

Eficiência das Escolas Públicas Urbanas das Regiões Nordeste e Sudeste do Brasil: Uma Abordagem em três Estágios

Luciana Duarte Bhering de Carvalho

Coord. de Avaliação Macrofiscal - Secretaria de Orçamento Federal do Ministério do Planejamento - SOF/MP - Endereço: SEPN 516, Lote 08 - Bloco D, Sala 418 - Asa Norte Brasília/DF - Brasil - CEP: 70770-524 - E-mail: lubhering@gmail.com

Maria da Conceição Sampaio de Sousa

Professora - Universidade Federal da Paraíba (UFPB)
Endereço: Cidade Universitária, s/n - Castelo Branco - João Pessoa/PB - Brasil
CEP: 58051-900 - E-mail: csampaiodesousa@gmail.com

Recebido em 01 de fevereiro de 2013. Aceito em 03 de maio de 2014.

Resumo

Nesse trabalho foi calculada uma medida de eficiência técnica para as escolas públicas urbanas das regiões Nordeste e Sudeste do Brasil, avaliadas pela Prova Brasil 2007, por meio da metodologia chamada de Análise da Envoltória de Dados (DEA) em três estágios, adotando-se a escola como unidade de análise. Em relação a todas as escolas estudadas, há indícios para afirmar que, mesmo descontando os fatores ambientais e aleatórios, ainda persiste a necessidade de melhorias ligadas à gestão de cada uma delas.

Palavras-Chave

Análise da Envoltória de Dados (DEA), eficiência, educação, Análise das Fronteiras Estocásticas (SFA), Nível Socioeconômico (NSE), Prova Brasil

Abstract

In this study, we calculated a measure of technical efficiency for urban public schools in the Northeast and Southeast of Brazil; evaluated by a battery of tests called Prova Brasil; year 2007, using Three-stage Data Envelopment Analysis (DEA) methodology; taking the school as unit of analysis. Based on our results we can say that there are margins for improvements in Brazilian urban public schools in terms of pure managerial ability.

Keywords

Data Envelopment Analysis (DEA), efficiency, education, Stochastic Frontier Analysis (SFA), Socioeconomic Status (SES), Prova Brasil

JEL Classification

I20, I21

1. Introdução

É patente a importância da educação na sociedade em que vivemos, seja do ponto de vista dos indivíduos, seja do ponto de vista coletivo. Hanushek e Kimko (2000), ao considerarem a qualidade da educação de cada país, além dos anos de escolaridade da população, que é a variável usual, mostraram que a educação pode ter efeitos até mesmo sobre as taxas de crescimento nacionais, o que vai ao encontro do preconizado pela teoria do capital humano. Além de todos esses benefícios diretos provenientes de melhores níveis educacionais, há também as chamadas externalidades positivas normalmente associadas à educação, tais como: redução nas taxas de criminalidade, melhorias nas condições de saúde da população e maior coesão social. Existe, pois, um grande incentivo para se investir em educação.

Em razão do exposto acima, a educação é um tema que recorrentemente figura como preocupação central dos governos. No Brasil, a situação não é diferente e fica evidente quando se examina o montante de recursos destinados a esse setor. Entre 2000 e 2008, o gasto com educação em termos percentuais do PIB passou de 3,5% para 5,3%, dos quais 2,6% e 4,1%, respectivamente, correspondem às aplicações nos níveis primário e secundário de educação (OECD, 2011). Porém, esse elevado aporte de recursos não se traduziu em melhorias substanciais dos resultados em educação, visto que a avaliação do país em estudos padronizados internacionais situa-se em patamares muito baixos, particularmente quando se leva em conta o seu nível de renda (Wilson, 2005). Temos, então, indícios para questionar se o problema da educação no Brasil se deve apenas ao volume insuficiente de investimentos na área, como afirmam alguns, ou também à má qualidade do gasto nessa área.

É nesse contexto que se situa o presente trabalho, cujo objetivo é calcular medidas de eficiência técnica das escolas públicas urbanas brasileiras das regiões Nordeste e Sudeste. Mais especificamente, buscamos responder às seguintes questões: quão eficientes são essas escolas? Que fatores estão relacionados ao seu grau de eficiência? Como os fatores que não estão sob seu controle afetam sua eficiência? Onde estão as escolas mais eficientes? As respostas para essas questões serão úteis para a identificação de práticas bem-sucedidas em algumas escolas, as quais poderão ser difundidas para as demais unidades.

Além disso, a obtenção de tais informações poderá servir de guia para uma aplicação mais racional dos recursos públicos na área da educação.

Para lidar com as questões levantadas acima, tomamos como unidade de análise a escola, a metodologia da Análise da Envoltória de Dados (DEA)¹ em Três Estágios, desenvolvida por Fried *et al.* (2002), que combina o método DEA com a Análise de Fronteira Estocástica (SFA).² Essa técnica foi concebida com o objetivo de separar a medida de eficiência resultante da aplicação do método DEA em três partes, a saber, uma atribuída a fatores externos, outra, a fatores aleatórios e a restante, ao nível de habilidade gerencial do responsável pelo processo em estudo.

