas. Na mediana, podem-se observar, na média geral e nas áreas de linguagens e códigos, matemática e redação, efeitos insignificantes para as EEEPs e positivos significativos para as privadas, enquanto nas áreas de ciências humanas e ciências da natureza os efeitos foram positivos e significantes nas EEEPs; nas escolas particulares, foram insignificantes e positivos significativos, respectivamente. No quantil 0.9, esse efeito se mostrou positivo e significativo em todas as áreas para as escolas de ensino privado e nas áreas de ciências humanas e linguagens e códigos para as EEEPs.

Além desses resultados, nota-se que alunos cujo estado civil é solteiro não apresentam efeito significativo sobre as notas do Enem no primeiro quantil da distribuição, na mediana, exceto em linguagens e códigos para as EEEPs e em matemática para as escolas particulares. Esse também é o caso referente ao quantil 0.9, nas áreas de ciências humanas e linguagens e códigos para ensino privado, bem como em todas as áreas nas EEEPs, exceto em ciências humanas. Diferentemente da variável que indica se o aluno concluiu o ensino fundamental em oito anos, cujo efeito foi positivo e significativo em todos os quantis analisados, áreas de conhecimentos e escolas, exceto no primeiro quantil de matemática para o ensino profissionalizante, cujo efeito foi insignificante. Já para a variável *Anos_medio*, o efeito foi insignificante para as EEEPs, em todas as áreas e quantis, exceto no quantil 0.1, em redação.

Referente às *dummies* de escolaridade dos pais dos alunos de ensino profissionalizante, observa-se que todos os efeitos foram insignificantes, nas partes inferiores e superiores da distribuição de notas, em todas as áreas, e na mediana, exceto na variável que indica se o pai tem ensino médio incompleto

ou completo para as áreas de conhecimento de média geral e de ciências da natureza, e nas variáveis E_{pai2} e E_{pai4} em ciências da natureza, cujos efeitos foram positivos. Porém, nas escolas privadas, verificam-se, predominantemente, efeitos negativos e significativos, exceto para as variáveis E_{pai1} , nos quantis 0.9 da média geral e 0.1 das ciências humanas, das ciências da natureza e de matemática, e E_{pai3} , na mediana em linguagens e códigos e no quantil 0.1 de linguagens e códigos, enquanto, para E_{pai4} , o efeito se mostrou, na maioria dos casos, insignificante.

Resultado semelhante pôde ser constatado para as escolaridades das mães dos alunos das escolas profissionalizantes, com exceção da média geral para E_{mae2} e E_{mae3} , no quantil 0.1, e da mediana, para E_{mae3} na média geral e E_{mae2} nas ciências da natureza. Já no ensino privado, os efeitos variaram entre insignificantes e negativos significativos, tanto por quantil quanto por área de conhecimento.

Quanto ao efeito das faixas de renda sobre as variações nas notas do Enem, verifica-se que, a partir da média geral, para as escolas profissionalizantes, todos os efeitos foram positivos e significativos, exceto na mediana para renda de 10 a 15 SMs e no quantil 0.9, para famílias que não possuem renda e para aquelas que possuem de 5 a 15 SMs. Nas escolas particulares, todavia, o efeito foi predominantemente negativo e significativo, com exceção de algumas faixas de renda no primeiro quantil e na mediana.

Ademais, efeitos insignificantes puderam ser observados ao longo da distribuição para o índice de condição socioeconômica e de infraestrutura das escolas, considerando a média geral para os alunos das escolas profissionalizantes, enquanto, nas escolas privadas, esse efeito foi positivo e significativo em todos os quantis analisados.

3.3 Decomposição da regressão FIR: efeito composição e estrutural

A decomposição tradicional de Oaxaca-Blinder e quantílica irá identificar quais fatores e efeitos composição (característicos) ou estrutural (coeficientes) determinam os diferenciais de notas por tipo de escola (regulares e profissionalizantes; privadas e EEEPs), na média e nos quantis da distribuição, utilizando as estimativas de MQO e FIR, respectivamente. Para tanto, esse procedimento foi estimado para os dezoito quantis estabelecidos, 0.05 a 0.95, cujos diferenciais de notas regulares e profissionalizantes se encontram nos gráficos A.8 a A.13, no apêndice A; os diferenciais entre as escolas privadas e as EEEPs, juntamente com os efeitos característicos e coeficientes nos quantis, são ilustrados através dos gráficos A.2 a A.7, embora somente alguns quantis foram selecionados e expostos na tabela 2, somados aos resultados da decomposição na média. Em todas as notas, as estimativas dos

diferenciais mostraram-se significativas a 1%; assim, é possível verificar vantagens dos alunos das escolares particulares em relação aos alunos do ensino profissional, bem como destas sobre os alunos das escolas regulares.¹³

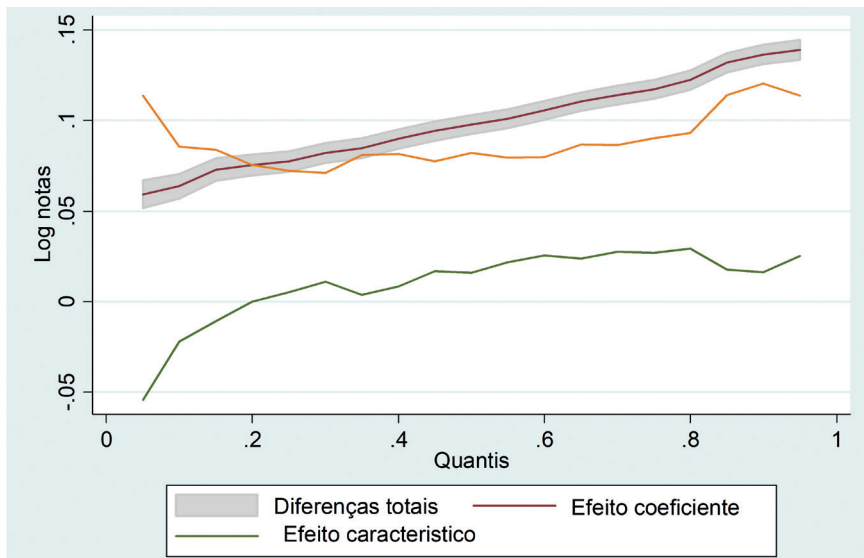
A partir da tabela 2, em matemática, nas ciências da natureza e na média geral, o aumento da diferença de notas na parte superior da distribuição indica o efeito teto de vidro (*glassceiling*). Verifica-se que o diferencial total entre as notas dos alunos das escolas profissionalizantes e privadas na média geral é monotonamente crescente nos quantis. No quantil 0.1, a diferença é de 0.0637, atingindo 0.0978 na mediana e 0.1365 no quantil 0.9 – ou seja, 6,57%, 10,27% e 14,62% a mais para os alunos das escolas privadas, respectivamente. Dessa forma, pode-se dizer que, em média, a diferença está aumentando à medida que se move a quantis mais elevados da distribuição – isto é, o diferencial de notas a favor dos alunos das escolas particulares é maior para aqueles que apresentam desempenhos superiores.

Interpretação semelhante pode ser utilizada para as áreas de ciências da natureza e matemática, cujo diferencial total também é positivo e crescente. Porém, observa-se que, em matemática, o hiato cresce com o quantil da distribuição. Nessa área, o quantil 0.10 apresenta uma diferença de 5,51% e atinge 21,55% no quantil 0.9, enquanto, em ciências da natureza, essas diferenças são de 6,2% e 13,63%, respectivamente.

Em ciências humanas e em linguagens e códigos, apesar de apresentarem diferenciais positivos, estes praticamente permanecem constantes ao longo da distribuição, com leves oscilações. No 10º, 50º e 90º quantis, as diferenças são 0.0698, 0.0602 e 0.0767, respectivamente, para ciências humanas, e 0.0606, 0.0637 e 0.0636, respectivamente, para linguagens e códigos. Apresentando diferenciais mais elevados ao longo de toda a distribuição, encontra-se a área de redação, na qual há a oscilação de 0.1504, equivalente a 16,22%, no quantil 0.1, para 0.1135 (12,01%) na mediana e 0.1503 (16,21%) no quantil 0.9.

13. Ver apêndice A.

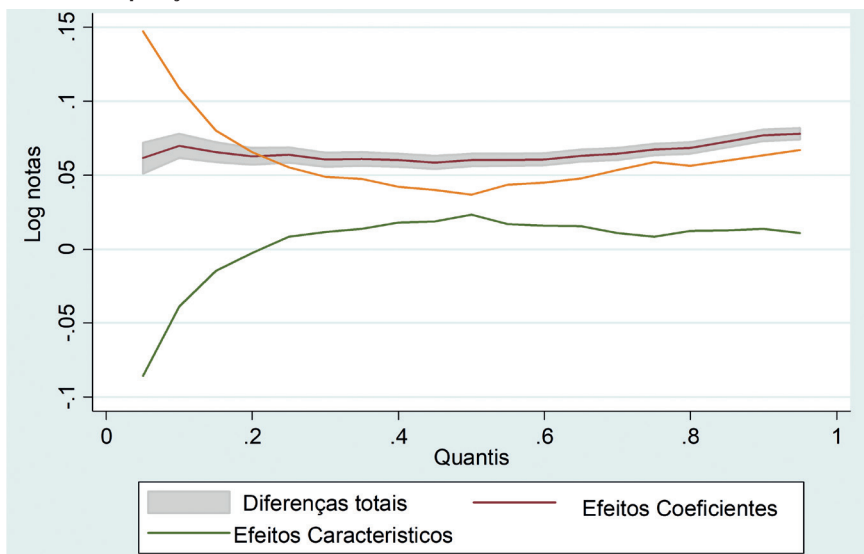
GRÁFICO 3
Decomposição dos efeitos: MG



Elaboração dos autores.

Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

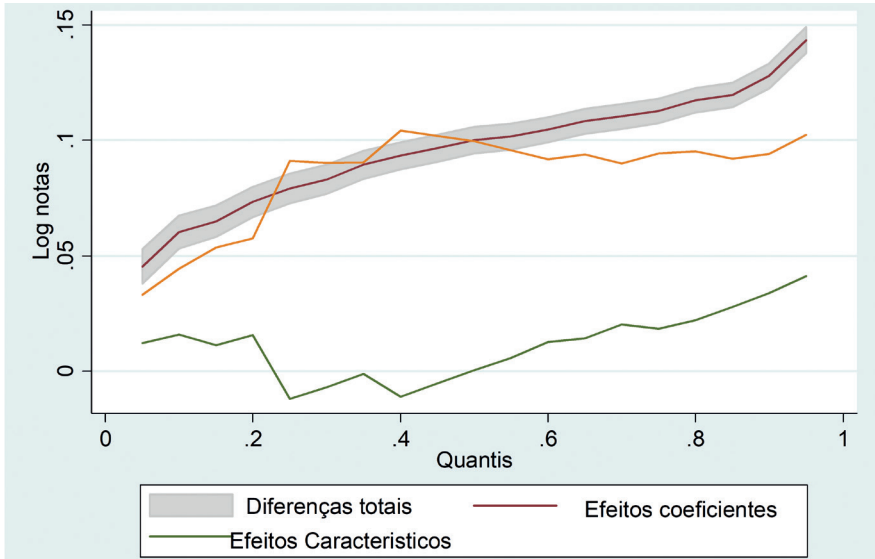
GRÁFICO 4
Decomposição dos efeitos: CH



Elaboração dos autores.

Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

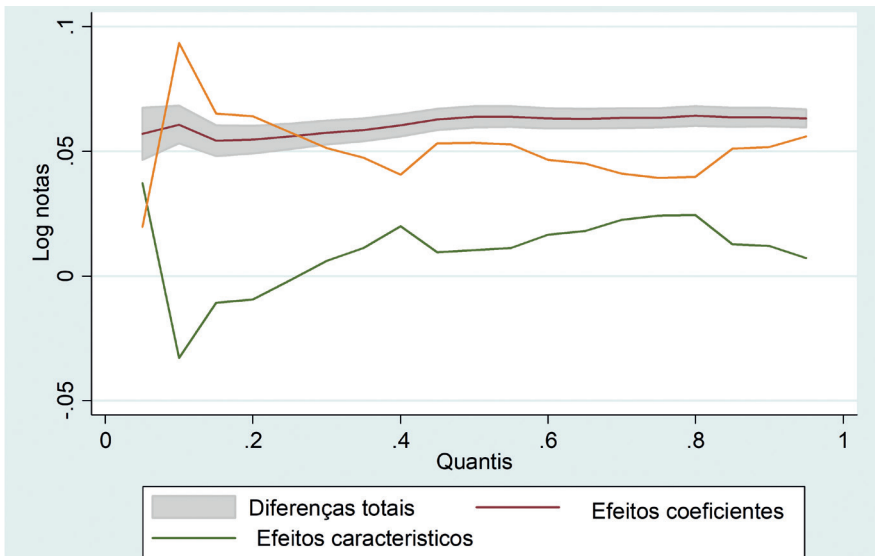
GRÁFICO 5
Decomposição dos efeitos: CN



Elaboração dos autores.

Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

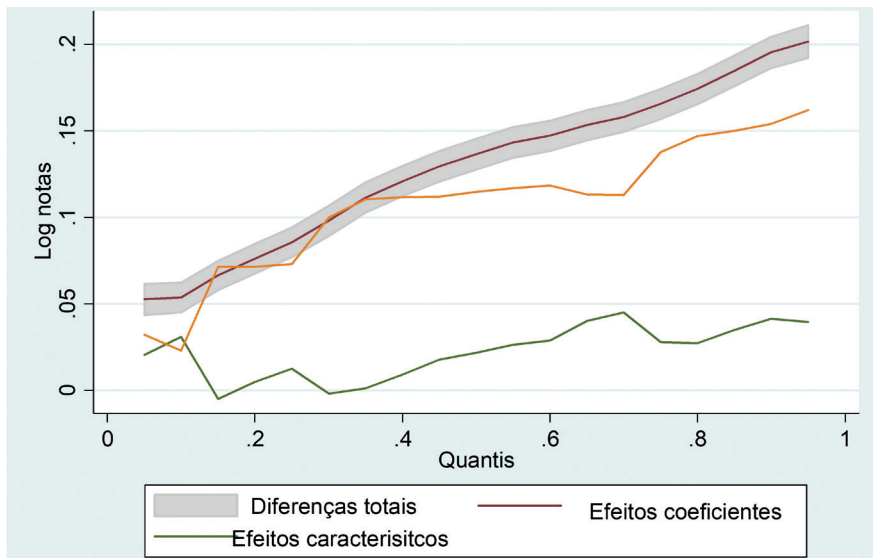
GRÁFICO 6
Decomposição dos efeitos: LC



Elaboração dos autores.

Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

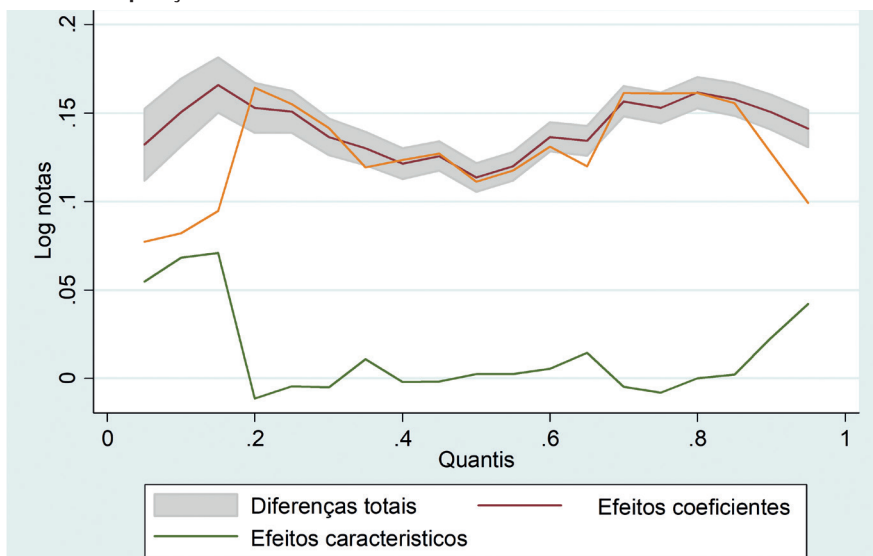
GRÁFICO 7
Decomposição dos efeitos: MT



Elaboração dos autores.

Obs.: Gráfico cujo layout e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

GRÁFICO 8
Decomposição dos efeitos: R



Elaboração dos autores.

Obs.: Gráfico cujo layout e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

Os resultados dos efeitos característicos para a média geral e a área das ciências humanas apresentam sinais negativos e significantes nos primeiros quantis (0.05, 0.10 e 0.15) da distribuição, nos quantis 0.10 e 0.15 de linguagens e códigos e no percentil 0.25 das ciências da natureza, indicando que estes agem no sentido de diminuir as diferenças entre as notas. Isso implica que as diferenças de notas entre as escolas em características observáveis preveem vantagens aos estudantes das EEEPs nessas áreas e na parte inferior da distribuição, com destaque para o quantil 0.1 de ciências humanas e de linguagens e códigos, que agem reduzindo o diferencial em 56% e 54%, respectivamente, como pode ser observado na tabela 2 e nos gráficos 3 a 8.

Dessa forma, os alunos das escolas profissionalizantes apresentam características produtivas melhores que os alunos das escolas particulares, e o efeito positivo para o efeito estrutural sugere que os retornos das características foram maiores para os alunos das escolas particulares. Com isso, nesses quantis, o que explica os diferenciais de notas são os fatores não observáveis, pois os alunos das EEEPs apresentaram atributos observáveis melhor do que os alunos das escolas privadas. Portanto, o sinal positivo da ação dos efeitos estruturais foi suficiente para anular as características (efeito composição).

Os efeitos composição nos quantis 0.2, 0.25, 0.35 e 0.5 da média geral, 0.2 e 0.25 das ciências humanas, 0.2, 0.25, 0.3, 0.45, 0.5 e 0.55 de linguagens e códigos, 0.15 a 0.6 de matemática, 0.15, 0.3 a 0.65 e 0.75 das ciências da natureza e 0.2 a 0.85 da redação não foram estatisticamente significantes. Para os demais, os efeitos composição foram positivos e significantes, atuando de forma a contribuir com o diferencial de notas – ou seja, os alunos das escolas privadas apresentam melhores características que ajudam a obter um bom desempenho, principalmente nos quantis mais elevados da distribuição, pois, em todas as áreas de conhecimento e na média geral, esses efeitos foram positivos e significativos nos percentis 0.9 e 0.95. Já nos menores quantis de matemática e de redação, observam-se os maiores efeitos, 57% e 45%, respectivamente, com destaque para o primeiro, o qual superou o efeito estrutural em contribuir para o aumento do diferencial de notas.

Contudo, os efeitos não observáveis (coeficientes) são todos positivos e significativos em todos os quantis da distribuição e em todas as áreas do Enem, sendo responsáveis por explicar a maior parte do diferencial das notas. Em matemática, o efeito da variável é crescente ao longo da curva; diferentemente das demais áreas, cujos efeitos oscilam entre os percentis. Esses resultados indicam que os alunos das escolas privadas são mais capazes de converter insumos educacionais em maiores pontuações em testes padronizados e que existem outros fatores não observáveis pertencentes a estes alunos que contribuem para melhores desempenhos.

TABELA 2
Decomposição agregada na distribuição das notas nas áreas do Enem

Quantil	MQO	%	0.1	%	0.25	%	0.5	%	0.75	%	0.9	%
MG												
Diferencial	0.0980*		0.0637*	100	0.0774*	100	0.0978*	100	0.1173*	100	0.1365*	100
Características	0.0833*		-0.022*	-35	0.0053	7	0.0158	16	0.0270**	23	0.0161**	12
Coeficientes	0.0146*		0.0857*	135	0.0720*	93	0.0820*	84	0.0903*	77	0.1204*	88
CH												
Diferencial	0.0644*		0.0698*	100	0.0635*	100	0.0602*	100	0.0672*	100	0.0767*	100
Características	0.0574*		-0.039**	-56	0.0084	-	0.0232*	39	0.0083**	12	0.0135*	18
Coeficientes	0.0070*		0.1088*	156	0.0551*	87	0.0369*	61	0.0588*	88	0.0632*	82
CN												
Diferencial	0.0965*		0.0602*	100	0.0791*	100	0.1000*	100	0.1127*	100	0.1278*	100
Características	0.0790*		0.0158*	26	-0.0120*	-15	0.0004	-	0.0184	-	0.0337***	26
Coeficientes	0.0174*		0.0443*	74	0.0912*	115	0.0996*	100	0.0943*	84	0.0940*	74
MT												
Diferencial	0.1273*		0.0537*	100	0.0855*	100	0.1365*	100	0.1655*	100	0.1952*	100
Características	0.1245*		0.0308**	57	0.0126	-	0.0216	-	0.0278*	17	0.0414*	21
Coeficientes	0.0028*		0.0228*	43	0.0728*	85	0.1148*	84	0.1377*	83	0.1538*	79
LC												
Diferencial	0.0605*		0.0606*	100	0.0558*	100	0.0637*	100	0.0633*	100	0.0636*	100
Características	0.0496*		-0.0328*	-54	-0.0018	-	0.0104	-	0.0240**	38	0.0120*	19
Coeficientes	0.0108*		0.0934*	154	0.0577*	103	0.0533*	84	0.0392*	62	0.0516*	81
R												
Diferencial	0.1429*		0.1504*	100	0.1506*	100	0.1135*	100	0.1528*	100	0.1503*	100
Características	0.1090*		0.0683*	45	-0.0044	-	0.0024	-	-0.0079	-	0.0230*	15
Coeficientes	0.0338*		0.0821*	55	0.1550*	103	0.1111*	98	0.1608*	105	0.1273*	85

Elaboração dos autores.

Obs.: Significância (*) = 1%; (**) = 5%; e (***) = 10%.

Os dois efeitos, característico e estrutural, podem ser decompostos de forma detalhada, como mostra a tabela 3 e os gráficos 9 e 10 para a média geral.

TABELA 3
Decomposição dos efeitos característico e retorno: MG

Variáveis	Característico					Estrutural				
	0.1	0.25	0.5	0.75	0.9	0.1	0.25	0.5	0.75	0.9
Total	-0.0222*	0.0053	0.0158	0.0270**	0.0161**	0.0857*	0.0720*	0.0820*	0.0903*	0.1204*
Idade	-0.0061*	-0.0053*	-0.0052*	-0.0037*	-0.0022*	0.2099*	0.2913*	0.3843*	0.3092	0.1951*
Sexo	0.0002	0.0007*	0.0012*	0.0012*	0.0011*	-0.0077*	-0.0093*	-0.0094*	-0.0075*	-0.007*
Raça	0.0012	0.0021**	0.0012	0.0027*	0.0003	-0.0002	-0.0018	0.0014	0.0002*	0.0052**
Estado civil	-0.001	-0.0002	-0.0001	-0.0000	-0.0000	-0.0855	-0.0452	-0.0325	-0.0473**	-0.0641*
Anos_Fund	0.0014*	0.0018*	0.0003*	0.0027*	0.0011*	0.0043	0.0048**	0.0055*	0.0088*	0.0042**
Anos_Médio	0.0007	-0.0001	0.0001	0.0015	-0.0009	0.0026	0.0080*	0.0029	0.0012	0.0107*
E_Pai1	0.0012	0.0001	-0.0006	0.0005	-0.0002	-0.0006	-0.0007	-0.0011*	-0.0008**	-0.0003
E_Pai2	0.0026	-0.0032	-0.0078	-0.0022	-0.0038	-0.0084	-0.0135**	-0.0163*	-0.0150*	-0.0115***
E_Pai3	0.0006	0.0028	0.0042**	0.0016	0.0016	-0.0138	-0.0209**	-0.0290*	-0.0256*	-0.0202**
E_Pai4	0.0009	0.0040	0.0075	0.0034	0.0029	-0.0042	-0.0073	-0.0118**	-0.0088	-0.0045
E_mae1	-0.0002	0.0002	-0.0001	0.0002	0.0004	-0.0004	-0.0002	-0.0000	-0.0000	-0.0000
E_mae2	-0.0078***	-0.0001	-0.0019	0.0066	0.0071	-0.0097*	-0.0056*	-0.0061*	-0.0021	-0.0013
E_mae3	0.0009***	0.0001	0.0007***	0.0000	-0.0000	-0.0130**	-0.0065	-0.0172*	-0.0109**	-0.0098
E_mae4	0.0037	0.0005	0.0031	-0.0009	-0.0002	-0.0068	0.0002	-0.0057	0.0002	-0.0002
Renda00	-0.0066*	-0.0030*	-0.0011*	-0.0005*	-0.0001	-0.0035*	-0.0017*	-0.0010	-0.0008*	-0.0005*
Renda01	-0.3294*	-0.01503*	-0.0666*	-0.0382*	-0.0176*	-0.0965*	-0.0516*	-0.0320*	-0.0263*	-0.0175*
Renda02	-0.0110*	-0.0054*	-0.0027*	-0.0017*	-0.0011*	-0.1909*	-0.1041*	-0.0685*	-0.0609*	-0.0447*
Renda03	0.1634*	0.0810*	0.0435*	0.0318*	0.0205*	-0.2054*	-0.1092*	-0.0693*	-0.0655*	-0.0499*
Renda04	0.1166*	0.0523*	0.0296*	0.0127*	0.0086	-0.1217*	-0.0554*	-0.0334*	-0.0216*	-0.0154*
Renda05	0.0475*	0.0277*	0.0105	0.0123	-0.0003	-0.0479*	-0.0280*	-0.0110	-0.0144	-0.0015
CSE	-0.0037	-0.0022	-0.0010	-0.0008	-0.0008	0.0382**	0.0326**	0.0546*	0.0555*	0.0563*
Infra	0.0016	0.0014	-0.0006	0.0005	-0.0000	0.0419*	0.0567*	0.0802*	0.0445*	0.0137
Constante	-	-	-	-	-	0.6056*	0.1398***	-0.1019***	-0.0212	0.0842

Elaboração dos autores.

Obs.: Significância (*) = 1%; (**) = 5%; (***) = 10%.

Na parte inferior da distribuição, no quantil 0.1, o hiato de notas foi reduzido devido a características referentes a idade, escolaridade da mãe e faixas de rendas para os que não possuem renda e os que possuem até 2 SMs, com destaque para a variável que indica se o aluno possui renda mensal domiciliar até 1 SM, cujo efeito foi responsável por tornar o efeito composição negativo. Resultado semelhante pode ser observado para os demais quantis, embora a característica do aluno ter mãe com o ensino fundamental completo ou incompleto não afete mais, no sentido de reduzir os diferenciais de notas e na parte superior da distribuição, quantil 0.9. Efeito semelhante é constatado ao se considerar família destituída de renda.

Por sua vez, a variável *gênero* age de forma a contribuir para o aumento do diferencial, positiva e significativa, exceto no quantil 0.1, cuja maior contribuição se encontra no quantil 0.25, aproximadamente 13. Em partes da distribuição, a raça também possui esse efeito, aumentando os diferenciais em 39,62 e 10, nos quantis 0.25 e 0.75, assim como as escolaridades da mãe e do pai, nas quais o efeito sobre o hiato no quantil 0.5 foi de 4,43 e 26,58, respectivamente.

Ao longo de toda a distribuição, a variável que indica se o aluno terminou o ensino fundamental em oito anos apresentou sinal positivo, contribuindo para o aumento das disparidades de notas entre os alunos das profissionalizantes e privadas, cujo efeito foi maior no quantil 0.25, aproximadamente 40. Desse modo, ter uma boa trajetória no ensino fundamental afeta o desempenho no Enem, favorecendo os alunos das escolas particulares. Além disso, possuir renda a partir de 2 SMs, para alguns percentis, também age de forma a aumentar as diferenças de notas, cujo maior efeito pode ser observado no quantil 0.25, nas três faixas de renda. Com isso, entre as variáveis que contribuem para elevar o hiato de notas na média geral do Enem, pertencer a elevadas faixas de rendas é o principal fator.

Como se pôde observar na tabela 2, a parte do diferencial atribuído a fatores não observáveis foi bastante elevada em todos os percentis: 135 no quantil 0.1; 93 no quantil 0.25; 84 na mediana; e 77 e 88 nos quantis 0.75 e 0.9, respectivamente. Isso pode ser comprovado por meio da decomposição detalhada desse efeito, apresentado na tabela 3, o qual mostra a influência significativa de fatores não observáveis na maioria das variáveis em boa parte da distribuição, principalmente em percentis mais elevados.

Na parte inferior da distribuição, o efeito não observável seria menor e contribuiria para a redução do hiato entre as notas na média geral, caso os alunos das escolas profissionalizantes tivessem um aumento no desempenho para os alunos do sexo masculino, com mães que possuem ensino fundamental completo ou incompleto, e um maior retorno em todas as faixas de rendas, principalmente para aqueles com renda familiar entre 2 e 5 SMs. Esse efeito não observável age de forma a reduzir o efeito estrutural em aproximadamente 240.

Considerando-se a mediana da distribuição, percebe-se que o efeito estrutural de notas poderia ser reduzido e, assim, contribuir para a queda da desigualdade de notas entre as escolas. Isso ocorreria se os alunos do sexo masculino das escolas profissionalizantes – entre aqueles com pais em todos os níveis de escolaridade, exceto os que possuem nível superior completo e incompleto – melhorassem o desempenho. O destaque seria para aqueles alunos cujos pais possuem ensino médio completo ou incompleto; nesse segmento, a redução sobre os efeitos não observáveis seria 35, 36. Além destes, o efeito estrutural reduziria o aumento do efeito dos alunos com mães que possuem ensino fundamental e médio completo

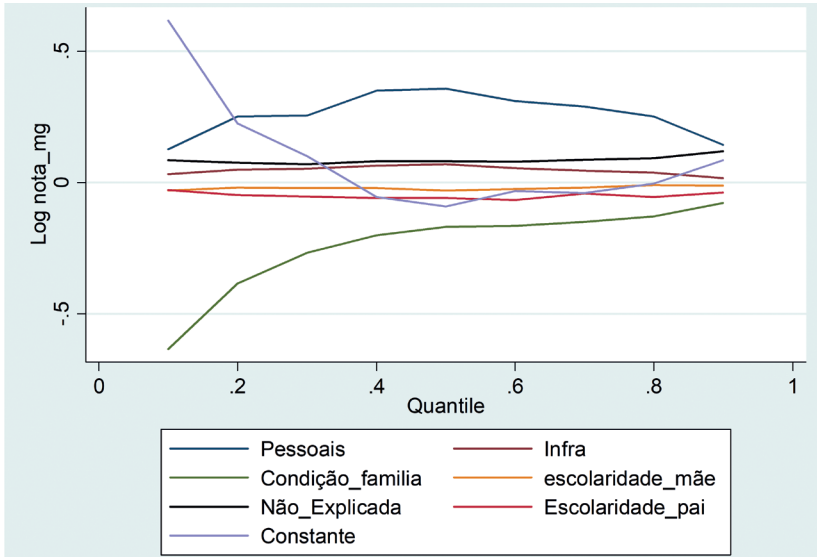
ou incompleto em, aproximadamente, 7 e 21, respectivamente; com um maior retorno em notas na média geral dos alunos em todas as faixas de rendas, exceto na *Renda05*, cujo efeito mais expressivo foi para aqueles com renda familiar entre 2 e 5 SMs, reduzindo o efeito não observado em aproximadamente 85.

Resultado similar pode ser verificado na parte superior da distribuição. Porém, o efeito estrutural das notas seria reduzido com o aumento no desempenho para os alunos das EEEPs do sexo masculino, solteiros, com pais que possuem ensino fundamental e médio completo ou incompleto. Além disso, esse efeito também se reduziria com um maior retorno em notas na média geral dos alunos em todas as faixas de rendas, exceto para os que possuem renda familiar entre 10 e 15 SMs, cujo efeito mais expressivo foi para aqueles com renda familiar entre 2 e 5 SMs, diminuindo o efeito não observado em, aproximadamente, 41.

Assim, é importante verificar a distribuição de notas por quantis para que se possa identificar o efeito heterogêneo das características dos alunos e das escolas, bem como do retorno dessas características no hiato de notas entre as escolas particulares e profissionalizantes do Ceará. As contribuições das covariáveis na redução do hiato podem servir como forma de identificar os fatores em que as políticas públicas poderiam agir quando se pretende equiparar o ensino público ao privado. Porém, não se pôde analisar características relacionadas à habilidade inata do aluno e ao esforço exercido pelos estudantes de ambas as redes de ensino, uma vez que estes podem ser fatores relevantes para a determinação da nota do aluno e do diferencial, mas não são observáveis.

Para melhor visualização da contribuição desses efeitos, foram construídos os gráficos 9 e 10, nos quais se optou por agregar algumas variáveis em grupos, em que *idade*, *sexo*, *raça*, *estado_civil*, *Anos_Fund* e *Anos_Médio* pertencem às características pessoais; *E_Pai1*, *E_Pai2*, *E_Pai3* e *E_pai4* inserem-se no grupo que representa a escolaridade do *p*; e a escolaridade da mãe e as variáveis *renda* e *CSE* pertencem ao grupo da condição da família.

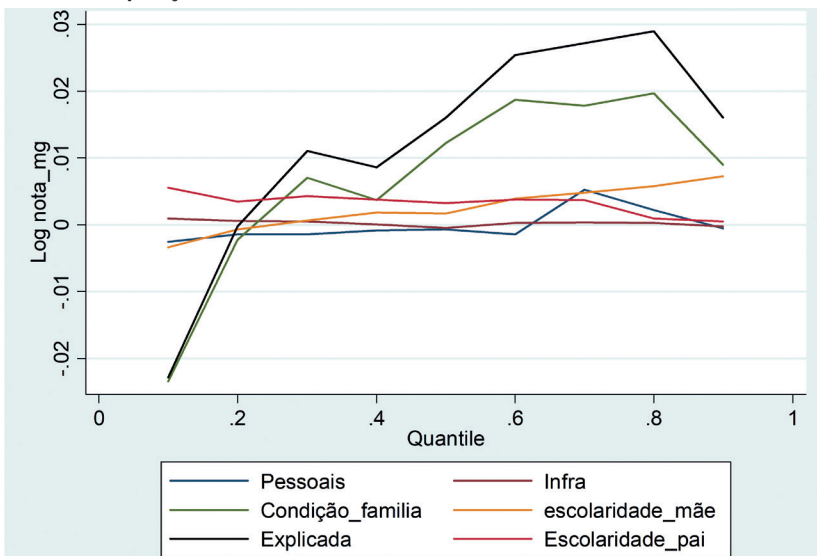
GRÁFICO 9
Decomposição detalhada do efeito característico



Elaboração dos autores utilizando os dados da tabela 3.

Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

GRÁFICO 10
Decomposição detalhada do efeito estrutural



Elaboração dos autores utilizando os dados da tabela 3.

Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

Observa-se que as condições socioeconômicas da família e as escolaridades dos pais são fatores determinantes para a redução dessa desigualdade. Nesse contexto, optou-se por fazer simulações para entender como uma melhoria de renda poderia compensar deficiências históricas no perfil educacional no desempenho dos alunos. Desse modo, serão simuladas distribuições com mudanças na escolaridade dos pais e na renda; fatores que individualmente aparecem com grande importância, de forma que os alunos das escolas profissionalizantes não possuam diferenças significativas nessas variáveis quando comparados aos alunos das escolas particulares.

Para tanto, aplicou-se o balanceamento por entropia.¹⁴ Conforme Hainmueller (2012), esse método permite ponderar um agrupamento de dados em que as distribuições das variáveis e as observações reponderadas satisfaçam um conjunto de condições especiais de momentos, de forma que exista equilíbrio exato sobre o primeiro (média), o segundo (variância) e o terceiro (assimetria) momentos das distribuições de variáveis independentes nos grupos de tratamento e controle. Assim, é possível se especificar um nível de equilíbrio desejável para as covariadas, usando um conjunto de condições associadas aos momentos da distribuição.

No cenário I, os alunos das escolas profissionalizantes apresentariam a mesma distribuição dos estudantes das escolas particulares nas variáveis referentes à renda da família. No cenário II, além dessas variáveis, foram balanceadas as que são relacionadas à educação da mãe. Por fim, no cenário III, acrescentaram-se ao cenário anterior as variáveis concernentes à educação do pai. Após a aplicação do balanceamento por entropia, estimou-se o diferencial total a partir das estimativas da FIR, ponderada pelos pesos da entropia, e comparou-se este aos resultados da tabela 2, cujos resultados se encontram na tabela 4.

TABELA 4
Diferencial total ponderado pela entropia: MG

Variáveis	Cenários											
	I				II				III			
	Antes		Após		Antes		Após		Antes		Após	
	EEEPs	Particulares	EEEPs	Particulares	EEEPs	Particulares	EEEPs	Particulares	EEEPs	Particulares	EEEPs	Particulares
Renda00	0.0155	0.0049	0.0049	0.0049	0.0155	0.0049	0.0049	0.0049	0.0155	0.0049	0.0049	0.0049
Renda01	0.6392	0.1323	0.1323	0.1323	0.6392	0.1323	0.1323	0.1323	0.6392	0.1323	0.1323	0.1323
Renda02	0.2797	0.2635	0.2635	0.2635	0.2797	0.2635	0.2635	0.2635	0.2797	0.2635	0.2635	0.2635
Mãe01	-	-	-	-	0.0279	0.0068	0.0068	0.0068	0.0279	0.0068	0.0068	0.0068
Mãe02	-	-	-	-	0.4944	0.1572	0.1572	0.1572	0.4944	0.1572	0.1572	0.1572
Pai01	-	-	-	-	-	-	-	-	0.0692	0.0149	0.0149	0.0149
Pai02	-	-	-	-	-	-	-	-	0.5787	0.2317	0.2317	0.2317

(Continua)

14. Pareamento desenvolvido por Hainmueller (2012), cuja especificação do método se encontra no apêndice C.

(Continuação)

Desigualdades						
Quantil 0.1	0.0637	0.0406	0.0637	0.0384	0.0637	0.038
Quantil 0.5	0.0978	0.0655	0.0978	0.0617	0.0978	0.0609
Quantil 0.9	0.1365	0.1072	0.1365	0.1044	0.1365	0.1028

Elaboração dos autores.

A partir da tabela 4, verifica-se que, antes do pareamento, os alunos das escolas profissionalizantes apresentavam desvantagens referentes às variáveis *renda* e *educação dos pais*. No cenário I, após o balanceamento, esses alunos passaram a evidenciar a renda da família equivalente aos alunos das escolas particulares, reduzindo, assim, a desigualdade na média geral em, aproximadamente, 36% no primeiro quantil das distribuições de notas, 33% na mediana e 21,4% no quantil 0.9. No cenário II, as variáveis da escolaridade da mãe também não apresentaram diferenças significativas entre os dois grupos, após o pareamento. Dessa forma, a desigualdade reduziria para 39,7%, 36,9% e 23,5%, respectivamente, nos quantis, 0.1, 0.5 e 0.9, enquanto, no cenário III, após balancear as variáveis *renda*, *escolaridade da mãe* e *escolaridade do pai*, o diferencial total diminuiria 40,3%, 37,7% e 24,6% nos quantis analisados.

Dessa forma, percebe-se a importância da decomposição detalhada dos fatores sobre o diferencial total; em destaque, a renda da família e a escolaridade dos pais, uma vez que estes fatores foram capazes de reduzir a desigualdade de notas dos alunos na média geral, em todos os quantis. Além disso, cabe destacar que esse efeito é maior nos quantis iniciais da distribuição – ou seja, entre os alunos com menores desempenho nos dois grupos.

4 CONCLUSÕES

Este trabalho procurou ampliar o debate existente na literatura da economia da educação sobre o ensino público, com ênfase no ensino profissionalizante desenvolvido no Ceará. Para tanto, buscou-se trazer novas evidências sobre os diferenciais de desempenho entre o ensino público e privado, ao analisar as diferenças nas distribuições de notas no Enem, para as cinco áreas de conhecimento e na média geral, entre alunos das EEEPs e das escolas privadas.

Esta análise foi feita visando reduzir a influência de fatores não observados, ao se construir grupos de alunos mais homogêneos para se aferir o diferencial de rendimento escolar líquido. A pertinência para tal escolha prende-se ao fato de as escolas de ensino profissional apresentarem desempenho superior às demais escolas públicas. Dessa maneira, busca-se verificar se esse novo modelo de escola pública provê uma estrutura de ensino equiparável ao da escola privada, ao analisar

como os alunos dos dois tipos de escola se comparam ao longo da distribuição de rendimento escolar e quais fatores observáveis mais influenciam os desempenhos dos alunos das duas instituições.

Além de propor uma nova construção para o grupo das escolas públicas, este trabalho buscou inovar ao utilizar o método RQI proposto por Firpo, Fortin e Lemieux (2007), o qual emprega regressões com base na FIR combinadas com um procedimento de reponderação inspirado em DiNardo, Fortin e Lemieux (1996) e permite aplicar o método de decomposição de Oaxaca-Blinder.

Como não se possuem informações identificadas dos alunos do ensino privado para uma análise longitudinal, utilizaram-se apenas as informações referentes ao desempenho dos alunos no Enem para 2014, as quais foram extraídas do Inep. Já a identificação das escolas profissionalizantes e de suas dotações de infraestrutura de todas as escolas ocorreu por meio de dados da Seduc e do Censo Escolar de 2014, respectivamente.