Este trabalho está dividido em cinco seções além desta introdução. A seção seguinte busca dar um panorama da literatura atinente ao tema; a seção 3 aborda a metodologia escolhida; as seções 4 e 5 descrevem os dados e os resultados obtidos e, finalmente, a última seção reúne as principais conclusões do trabalho.

2. Literatura

A produção dos primeiros trabalhos sobre eficiência na área educacional se confunde com a origem do método DEA. O aspecto essencial desse método consiste na construção de uma fronteira, formada pelas Unidades Tomadoras de Decisão (DMUs)³ mais eficientes, que, assim, passam a delimitar as demais unidades, de forma que aquelas mais eficientes estejam localizadas na fronteira, ou próximas a ela, e as menos eficientes, em posição mais afastada.

Convém lembrar que o método DEA originou-se dos esforços para avaliar os resultados de um programa educacional *Follow Through*, doravante mencionado como FT (Cooper; Seiford; Tone, 2007). Desde então, outros trabalhos nessa vertente metodológica vêm sendo desenvolvidos, tanto no Brasil, como no exterior, para tratar de temas ligados à eficiência no processo produtivo educacional. Bradley,

¹ A sigla corresponde à expressão *Data Envelopment Analysis*.

² A sigla corresponde à expressão *Stochastic Frontier Analysis*.

³ A sigla corresponde à expressão *Decision Making Units*.

Johnes e Millington (2001) cotejaram trabalhos produzidos desde os fins da década de 70 até o trabalho deles próprios. No Brasil, esse tipo de pesquisa começou a produzir seus primeiros resultados em meados da década de 1990 (Delgado; Machado, 2007). Porém, os trabalhos da literatura nacional, que adotam a escola como unidade de análise, lidam com um universo de escolas pertencentes a uma ou algumas cidades, com abrangência máxima de um estado completo. Em decorrência disso muitas das amostras estudadas estão na casa das dezenas ou centenas, o que pode gerar estimativas viesadas de eficiência devido à chamada “maldição da dimensionalidade.”⁴

Uma das restrições ao método DEA, em seu formato original, advém do fato de que ele não é capaz de distinguir a ineficiência atribuída às más práticas gerenciais no setor educacional da ineficiência decorrente de diferenças socioeconômicas e demais fatores não controláveis pelos gestores da educação. Ora, em um país grande e diverso como o Brasil, a consideração de diferenças ambientais, de acesso, econômicas, institucionais, culturais, entre outras, é indispensável. Ao passo que governos locais, ou regiões limitadas e geograficamente concentradas se defrontam com ambientes socioeconômicos relativamente similares, grandes nações e continentes usualmente diferem de forma substancial no âmbito interno; por essa razão, avaliações que não levam em conta essas diferenças são questionáveis, pois responsabilizam os gestores por todas as diferenças constatadas nos indicadores de desempenho educacional. Por fim, o método DEA não leva em conta fatores aleatórios, por isso os escores de eficiência computados por essa abordagem podem ser fortemente enviesados se o processo produtivo for caracterizado por elementos estocásticos.

Concluimos, pois, que o uso de medidas de eficiência baseadas unicamente nos insumos e produtos gerenciáveis pelas escolas pode distorcer o desempenho observado de uma dada escola. É necessário, dessa forma, levar em conta, na análise da eficiência técnica, tanto os fatores aleatórios como os elementos exógenos, para melhorar a qualidade e a representatividade dos indicadores de eficiência e permitir, assim, um ordenamento inequívoco das escolas.

A inclusão de variáveis exógenas na análise não paramétrica de eficiência é um tópico importante da literatura especializada.

⁴ A “maldição da dimensionalidade” refere-se ao fato de que o aumento do número de variáveis conduz à superestimação dos escores de eficiência e reduz o poder discriminatório da técnica DEA.

Diferentes modelos se propõem a tratar esse tema. Nos modelos não paramétricos de Banker e Morey (1986) e de Ruggiero (1996), as variáveis exógenas são consideradas por meio de modificações nas restrições do sistema de programação linear. No entanto, devido aos seus muitos inconvenientes, destacando-se, entre eles, o já mencionado problema da maldição da dimensionalidade, essa abordagem foi praticamente descartada e substituída pela análise em dois ou mais estágios que combinam o método DEA com técnicas econométricas (Ray; 1991, Ruggiero; 1998, Banker; Natarajan, 2008). Nesses modelos, computam-se medidas de eficiência não paramétricas do tipo DEA, em que todos os insumos são tratados como controláveis. Nos estágios subsequentes, regridem-se os escores de eficiência nos fatores exógenos para se obter os índices de eficiência técnica “puros”. Embora as abordagens acima mencionadas considerem os fatores ambientais, elas não permitem distinguir a ineficiência dos fatores aleatórios existentes na produção de serviços educacionais. Nosso trabalho representa uma contribuição à literatura, na medida em que propõe estudar a eficiência das escolas públicas brasileiras de duas importantes regiões geográficas, o Nordeste e o Sudeste, em presença de fatores exógenos e aleatórios.