Os resultados encontrados mostram que, em todas as áreas, as estimativas dos diferenciais foram significantes, apresentando vantagens aos alunos das escolas particulares em relação aos alunos do ensino profissional. Nas áreas de matemática, ciências da natureza e na média geral, o aumento da diferença de notas na parte superior da distribuição indica o efeito *glassceiling*, pois observa-se que o diferencial total entre as notas dos alunos das escolas profissionalizantes e privadas é crescente nos quantis. Dessa forma, a diferenciação de notas, a favor dos alunos das escolas particulares, é maior para aqueles que apresentam desempenho superior. Nas áreas de ciências humanas e linguagens e códigos, apesar de apresentarem diferenciais positivos, estes praticamente permanecem constantes ao longo da distribuição, com leves oscilações; enquanto, em redação, os diferenciais foram os mais elevados ao longo de toda a distribuição.

A decomposição quantílica de Firpo, Fortin e Lemieux (2007) permite identificar os fatores que contribuem para esse diferencial de notas, os quais podem ser observados através dos efeitos composição (característicos) e estruturais (coeficientes). Dos resultados, percebeu-se que os efeitos não observáveis (coeficientes) são todos positivos e significativos em todos os quantis da distribuição e em todas as áreas do Enem, sendo responsáveis por explicar a maior parte da diferença das notas, com destaque para matemática, cujo efeito é crescente ao longo da curva. Esses resultados indicam que os alunos das escolas particulares são mais capazes de converter insumos educacionais em maiores pontuações em testes padronizados e que existem outros fatores não observáveis pertencentes a estes alunos que contribuem para melhores desempenhos.

Além disso, verificou-se que o efeito característico age no sentido de diminuir as diferenças entre as notas apenas em alguns quantis da parte inferior da distribuição destas, principalmente para o quantil 0.1 de ciências humanas e de linguagens e códigos, cujo efeito reduz o diferencial em 56 e 54, respectivamente. Porém,

na maioria dos casos, esse efeito foi positivo e significativo, atuando de maneira a contribuir com o diferencial de notas. Ou seja, os alunos das escolas privadas apresentam melhores características que contribuem para um bom desempenho, principalmente nos quantis mais elevados da distribuição, pois, em todas as áreas de conhecimento e na média geral, esses efeitos foram positivos e significativos nos percentis 0.9 e 0.95.

A partir da decomposição detalhada dos fatores que afetam o diferencial entre os grupos, identificou-se que melhorias na renda da família e na escolaridade dos pais poderiam minimizar as deficiências no perfil educacional no desempenho dos alunos. Dessa forma, foram construídos três cenários com mudanças nessas variáveis, e utilizou-se o balanceamento por entropia para equiparar o grupo dos alunos das escolas profissionalizantes aos alunos das escolas particulares. Após esse procedimento, estimou-se o diferencial total a partir da FIR ponderada pelos pesos da entropia. Dos resultados, observou-se que, ao se melhorar as faixas de renda e a educação dos pais, a desigualdade entre esses grupos poderia reduzir 40,3%, 37,7% e 24,6%, respectivamente, no primeiro quantil, na mediana e no extremo da distribuição das notas na média geral.

Embora este estudo tenha buscado inovar ao comparar grupos teoricamente mais homogêneos, alunos das escolas profissionalizantes com alunos das escolas particulares, o diferencial de notas no desempenho do Enem ainda é significativo, ao longo de toda a distribuição. Com isso, políticas públicas educacionais, como o ensino profissionalizante, devem ser complementadas com políticas sociais que não só visem à melhoria no tipo de ensino que é ofertado, mas também foquem no desenvolvimento das famílias, por meio de melhorias na renda e na escolaridade dos pais.

REFERÊNCIAS

- BARSKY, R. *et al.* Accounting for the black-white wealth gap: a nonparametric approach. **Journal of the American Statistical Association**, v. 97, n. 459, p. 663-673, 2002.
- BORAH, B. J.; BASU, A. Highlighting differences between conditional and unconditional quantile regression approaches through an application to assess medication adherence. **Health Economics**, v. 22, n. 9, p. 1052-1070, 2013.
- CEARÁ. Secretaria da Educação do Estado do Ceará. **O pensar e o fazer da educação profissional no Ceará: 2008-2014**. Relatório de gestão. Fortaleza: Seduc, 2014. (Coordenadoria de Educação Profissional, 1. ed.).
- CHERNOZHUKOV, V.; FERNANDEZ-VAL, I.; MELLY, B. Inference on counterfactual distributions. **Econometrica**, v. 81, n. 6, p. 2205-2268, 2013.

COSTA, L.; ARRAES, R. Identificação parcial do efeito das escolas privadas brasileiras. **Pesquisa e Planejamento Econômico**, v. 44, n. 2, p. 337-371, ago. 2014.

DINARDO, J.; FORTIN, N. M.; LEMIEUX, T. Labor market institutions and the distribution of wages, 1973-1992: a semiparametric approach. **Econometrica**, v. 64, n. 5, p. 1001-1044, Sept. 1996.

DONALD, S. G.; GREEN, D. A.; PAARSCH, H. P. Differences in wage distributions between Canada and the United States: an application of a flexible estimator of distribution functions in the presence of covariates source. **Review of Economic Studies**, v. 67, n. 4, p. 609-633, 2000.

EDOKA, I. P. **Decomposing differences in cotinine distribution between children and adolescents from different socioeconomic backgrounds**. Heslington: HEDG/ University of York, 2012. (Working Paper, n. 12/29).

FRANÇA, M. T. A.; GONÇALVES, F. O. Provisão pública e privada de educação fundamental: diferenças de qualidade medidas por meio de propensity score. **Economia Aplicada**, v. 14, n. 4, p. 373-390, 2010.

FISHER, J.; MARCHAND, J. Does the retirement consumption puzzle differ across the distribution? **The Journal of Economic Inequality**, v. 12, p. 279-296, 2014.

FIRPO, S.; FORTIN, N.; LEMIEUX, T. **Decomposing wage distributions using recentered influence function regressions**. Vancouver: University of British Columbia, 2007. Mimeographed.

_____. Unconditional quantile regressions. **Econometrica**, v. 77, n. 3, p. 953-973, 2009.

FORTIN, N. M. The gender wage gap among young adults in the United States: the importance of money vs. people. **The Journal of Human Resources**, v. 43, n. 4, p. 886-920, 2008.

FORTIN, N. M.; LEMIEUX, T. Rank regressions, wage distributions, and the gender gap. **The Journal of Human Resources**, v. 33, n. 3, p. 610-643, 1998.

FOURNIER, J.; KOSKE, I. **Less income inequality and more growth: are they compatible? Part 7. The drivers of labour earnings inequality: an analysis based on conditional and unconditional quantile regressions**. Paris: OECD Publishing, 2012. (Working Papers, n. 930).

HAINMUELLER, J. Entropy balancing for causal effects: a multivariate reweighting method to produce balanced samples in observational studies. **Political Analysis**, v. 20, n. 1, p. 25-46, 2012.

HAMPEL, F. R. *et al.* **Robust statistics: the approach based on influence functions**. New York: Wiley, 1986.

HIRSCH, B. T.; WINTERS, J. V. An anatomy of racial and ethnic trends in male earnings in the U.S. **Review of Income and Wealth**, v. 60, n. 4, p. 930-947, 2014.

IPECE – INSTITUTO DE PESQUISA E ESTRATÉGIA ECONÔMICA DO CEARÁ. **Análise da participação das escolas públicas estaduais cearenses no Exame Nacional do Ensino Médio (Enem): 2009-2011**. Fortaleza: Ipece, fev. 2013 (Ipece Informe, n. 54). Disponível em: <<https://bit.ly/2PQuLYa>>.

JUHN, C.; MURPHY, K. M.; PIERCE, B. Wage inequality and the rise in returns to skill. **Journal of Political Economy**, v. 101, n. 3, p. 410-442, June 1993.

KASSENBOEHMER, S. C.; SINNING, M. G. Distributional changes in the gender wage gap. **Industrial & Labor Relations Review**, v. 67, n. 2, p. 335-361, Apr. 2014.

KOENKER, R.; BASSETT, J. R. Regression quantiles. **Econometrica**, v. 46, n. 1, p. 33-50, 1978.

_____. Robust tests for heteroscedasticity based on regression quantiles. **Econometrica**, v. 50, n. 1, p. 43-61, 1982.

LE, H. T.; BOOTH, A. L. Inequality in vietnamese urban-rural living standards, 1993-2006. **Review of Income and Wealth**, v. 60, n. 4, p. 862-886, Dec. 2014.

MACHADO, J. F.; MATA, J. Counterfactual decomposition of changes in wage distributions using quantile regression. **Journal of Applied Econometrics**, v. 20, n. 4, p. 445-465, May/June 2005.

MANSKI, C. Anatomy of the selection problem. **The Journal of Human Resource**, v. 24, n. 3, p. 343-360, 1989.

MELLY, B. Decomposition of differences in distribution using quantile regression. **Labour Economics**, v. 12, n. 4, p. 577-590, Aug. 2005.

MORAES, A. G. E.; BELLUZZO, W. O diferencial de desempenho escolar entre escolas públicas e privadas no Brasil. **Nova Economia**, v. 24, n. 2, p. 409-430, maio/ago. 2014.

MORIN, L. P. Do men and women respond differently to competition? Evidence from a major education reform. **Journal of Labor Economics**, v. 33, n. 2, p. 443-491, 2015.

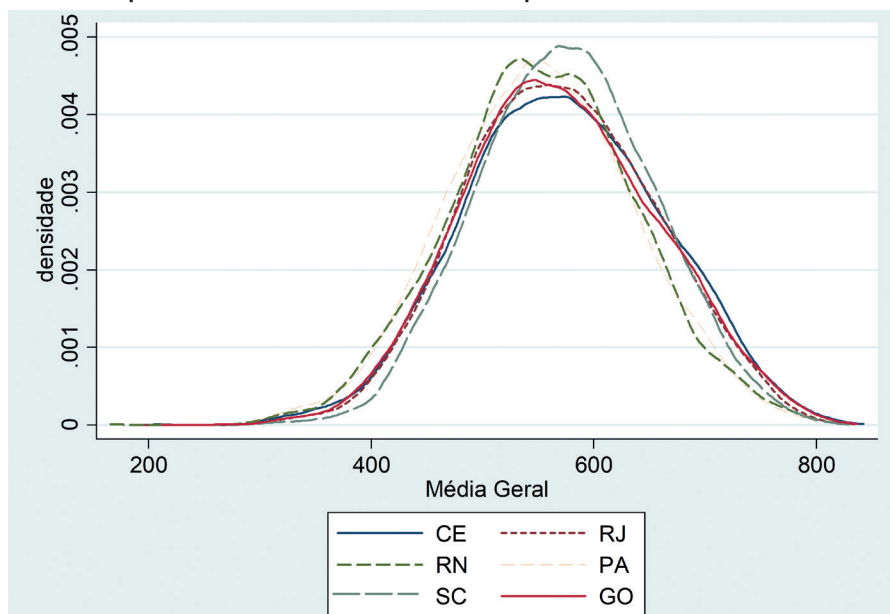
NGUYEN, H. T. **The evolution of the gender test score gap through seventh grade: new insights from Australia using quantile regression and decomposition**. Bentley: BCEC, 2015. (Working Paper, n. 15/7).

RUBIL, I. **The great recession and the public-private wage gap**: distributional decomposition evidence from Croatia 2008-2011. MPRA Paper, 2013. (Working Paper, n. 46798).

APÊNDICE A

GRÁFICO A.1

Desempenho no Enem dos alunos das escolas particulares: CE, RJ, RN, PA e GO

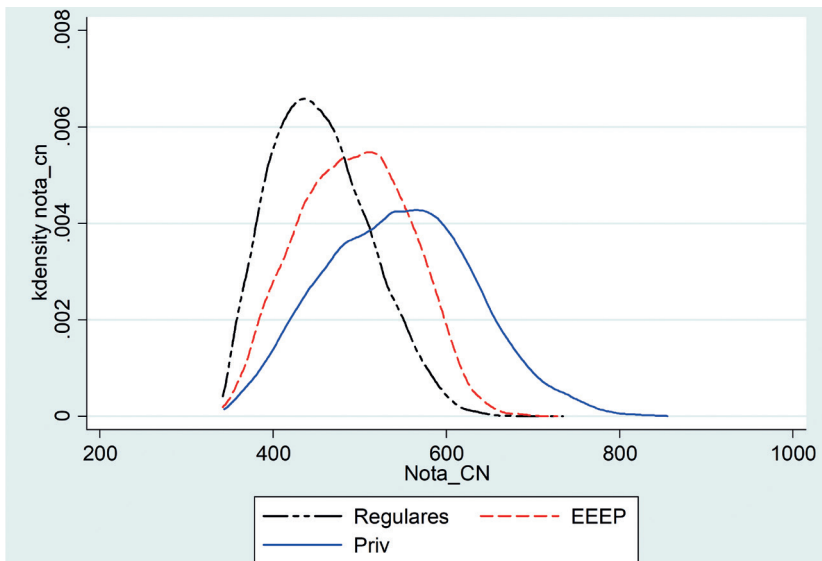


Elaboração dos autores.

Obs.: 1. Enem – Exame Nacional do Ensino Médio; CE – Ceará; RJ – Rio de Janeiro; RN – Rio Grande do Norte; PA – Pará; e GO – Goiás.

2. Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

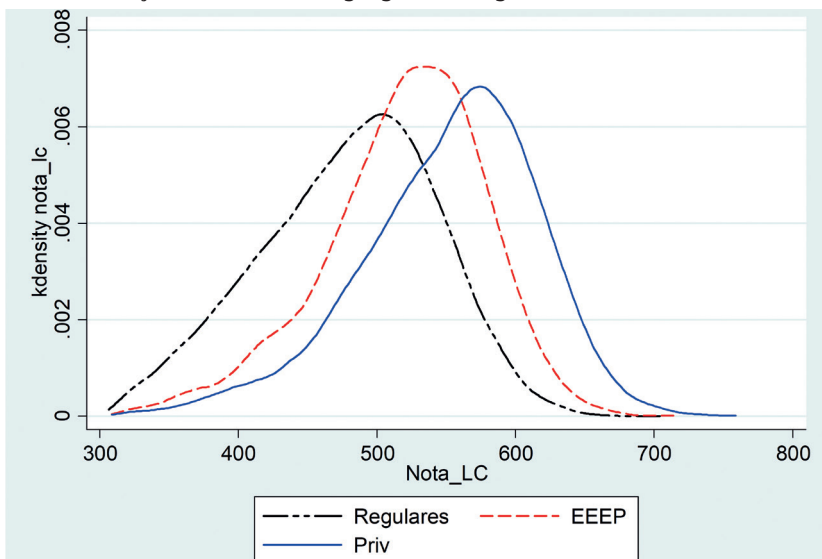
GRÁFICO A.2
Distribuição das notas em ciências naturais



Elaboração dos autores.

Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

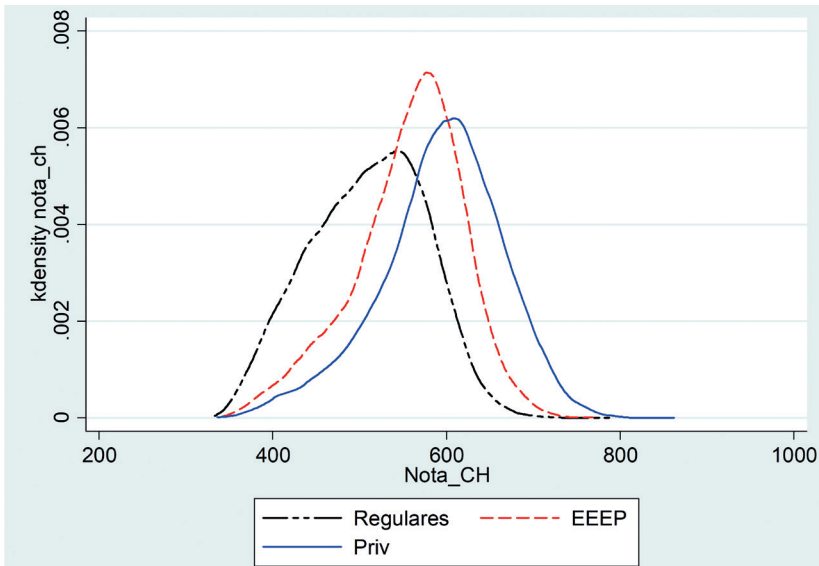
GRÁFICO A.3
Distribuição das notas em linguagens e códigos



Elaboração dos autores.

Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

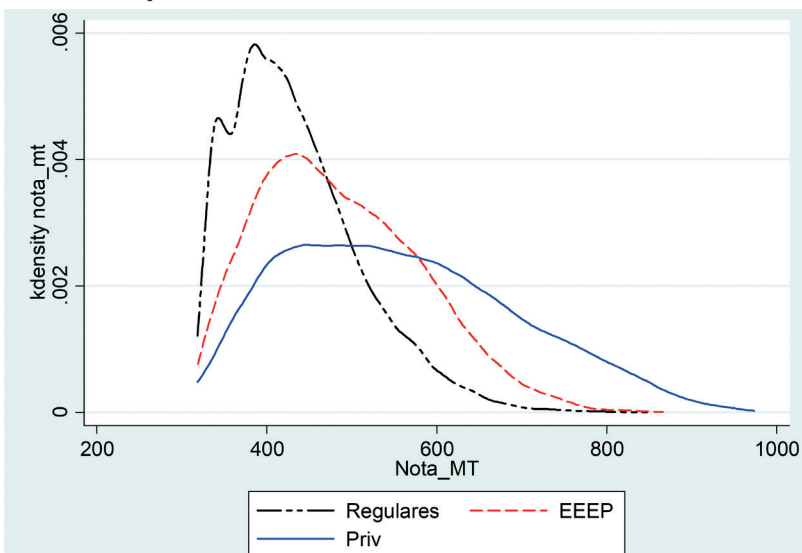
GRÁFICO A.4
Distribuição das notas em ciências humanas



Elaboração dos autores.

Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

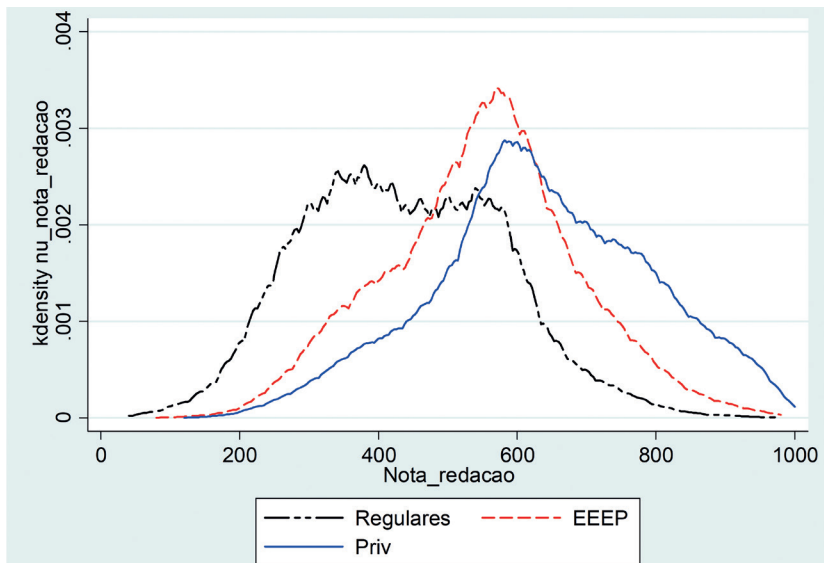
GRÁFICO A.5
Distribuição das notas em matemática



Elaboração dos autores.

Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

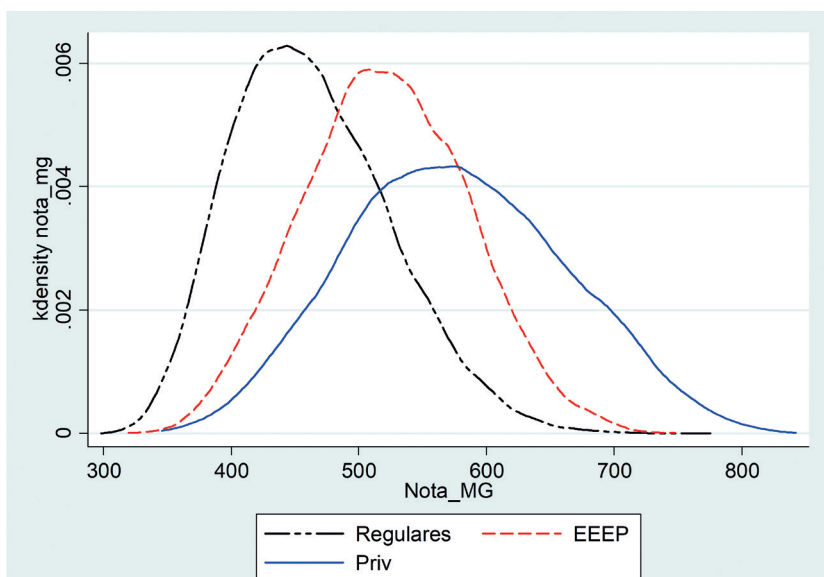
GRÁFICO A.6
Distribuição das notas em redação



Elaboração dos autores.

Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

GRÁFICO A.7
Distribuição das notas em média geral

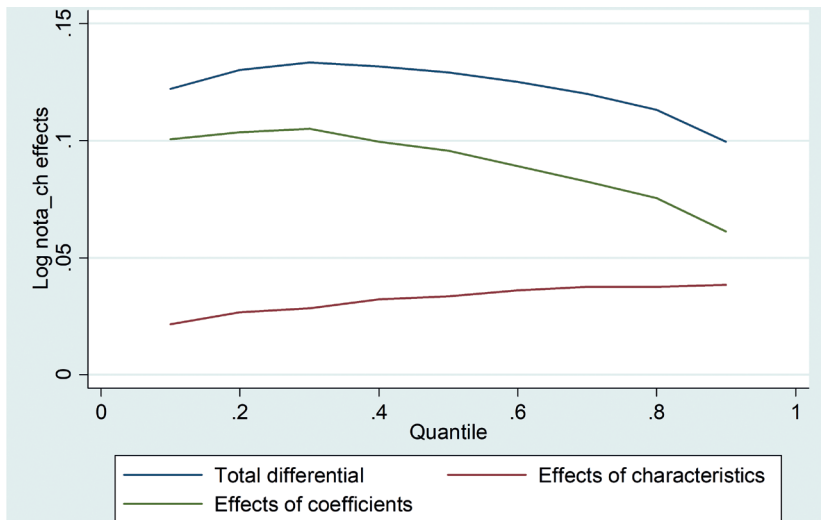


Elaboração dos autores.

Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

GRÁFICO A.8

Decomposição dos efeitos: escolas regulares e profissionalizantes – média geral

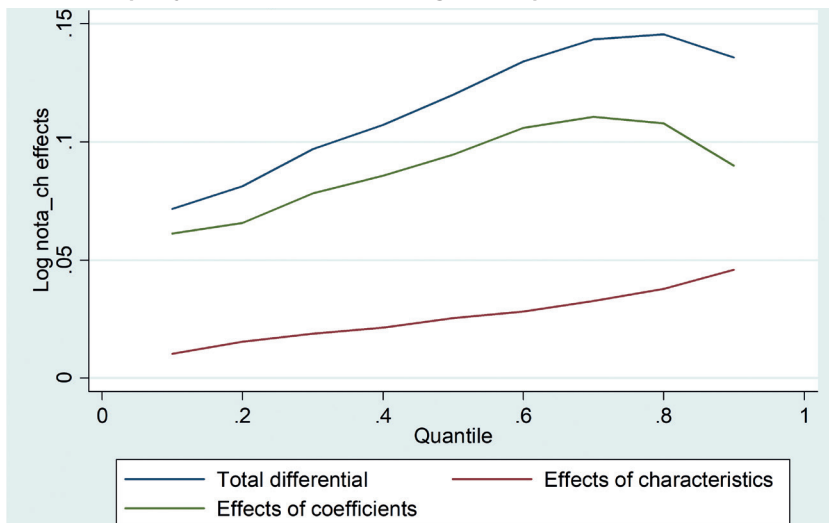


Elaboração dos autores.

Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

GRÁFICO A.9

Decomposição dos efeitos: escolas regulares e profissionalizantes – matemática

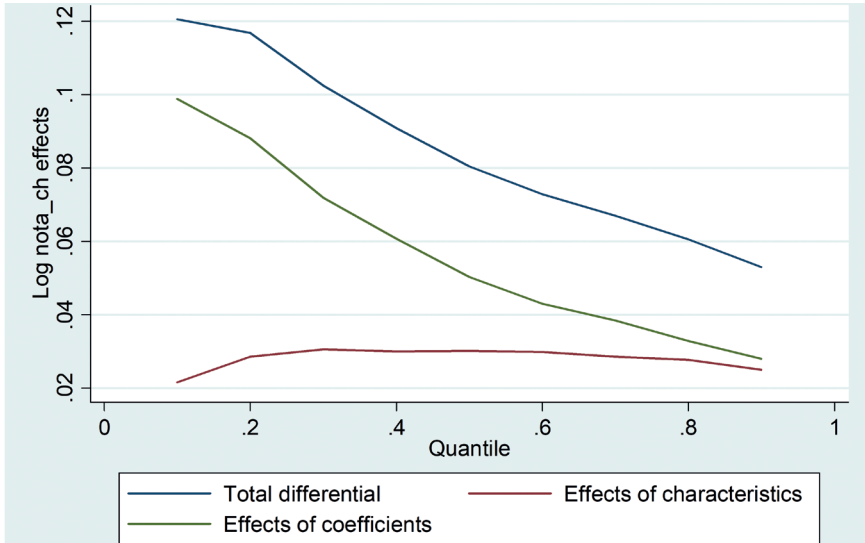


Elaboração dos autores.

Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

GRÁFICO A.10

Decomposição dos efeitos: escolas regulares e profissionalizantes – linguagens e códigos

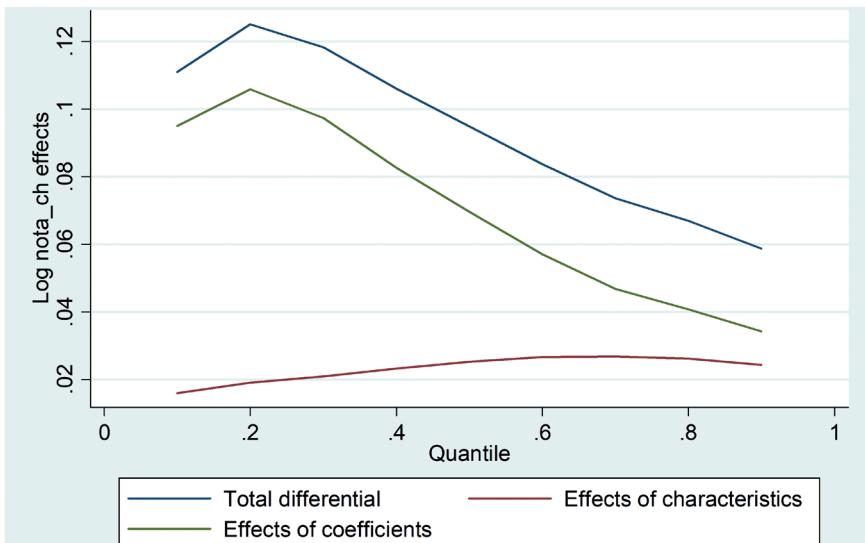


Elaboração dos autores.

Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

GRÁFICO A.11

Decomposição dos efeitos: escolas regulares e profissionalizantes – ciências humanas

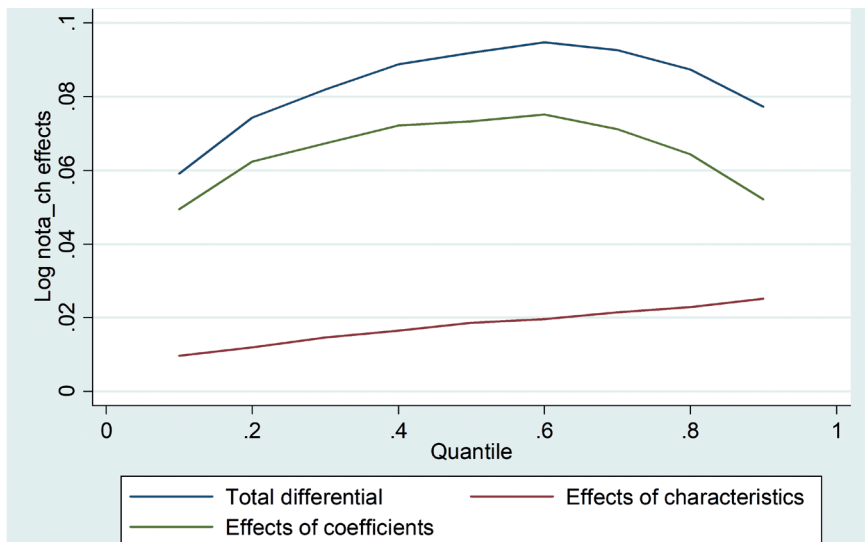


Elaboração dos autores.

Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

GRÁFICO A.12

Decomposição dos efeitos: escolas regulares e profissionalizantes – ciências da natureza

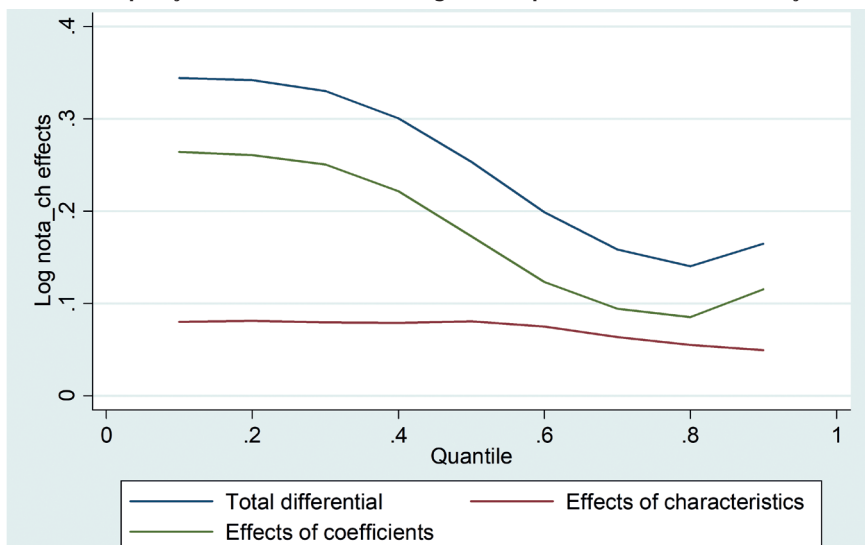


Elaboração dos autores.

Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

GRÁFICO A.13

Decomposição dos efeitos: escolas regulares e profissionalizantes – redação



Elaboração dos autores.

Obs.: Gráfico cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

APÊNDICE B

TABELA B.1
Estimativas da regressão quantílica incondicional por escola: média geral

Variáveis	MQQ ²											
	0.1				0.5				0.9			
	EEEPs ¹	Particulares	Regulares	EEEPs	Particulares	Regulares	EEEPs	Particulares	Regulares	EEEPs	Particulares	Regulares
Idade	-0.024*	-0.0097*	-0.0037*	-0.0348*	-0.0229*	-0.0029*	-0.0296*	-0.0078*	-0.0042*	-0.0125*	-0.0015*	-0.0036*
Sexo	0.020*	0.0033	-0.0058*	0.0064	-0.0102**	-0.0209*	0.0291*	0.0088**	-0.0076*	0.0276*	0.0123*	0.0088*
Raça	0.005***	0.0090*	0.0096*	0.0054	0.0049	0.0056**	0.0054	0.0088**	0.0110*	0.0016	0.0140*	0.0136*
Estrado civil	0.283	-0.0128	-0.0172*	0.0426	-0.0436	0.0269*	0.0271	-0.0057	-0.0226*	0.0251**	-0.0396**	-0.0028
Anos_Fund	0.0192*	0.0303*	0.0156*	0.0187*	0.0282*	0.0089*	0.0253*	0.0374*	0.0170*	0.0149*	0.0242*	-0.0230*
Anos_Médio	-0.0007	0.0093*	-0.0007	0.0043	0.0093**	-0.0062*	0.0011	0.0067***	-0.0013	-0.0055	0.0148*	0.0031
E_Pai1	0.0005	-0.0494*	-0.0365*	-0.0237	-0.0644**	-0.0209**	0.0115	-0.0665*	-0.0428*	0.0045	-0.0180	-0.0509**
E_Pai2	0.0098	-0.0458*	-0.0291*	-0.0077	-0.0441*	-0.0077	0.0227	-0.0479*	-0.0344*	0.0111	-0.0386*	-0.0491**
E_Pai3	0.0233	-0.0325*	-0.0039	0.0060	-0.0291*	0.0081	0.0425**	-0.0311*	-0.0057	0.0170	-0.0343*	-0.0204
E_Pai4	0.0208	-0.0100**	0.0055	0.0041	-0.0118**	0.0068	0.0351	-0.0096	-0.0004	0.0136	-0.0033	0.0049
E_mae1	-0.0045	-0.0272***	-0.0462*	0.0117	-0.0581	-0.0187*	0.0093	0.0001	-0.0439*	-0.0197	-0.0243***	-0.0644*
E_mae2	-0.0035	-0.0330*	-0.0395*	0.0232***	-0.0387*	-0.0151*	0.0057	-0.0336*	-0.0330*	-0.0212	-0.0298*	-0.0590*
E_mae3	0.0086	-0.0190*	-0.013**	0.0274**	-0.0058	0.0040	0.0198***	-0.0241*	-0.0014	-0.0025	-0.0275*	-0.0328*
E_mae4	0.0042	-0.0026	-0.0071	0.0183	-0.0047	0.0001	0.0153	-0.0042	0.0044	-0.0010	-0.0020	-0.0166
Renda00	0.1937*	-0.0931*	0.0353	0.6233*	-0.0928**	0.0022	0.1094*	-0.0983*	0.0103	0.0168	-0.1034*	0.0016
Renda01	0.2117*	-0.0992*	0.0448	0.6499*	-0.0796*	0.0130	0.1314*	-0.1107*	0.0179	0.0348*	-0.0978*	0.0133
Renda02	0.2456*	-0.0780*	0.0685	0.6789*	-0.0454*	0.0319	0.1720*	-0.0881*	0.0449	0.0706*	-0.0993*	0.0384
Renda03	0.2561*	-0.0458*	0.0797	0.6838*	-0.0032	0.0231	0.1821*	-0.0499*	0.0498	0.0858*	-0.0810*	0.0836
Renda04	0.2837*	-0.0220*	0.0777	0.6861*	-0.0086	0.0261	0.1742*	-0.0165**	0.0460	0.0507	-0.0375**	0.0987
Renda05	-0.0054	0.0575*	0.0810*	-0.0196	0.0371**	0.0371	0.1707	-0.0072	0.0208	-0.0056	-0.0307***	0.1424
CSE	0.0103	0.1304*	0.1165*	0.0269	0.1282*	0.0724*	-0.0107	0.1828*	0.0973*	-0.0042	0.0792*	0.0785*
Constante	6.3922	6.5196*	6.1587*	5.9553*	6.5609*	5.9992*	6.5421*	6.4402*	6.1906*	6.5380*	6.6223*	6.3618*

Elaboração dos autores.

Notas: ¹EEEPs = Escolas de educação de ensino profissional.

²MQQ = Mínimos quadrados ordinários.

Obs.: Significância (*) = 1%; (**) = 5%; e (***) = 10%.

TABELA B.2
Estimativas da regressão quantílica incondicional por escola: ciências humanas

Variáveis	MQO				0.1				0.5				0.9			
	EEEPs	Particulares	Regulares	EEEPs	Particulares	Regulares	EEEPs	Particulares	Regulares	EEEPs	Particulares	Regulares	EEEPs	Particulares	Regulares	
Idade	-0.0154	-0.00483	-0.0007	-0.0219*	-0.0087*	-0.0010*	-0.0198*	-0.0047*	-0.0008*	-0.0074*	-0.0010*	-0.0008*	-0.0074*	-0.0010*	-0.0010*	
Sexo	0.0253	0.01350	0.00791	0.0209*	-0.0022	-0.004***	0.0276*	0.0175*	0.0090*	0.0306*	0.0184*	0.0090*	0.0306*	0.0184*	0.0192*	
Estrado civil	0.0260	-0.00730	-0.0169	0.0885	0.0028	-0.0294*	-0.0004	-0.0045	-0.0187*	0.0166***	-0.0042	-0.0187*	0.0166***	-0.0042	-0.006***	
Raça	0.0058	0.00511	0.00797	0.0045	0.0078	0.0028	0.0081**	0.0028	0.0101*	0.0103*	0.0050***	0.0101*	0.0103*	0.0050***	0.0082*	
Anos_Fund	0.0152	0.02224	0.01351	0.0197*	0.0289*	0.0121*	0.0178*	0.0245*	0.0161*	0.0101*	0.0151*	0.0161*	0.0101*	0.0151*	0.0137*	
Anos_Médio	0.0012	0.00396	-0.0022	-0.0036	-0.0001	-0.0058**	0.0049	0.0050***	-0.0020	0.0011	0.0074*	-0.0020	0.0011	0.0074*	0.0019	
E_Pai1	0.0061	-0.02949	-0.0290	0.0087	-0.0479	-0.023***	0.0022	-0.0405*	-0.0472*	-0.0045	-0.0295*	-0.0472*	-0.0045	-0.0295*	-0.0327**	
E_Pai2	0.0085	-0.03102	-0.0247	-0.0002	-0.0298*	-0.0214	0.0053	-0.0332*	-0.0401*	-0.0042	-0.0378*	-0.0401*	-0.0042	-0.0378*	-0.030***	
E_Pai3	0.0234	-0.0212	-0.0025	0.0249	-0.0218**	-0.0016	0.0210	-0.0229*	-0.0125	0.0056	-0.0314*	-0.0125	0.0056	-0.0314*	-0.0114	
E_Pai4	0.0175	-0.00558	0.0034	0.0122	-0.0112	0.0007	0.0170	-0.0050	-0.0034	-0.0075	-0.0125***	-0.0034	-0.0075	-0.0125***	0.0073	
E_mae1	0.0043	0.00059	-0.0279	0.0191	0.0183	-0.0201**	-0.0053	-0.0148	-0.0300*	-0.0033	0.0161	-0.0300*	-0.0033	0.0161	-0.0316*	
E_mae2	0.0003	-0.02176	-0.0224	0.0242	-0.0251**	-0.016***	-0.0058	-0.0248*	-0.0195*	0.0084	-0.0036	-0.0195*	0.0084	-0.0036	-0.0287*	
E_mae3	0.0084	-0.0129	-0.0019	0.0284	-0.0058	0.0041	0.0070	-0.0171*	0.0041	0.0027	-0.0044	0.0041	0.0027	-0.0044	-0.0088	
E_mae4	0.0016	-0.00063	0.0019	0.0108	-0.0002	0.0031	0.0016	-4.65e-06	0.0071	0.0040	0.0048	0.0071	0.0040	0.0048	-0.0023	
Renda00	0.31398	-0.07255	0.0075	0.8756*	-0.1121**	-0.0762*	0.1132*	-0.0619*	0.0965***	0.0029	-0.0598*	0.0965***	0.0029	-0.0598*	0.0139	
Renda01	0.32080	-0.07152	0.0179	0.8716*	-0.0878*	-0.0663*	0.1181*	-0.0704*	0.1110***	0.0294*	-0.0694*	0.1110***	0.0294*	-0.0694*	0.0157	
Renda02	0.35243	-0.0525	0.03856	0.9167*	-0.0468*	-0.0515*	0.1529*	-0.0527*	0.1359**	0.0451*	-0.0648*	0.1359**	0.0451*	-0.0648*	0.0365	
Renda03	0.36238	-0.02764	0.0466	0.9392*	-0.0031	-0.0524*	0.1604*	-0.0303*	0.1429**	0.0559*	-0.0576*	0.1429**	0.0559*	-0.0576*	0.0581	
Renda04	0.34520	-0.01205	0.04624	0.9109*	-0.0048	-0.0518*	0.1414*	-0.0104***	0.1408**	0.0341	-0.0263*	0.1408**	0.0341	-0.0263*	0.0720	
Renda05	0.39146	-0.08889	0.06258	1.0233*	-0.0063	-0.0745	0.2959*	-0.0035	0.1707**	-0.0150	-0.0299*	0.1707**	-0.0150	-0.0299*	0.0502	
CSE	-0.0156	0.02750	0.06376	-0.0530**	0.0202	0.0545*	-0.0104	0.0455*	0.0837*	0.0116	0.0464*	0.0837*	0.0116	0.0464*	0.0464*	
Infra	-0.0005	0.0807	0.09219	-0.0540	0.1207*	0.0540*	-0.0006	0.0914*	0.1002*	0.0150	0.0299*	0.1002*	0.0150	0.0299*	0.1339*	
Constante	6.1976	6.4667	6.2148	5.5466*	6.3368*	6.1360*	6.5254*	6.4567*	6.1399*	6.4987*	6.5563*	6.1399*	6.4987*	6.5563*	6.3764*	

Elaboração dos autores.

Obs.: Significância (*) = 1%; (**) = 5%; (***) = 10%.

TABELA B.3
Estimativas da regressão quantílica incondicional por escola: ciências da natureza

Variáveis	MQO				0,1				0,5				0,9			
	Particulares		Regulares		Particulares		Regulares		Particulares		Regulares		Particulares		Regulares	
	EEEPs	Regulares	EEEPs	Regulares	EEEPs	Regulares	EEEPs	Regulares	EEEPs	Regulares	EEEPs	Regulares	EEEPs	Regulares	EEEPs	Regulares
Idade	-0,01619	-0,00665	-0,00141	-0,0111*	-0,014*	-0,0008*	-0,0208*	-0,0074*	-0,008*	-0,0017*	-0,008*	-0,0011*	-0,0011*	-0,0020*	-0,0011*	-0,0020*
Sexo	0,04458	0,030123	0,01743	0,0091***	0,0300*	0,0035**	0,0564*	0,0409*	0,0453*	0,0174*	0,0453*	0,0363*	0,0363*	0,0453*	0,0363*	0,0325*
Estado civil	0,02499	-0,01561	-0,00872	-0,0451	0,0509	-0,0116**	0,0251	-0,0119	0,0093	-0,0131*	0,0093	-0,0210***	-0,0210***	0,0093	-0,0210***	-0,0032
Raça	0,00415	0,008923	0,00697	0,0019	-0,0042	0,0030	0,0089***	0,0106*	0,0022	0,0075*	0,0022	0,0183*	0,0183*	0,0022	0,0099*	0,0099*
Anos_Fund	0,01332	0,029400	0,01038	0,0218*	0,0149*	0,0062*	0,0198*	0,0389*	0,0190*	0,0098*	0,0190*	0,0232*	0,0232*	0,0190*	0,0180*	0,0180*
Anos_Médio	0,00289	0,008523	0,00087	0,0066	0,0049	0,0004	0,0025	0,0061	-0,0000	-0,0002	-0,0000	0,0145*	0,0145*	-0,0000	0,0015	0,0015
E_Pai1	0,02350	-0,045682	-0,01941	-0,0296	0,0436	-0,0137	0,0323	-0,0582*	-0,0184	-0,0275**	-0,0184	-0,0375*	-0,0375*	-0,0184	-0,034***	-0,034***
E_Pai2	0,02603	-0,042944	-0,01779	-0,0325*	0,0449	-0,016***	0,0405***	-0,0457*	-0,0223	-0,022***	-0,0223	-0,0396*	-0,0396*	-0,0223	-0,035***	-0,035***
E_Pai3	0,03070	-0,036050	-0,00682	-0,0286*	0,0517	-0,0093	0,0474**	-0,0370*	-0,0198	-0,0116	-0,0198	-0,0381*	-0,0381*	-0,0198	-0,0140	-0,0140
E_Pai4	0,032032	-0,009626	0,00791	-0,0120	0,0464	-0,00007	0,0517**	-0,0042	-0,0075	0,0073	-0,0075	-0,0053	-0,0053	-0,0075	0,0074	0,0074
E_mae1	-0,00394	-0,036917	-0,03318	-0,0971**	0,0182	-0,0246*	-0,0156	-0,0113	-0,0346*	-0,0346*	-0,0346*	-0,0117	-0,0117	-0,0346*	-0,0428*	-0,0428*
E_mae2	-0,0054635	-0,029006	-0,02840	-0,0222**	0,0230	-0,011***	-0,021***	-0,0367*	-0,0323*	-0,0323*	-0,0323*	-0,0267*	-0,0267*	-0,0323*	-0,0400*	-0,0400*
E_mae3	-0,0016787	-0,0186841	-0,01423	-0,0076	0,0218	-0,0042	-0,0171	-0,0226*	-0,0025	-0,0171**	-0,0025	-0,0167**	-0,0167**	-0,0025	-0,0215**	-0,0215**
E_mae4	0,0018264	-0,0012844	-0,00943	0,0030	0,0268	-0,0054	-0,0122	-0,0005	-0,0034	-0,0108	-0,0034	0,0025	0,0025	-0,0034	-0,0119	-0,0119
Renda00	0,0653425	-0,090357	0,01734	-0,114*	-0,114*	0,0657	0,1368*	-0,1103*	0,0216	0,0022	0,0216	-0,1152*	-0,1152*	0,0216	0,0776*	0,0776*
Renda01	0,0797549	-0,09416	0,02013	-0,0659*	-0,113*	0,0644	0,1526*	-0,1034*	0,0113	0,0113	0,0237*	-0,1081*	-0,1081*	0,0237*	0,0807*	0,0807*
Renda02	0,1014531	-0,07831	0,03431	-0,107*	-0,107*	0,0729	0,1816*	-0,0892*	0,0269	0,0269	0,0498*	-0,1114*	-0,1114*	0,0498*	0,0993*	0,0993*
Renda03	0,1083583	-0,04581	0,04464	-0,092*	-0,092*	0,0744	0,1864*	-0,0481*	0,0354	0,0354	0,0679*	-0,0938*	-0,0938*	0,0679*	0,1272*	0,1272*
Renda04	0,0788905	-0,02117	0,05204	-0,081*	-0,081*	0,0816	0,1653*	-0,0193**	0,0090	0,0090	0,0090	-0,0526*	-0,0526*	0,0090	0,1455*	0,1455*
Renda05	0,2043453	-0,01205	0,03445	-0,0117	-0,0117	0,0707	0,1891	-0,0120	0,0642	0,0642	0,0642	-0,0336**	-0,0336**	0,0642	0,0924	0,0924
CSE	-0,012838	0,05229	0,0318113	-0,0133	-0,0133	0,0108***	-0,0416*	0,0552*	0,0386*	0,0386*	0,0026	0,0681*	0,0681*	0,0026	0,0363*	0,0363*
Infra	0,00586	0,13008	0,086118	-0,0084	-0,0084	0,1288*	-0,0109	0,1705*	0,0868*	0,0868*	0,0111	0,0368**	0,0368**	0,0111	0,1292*	0,1292*
Constante	6,31781	6,401588	6,11947	6,2495*	6,2495*	5,9143*	6,3473*	6,3909*	6,4610*	6,1419*	6,4610*	6,5434*	6,5434*	6,4610*	6,2345*	6,2345*

Elaboração dos autores.

Obs.: Significância (*) = 1%; (**) = 5%; (***) = 10%.

TABELA B.4
Estimativas da regressão quantílica incondicional por escola: linguagens e códigos

Variáveis	MQO														
	0.1					0.5					0.9				
	EEEPs	Particulares	Regulares	EEEPs	Particulares	Regulares	EEEPs	Particulares	Regulares	EEEPs	Particulares	Regulares	EEEPs	Particulares	Regulares
Idade	-0.0183	-0.00688	-0.00312	-0.0291*	-0.0134*	-0.0034*	-0.0210*	-0.0057*	-0.0035*	-0.0101*	-0.0010*	-0.0024*	-0.0101*	-0.0010*	-0.0024*
Sexo	-0.0091	-0.01930	-0.02130	-0.0265*	-0.0347*	-0.0361*	-0.0035	-0.0167*	-0.0232*	0.0018	-0.0017	-0.0070*	0.0018	-0.0017	-0.0070*
Estado civil	0.03945	0.00691	-0.01398	0.0715	0.0148	-0.0361*	0.0347***	-0.0004	-0.0091**	-0.0036	-0.0097	-0.0041	-0.0036	-0.0097	-0.0041
Raça	0.00595	0.00489	0.00851	0.0044	0.0057	0.0100*	0.0051	0.0060**	0.0084*	0.0096**	0.0069*	0.0080*	0.0096**	0.0069*	0.0080*
Anos_Fund	0.01377	0.01925	0.01065	0.0116**	0.0241*	0.0017	0.0156*	0.0198*	0.0119*	0.0123*	0.0114*	0.0132*	0.0123*	0.0114*	0.0132*
Anos_Médio	-0.0012	0.00358	-0.00225	-0.0069	0.0118**	-0.0020	0.0016	0.0015	-0.003***	-0.0014	0.0009	-0.0015	-0.0014	0.0009	-0.0015
E_Pai1	-0.0107	-0.03899	-0.02176	-0.0454	-0.0593**	-0.0436*	-0.0093	-0.0438*	-0.0141	-0.0020	-0.0393*	-0.028***	-0.0020	-0.0393*	-0.028***
E_Pai2	-0.0038	-0.03556	-0.01487	-0.0204	-0.0273*	-0.0324**	-0.0116	-0.0411*	-0.0095	0.0011	-0.0374*	-0.027***	0.0011	-0.0374*	-0.027***
E_Pai3	0.0078	-0.0255	0.00750	-0.0111	-0.0220*	-0.0098	0.0044	-0.0274*	0.0174	0.0087	-0.0302*	-0.0061	0.0087	-0.0302*	-0.0061
E_Pai4	-0.0018	-0.00943	0.01352	-0.0354	-0.0100	-0.0135	-0.0027	-0.0027	0.0248***	0.0033	-0.0140**	0.0025	0.0033	-0.0140**	0.0025
E_mae1	-0.0041	-0.00460	-0.03291	-0.0166	-0.0335	-0.0060	0.0055	0.0055	-0.0450*	-0.0059	0.0028	-0.0352*	-0.0059	0.0028	-0.0352*
E_mae2	-0.0064	-0.01651	-0.02888	-0.0022	-0.0234**	-0.0104	-0.0014	-0.0014	-0.0355*	-0.0117	-0.0093**	-0.0310*	-0.0117	-0.0093**	-0.0310*
E_mae3	0.00486	-0.0100	-0.00791	0.0146	-0.0101	0.0101	0.0109	0.0109	-0.0113	-0.0052	-0.0077***	-0.0135	-0.0052	-0.0077***	-0.0135
E_mae4	-0.0033	-0.00050	-0.00458	0.0048	-0.0020	0.0068	0.0012	0.0012	-0.0100	-0.0085	0.0037	-0.0037	-0.0085	0.0037	-0.0037
Renda00	0.17614	-0.06865	0.03453	0.8101*	-0.0216	0.1040	0.1198*	0.1198*	-0.060	0.0121***	-0.0690*	0.0466*	0.0121***	-0.0690*	0.0466*
Renda01	0.19784	-0.06354	0.04222	0.8470*	-0.0484*	0.1124	0.1353*	0.1353*	0.056	0.0378*	-0.0514*	0.0582*	0.0378*	-0.0514*	0.0582*
Renda02	0.22196	-0.043608	0.06379	0.8742*	-0.0125	0.1340	0.1634*	0.1634*	0.305	0.0652*	-0.0512*	0.0771*	0.0652*	-0.0512*	0.0771*
Renda03	0.23301	-0.02629	0.06658	0.8804*	0.0009	0.1269	0.1753*	0.1753*	0.321	0.0728*	-0.0399*	0.1011*	0.0728*	-0.0399*	0.1011*
Renda04	0.22363	-0.013051	0.07079	0.8960*	0.0028	0.1492	0.1512*	0.1512*	0.275	0.0449**	-0.0191**	0.1208*	0.0449**	-0.0191**	0.1208*
Renda05	0.27203	-0.011407	0.07223	0.9740*	0.0001	0.1046	0.1610	0.1610	0.0211	0.0073	-0.0258*	0.1315**	0.0073	-0.0258*	0.1315**
CSE	-0.0086	0.032172	0.06686	-0.0246	0.0473*	0.0795*	-0.0065	-0.0065	0.0765*	-0.0002	0.0389*	0.0401*	-0.0002	0.0389*	0.0401*
Infra	0.04038	0.04917	0.1042	0.0588***	0.0636*	0.0553*	0.0423**	0.0423**	0.1152*	0.0368**	0.0086	0.1219*	0.0368**	0.0086	0.1219*
Constante	6.3137	6.4396	6.1766	5.6827*	6.3425*	5.9520*	6.4370*	6.4370*	6.2322*	6.4977*	6.3135*	6.3135*	6.4977*	6.3135*	6.3135*

Elaboração dos autores.

Obs.: Significância (*) = 1%; (**) = 5%; (***) = 10%.

TABELA B.5
Estimativas da regressão quantílica incondicional por escola: matemática

Variáveis	MQO				0,1				0,5				0,9								
	Particulares		Regulares		Particulares		Regulares		Particulares		Regulares		Particulares		Regulares						
	EEEPs	Regulares	EEEPs	Regulares	EEEPs	Regulares	EEEPs	Regulares	EEEPs	Regulares	EEEPs	Regulares	EEEPs	Regulares	EEEPs	Regulares					
Idade	-0.03228	-0.01073	-0.0270*	-0.0155*	-0.0428*	-0.0014*	-0.0140*	-0.0040*	-0.0204*	-0.0023*	-0.0052*	0.09407	0.07791	0.0537*	0.0534*	0.1180*	0.0966*	0.0556*	0.1058*	0.0833*	0.0780*
Sexo	-0.00199	-0.04258	0.0270	-0.0806	-0.0197	-0.0069	-0.0647**	-0.0187*	-0.0159	-0.0423**	-0.0133**	0.00296	0.01344	0.00697	0.0021	0.0004	0.0126**	0.0087*	-0.0020	0.0239*	0.0136*
Raça	0.02065	0.04652	0.0089	0.0199*	0.0278*	0.0037***	0.0605*	0.0164*	0.0264*	0.0458*	0.0332*	-0.00138	0.01945	-0.00036	0.0029	-0.0016	0.0238*	0.0006	0.0012	0.0135**	-0.0028
Anos_Médio	-0.00360	-0.05130	0.0085	-0.0139	-0.0202	0.0002	-0.0738*	-0.0226	0.0128	-0.0431***	-0.0014	0.00785	-0.05358	-0.01513	0.0125	-0.0214***	-0.0688*	-0.0210	0.0238	-0.0736*	-0.0004
E_Pai1	0.01911	-0.04084	0.0214	-0.0034	0.0075	0.0047	-0.0548*	-0.0094	0.0178	-0.0603*	0.0204	0.0332	-0.00840	0.00484	0.0454	-0.0167	-0.0588*	-0.0058	0.0544	-0.0089	0.0497
E_Pai2	-0.01755	-0.0497	-0.0127	-0.0289	-0.0281*	-0.0231*	-0.0446*	-0.0260	-0.0277	-0.0577*	-0.1057*	-0.01201	-0.05807	-0.04708	0.0061	-0.0446*	-0.0629*	-0.0412*	-0.0166	-0.0590*	-0.0963*
E_Pai3	0.00469	-0.03321	0.0198	-0.0178**	0.0084	-0.0084	-0.0405*	0.0018	0.0054	-0.0527*	-0.0531*	-0.00002	0.00030	-0.01634	0.0030	-0.0130	0.0017	-0.0152	0.0165	-0.0059	-0.037***
E_Pai4	0.13414	-0.11765	0.07438	-0.126*	0.0753	0.1580*	-0.1573*	0.0300	0.0300	-0.1518*	0.0073	0.15550	-0.16037	0.08341	-0.1390*	-0.0810*	-0.2002*	0.0647	0.0259**	-0.1607*	0.0206
Renda00	0.20208	-0.1394	0.105161	-0.1123*	0.0926	0.2546*	-0.1646*	0.0877	0.0826*	-0.1751*	0.0542	0.21470	-0.09055	0.11875	-0.1134*	-0.0144	-0.0947*	0.0921	0.0993*	-0.1529*	0.0818
Renda01	0.17182	-0.04954	0.13176	-0.1426*	0.1076	0.2182*	-0.0484*	0.1047	0.0775	-0.0840*	0.1072	0.19145	-0.02332	0.16781	-0.0267	-0.0087	-0.0169	0.1612***	-0.0450	-0.0488***	0.1753
Renda02	0.00407	0.07704	0.06268	0.0261	-0.0239	0.0828*	0.0748*	0.0769*	0.0343	0.0828*	0.0748*	0.01488	0.18425	0.09563	-0.0087	0.1327*	0.0245***	0.0093	0.0343	0.0665*	0.2022*
Renda03	6.50249	6.4935	6.0235	6.4286*	5.7783*	6.6696*	6.2523*	6.0425*	6.6641*	6.7093*	6.3245*	0.01488	0.18425	0.09563	-0.0087	0.1327*	0.0245***	0.0093	0.0343	0.0665*	0.2022*
Renda04																					
Renda05																					
CSE																					
Infra																					
Constante																					

Elaboração dos autores.

Obs.: Significância (*) = 1%; (**) = 5%; (***) = 10%.

TABELA B.6
Estimativas da regressão quantílica incondicional por escola: redação

Variáveis	MQO						0.1						0.5						0.9					
	Particulares		Regulares		EEEPs		Particulares		Regulares		EEEPs		Particulares		Regulares		EEEPs		Particulares		Regulares			
	EEEPs	Particulares	Regulares	EEEPs	Particulares	Regulares	EEEPs	Particulares	Regulares	EEEPs	Particulares	Regulares	EEEPs	Particulares	Regulares	EEEPs	Particulares	Regulares	EEEPs	Particulares	Regulares			
Idade	-0.04414	-0.01073	-0.01197	-0.0749*	-0.0605*	-0.0156*	-0.0356*	-0.0103*	-0.0130*	-0.0227*	-0.0034*	-0.0052*	-0.04618	0.07791	-0.10008	-0.0602*	-0.0998*	-0.1309*	-0.0471*	-0.0402*	-0.0589*	-0.0597*		
Sexo	0.04923	-0.04258	-0.04391	0.0853	0.0314	-0.1038*	0.0605	0.0078	-0.030**	0.0606	-0.0434***	-0.0015	0.00923	0.01344	0.01889	0.0485*	0.0263**	0.0130***	0.0037	0.0121**	0.0193*	0.0207*		
Raça	0.03442	0.04652	0.02935	0.0505*	0.0429*	0.0383*	0.0264*	0.0390*	0.0334*	0.0229*	0.0269*	0.0248*	-0.00039	0.01945	-0.00145	0.0263***	0.0230***	-0.0198*	0.0000	0.0136**	0.0026	0.0068***		
Anos_Médio	-0.00842	-0.05130	-0.10362	-0.0352	-0.1461***	-0.0728**	-0.0238	-0.0775*	-0.1067*	-0.0584	-0.0513**	-0.0704**	0.01465	-0.05358	-0.07663	0.0117	-0.1112*	-0.0408	-0.113	-0.0677*	-0.0233	-0.062***		
E_Pai1	0.04047	-0.04084	-0.01195	0.0477	-0.0769*	0.0244	0.0078	-0.0352*	0.0275	0.0008	-0.0288**	-0.0185	0.02934	-0.000840	-0.00007	0.0489	-0.0537*	0.0217	-0.034	-0.0152	-0.0022	0.0090		
E_Pai2	-0.00935	-0.0497	-0.08757	-0.0880	-0.3109**	-0.0734*	-0.0125	-0.0202	-0.1065*	0.0266	-0.0099	-0.1103*	0.00237	-0.05807	-0.07848	-0.0065	-0.0762*	-0.0516*	-0.042	-0.0380*	0.0146	-0.1048*		
E_mae1	0.02456	-0.03321	-0.02257	0.0204	-0.0073	0.0032	0.0205	-0.0264*	-0.0272	0.0302	-0.0373*	-0.0668*	0.01951	0.000030	-0.00823	0.0223	-0.0088	0.0024	0.030	-0.0162***	0.0340	-0.042***		
E_mae2	0.23637	-0.011765	0.06942	-0.2363*	-0.1798	0.2184	0.2844*	-0.0867*	-0.0905	0.1211*	-0.1045*	0.0045	0.26106	-0.016037	0.09074	-0.2192*	-0.1468*	0.2403	0.2844*	-0.0867*	0.1689*	0.0195		
E_mae3	0.30796	-0.01394	0.13654	-0.1581*	-0.0932*	0.2843	0.3227*	-0.0583*	-0.0375	0.1689*	-0.0966*	0.0477	0.31917	-0.09055	0.16031	-0.1710*	-0.0148	0.2761	0.3301*	-0.0273**	0.2225*	0.0897		
E_mae4	0.29146	-0.04954	0.11805	-0.3329*	-0.0224	0.1828	0.3209*	-0.0025	-0.0305	0.2209*	-0.0682*	0.1064	0.29146	-0.04954	0.11805	-0.3329*	-0.0224	0.1828	0.3209*	-0.0025	0.2209*	0.1064		
Renda00	0.01120	0.007704	0.20595	0.0946***	0.1780*	0.2451*	0.0073	0.1106*	0.2328*	-0.0335	0.0959*	0.1102*	0.34757	-0.02332	0.18990	0.0329	-0.0071	0.1755	0.3123***	-0.0114	0.0663	0.2382		
Renda01	-0.00549	0.018425	0.22785	0.0535	0.2588*	0.2727*	-0.0392	0.269*	0.2204*	-0.0403	0.1097*	0.1832*	0.01120	0.007704	0.20595	0.0946***	0.1780*	0.2451*	0.0073	0.1106*	0.2328*	0.1102*		
Renda02	6.68635	6.49352	6.20832	7.2347*	7.0221*	5.6730*	6.6280*	6.5642*	6.4721*	6.8421*	6.8747*	6.5846*	6.68635	6.49352	6.20832	7.2347*	7.0221*	5.6730*	6.6280*	6.5642*	6.4721*	6.8421*		

Elaboração dos autores.

Obs.: Significância (*) = 1%; (**) = 5%; (***) = 10%.

APÊNDICE C

BALANCEAMENTO POR ENTROPIA

Considere amostras aleatórias extraídas de unidades de tratados e de controle, nas quais se inserem as definições a seguir.

- 1) D_i é uma variável binária que assume o valor 1, se a unidade pertence ao tratamento, e 0, se pertence ao controle.
- 2) X é uma matriz composta pelos elementos x_{ij} , referentes aos valores da variável exógena predeterminada j na unidade i .
- 3) A densidade das covariadas nas amostras de tratamento e controle são dadas por $f_{X|D_i=1}(x)$ e $f_{X|D_i=0}(x)$, respectivamente.
- 4) A variável resultado observado é expressa por $Y_i = Y_i(1)D_i + (1 - D_i)Y_i(0)$.

O efeito médio do tratamento sobre os tratados (EMTT) é dado por $\tau = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1]$, em que a primeira média pode ser diretamente identificada do grupo de tratados, mas a segunda corresponde ao contrafactual, o qual não é observado. Rosenbaum e Rubin (1983) apontam que, assumindo seleção nos observáveis, $Y(0) \perp D|X$, e sobreposição, $\Pr(D_i = 1|X_i = x) < 1$ para todo x no suporte de $f_{X|D_i=1}$, o EMTT é identificado como:

$$\tau = E[Y_i|D_i = 1] - \int E[Y_i|X_i = x, D_i = 0] f_{X|D_i=1}(x) dx. \quad (1)$$

Para estimar o contrafactual, a distribuição da covariável no grupo de controle necessita ser ajustada para torná-la semelhante à distribuição no grupo de tratamento, tal que o indicador de tratamento D se torne mais perto de ser ortogonal em relação às covariáveis. Uma variedade de métodos de pré-processamento de dados – tais como pareamento pelo vizinho mais próximo, cem e escore de propensão – tem sido proposta para reduzir o desequilíbrio na distribuição de variáveis de controle. Uma vez ajustadas, as distribuições de variáveis independentes e os métodos de análise padrão, tais como a regressão, podem ser posteriormente utilizados para estimar o tratamento com menor erro e o modelo de dependência (Imbens, 2004; Rubin, 2006; Ho *et al.*, 2007; Sekhon, 2009).

Considere o caso mais simples em que o efeito do tratamento nos dados pré-processados é estimado usando a diferença nos resultados médios entre os grupos de tratados e o controle ajustado, cujo método muito utilizado na literatura é o escore de propensão ponderado (Hirano, Imbens e Ridder, 2003). A média contrafactual é estimada como segue:

$$E[Y_i(0)|\widehat{D}_i = 1] = \frac{\sum_{\{i|D = 0\}} Y_i d_i}{\sum_{\{i|D = 0\}} d_i} \quad (2)$$

As unidades de controle recebem um peso dado por $d_i = \frac{\hat{p}(x_i)}{1 - \hat{p}(x_i)}$, $\hat{p}(x_i)$ na equação (2), que é o escore de propensão, o qual é comumente estimado através de uma regressão *probit* ou *logit*. Se este modelo estiver corretamente especificado, então o peso estimado d_i assegura que a distribuição da covariável das unidades de controle reponderadas corresponda à distribuição no grupo de tratamento. No entanto, na prática, essa abordagem muitas vezes não consegue equilibrar conjuntamente todas as covariáveis.

O balanceamento por entropia generaliza a abordagem de ponderação do escore de propensão, ao estimar os pesos diretamente de um conjunto de restrições de equilíbrio que exploram o conhecimento do pesquisador sobre os momentos amostrais. Considere w_i o peso do balanceamento por entropia escolhido para cada unidade de controle, o qual foi encontrado pelo seguinte esquema de reponderação que minimiza a distância métrica de entropia:

$$\min_{w_i} H(w) = \sum_{\{i|D=0\}} w_i \log(w_i/q_i) \quad (3)$$

Sujeito às restrições de equilíbrio e normalização:

$$\sum_{\{i|D = 0\}} w_i c_{ri}(X_i) = m_r \quad \text{com } r \in 1, \dots, R \quad (4)$$

$$\sum_{\{i|D = 0\}} w_i = 1 \quad (5)$$

$$w_i \geq 0 \text{ para todo } i, \text{ tal que } D = 0, \quad (6)$$

em que $q_i = 1/n$ é um peso base, sendo n o tamanho da amostra das unidades de controle, e $c_{ri}(X_i) = m_r$ descreve um conjunto de R restrições referentes aos momentos das covariadas no grupo de controle reponderados. Inicialmente, escolhe-se a covariada que será incluída na reponderação. Para cada covariada, especifica-se um conjunto de restrições de balanceamento (equação 4) para equiparar os momentos das distribuições das covariadas entre os grupos de tratamento e controles reponderados. As restrições de momentos podem ser a média (primeiro momento), a variância (segundo momento) e a assimetria (terceiro momento).

Dessa forma, o balanceamento por entropia procura, para um conjunto de unidades, pesos $W = [w_1, \dots, w_{n_0}]'$, nos quais minimiza a equação (3), distancia-se de entropia entre W e o vetor-base de pesos $Q = [q_1, \dots, q_{n_0}]'$ e sujeita as

restrições de balanceamento (equação 4), de normalização (equação 5) e de não negatividade (equação 6). Devido à sua propriedade de minimizar os desequilíbrios entre os grupos, esse método foi aqui aplicado.

REFERÊNCIAS

- HIRANO, K.; IMBENS, G.; RIDDER, G. Efficient estimation of average treatment effects using the estimated propensity score. **Econometrica**, v. 71, n. 4, p. 1161-1189, 2003.
- HO, D. *et al.* Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. **Political Analysis**, v. 15, n. 3, p. 199-236, 2007.
- IMBENS, G. Nonparametric estimation of average treatment effects under exogeneity: a review. **Review of Economics and Statistics**, v. 86, n. 1, p. 4-29, 2004.
- ROSENBAUM, P.; RUBIN, D. Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. **Journal of the Royal Statistical Society**, v. 45, n. 2, p. 212-218, 1983.
- RUBIN, D. B. **Matched sampling for causal effects**. Cambridge: Cambridge University Press, 2006.
- SEKHON, J. S. Opiates for the matches: matching methods for causal inference. **Annual Review of Political Science**, v. 12, p. 487-508, 2009.