A metodologia utilizada combina o método DEA no primeiro estágio com a abordagem da SFA no segundo estágio. A vantagem dessa combinação – quando comparada com as análises já mencionadas - advém da utilização de duas técnicas que envolvem estimações de fronteiras de eficiência e permitir a distinção entre ineficiência gerencial e a decorrente de fatores externos e aleatórios. A justificativa principal para o uso da SFA no segundo estágio, ao invés dos modelos Tobit e Mínimos Quadrados Ordinários (MQO), é a existência de um termo de erro assimétrico que permite identificar o componente assimétrico, que corresponde à ineficiência gerencial, e o componente simétrico (ruído estatístico) captando, assim, de forma mais acurada, o conceito de fronteira. Ademais, a abordagem DEA-SFA, além de acomodar facilmente discrepâncias estatísticas e fatores exógenos, permite que testes estatísticos padrão sejam usados para se fazer inferência. Dessa forma, é possível contornar uma limitação importante do método DEA que decorre da hipótese da inexistência de erros de mensuração e, ainda, considerar os efeitos das variáveis ambientais. A favor do método DEA-SFA lista-se, ainda, o fato de que esses modelos têm obtido melhor desempenho em simulações

de Monte Carlo (Cordero-Ferrera *et al.*, 2009). No que se segue, detalharemos essa metodologia.

3. Metodologia

No primeiro estágio aplica-se o método DEA - BCC⁵ aos insumos e produtos relativos ao processo em análise. No segundo estágio, utiliza-se a técnica SFA, (Aigner *et al.*, 1977). Nesse passo, a insuficiência de cada produto,⁶ identificada no primeiro estágio, é considerada variável dependente, enquanto as variáveis chamadas de ambientais, que não estão sob controle do gestor do processo, são consideradas variáveis independentes. A partir dos resultados desse estágio as variáveis de produto são ajustadas de modo a expurgar os fatores aleatórios e ambientais que as influenciam. Feita essa correção, recalculam-se os escores DEA utilizando os dados ajustados. Desse novo cálculo, espera-se que as medidas obtidas reflitam primordialmente o grau de eficiência gerencial das unidades em estudo.

3.1 1^o Estágio – O modelo DEA-BCC produto orientado

Nessa metodologia, o Conjunto de Possibilidade de Produção (CPP) é definido a partir do conhecimento das variáveis relativas aos insumos e produtos utilizados pelas DMUs em seu processo produtivo. Tal conjunto, representado por T, abrange todas as correspondências factíveis entre os insumos $x \in \mathfrak{R}_+^n$ capazes de produzir os produtos $y \in \mathfrak{R}_+^m$, e pode ser representado como:

$$T = \{(x, y) \in \mathfrak{R}_+^{n+m} \mid x \text{ pode produzir } y\} \quad (1)$$

Nesse conjunto, cada DMU i se depara com o seguinte problema:

$$\begin{aligned} & \max_{\theta, \lambda} \theta \\ & \text{s. a. } X\lambda \leq x_i \\ & \theta y_i - Y\lambda \leq 0 \\ & e\lambda = 1 \\ & \lambda \geq 0 \end{aligned} \quad (2)$$

⁵ Formulação proposta por Banker, Charnes e Cooper (1984) que considera retornos variáveis de escala. Por essa razão, tal modelo é também chamado DEA - VRS, *Variable Returns Scale*.

⁶ Devido à escolha da orientação da análise para produto. No caso da orientação para insumo, a variável dependente seria o excesso de cada insumo.

Em que Y e X são os vetores dos produtos e dos insumos de todas as DMUs, com dimensões $M \times I$ e $N \times I$, respectivamente. O escalar θ deve ser maior ou igual a um e é limitado inferiormente por um; quando assume valor igual a 1 indica eficiência técnica plena da DMU. O escalar θ é chamado de medida de eficiência de Farrell (1957). O vetor λ expressa a intensidade com que insumos e produtos devem ser empregados e produzidos, respectivamente, para que a DMU se torne tecnicamente eficiente. As quantidades de insumos e produtos multiplicadas por essa intensidade, portanto, representam as metas (*targets*) que as DMUs devem buscar alcançar a fim de se tornarem eficientes. Esses *targets* são as projeções dos pontos que representam o processo produtivo de cada unidade na fronteira, que é o *locus* das combinações de insumos e produtos consideradas tecnicamente eficientes.

Em relação à interpretação da medida de eficiência de Farrell (1957), caso a DMU apresente uma medida de eficiência $\theta=1,25$, por exemplo, isso significa que ela deve aumentar em 25% o seu mix de produtos, mantidas as proporções entre eles, e fixados seus insumos, a fim de se tornar tecnicamente eficiente. Essa medida também é apresentada pelo seu inverso, nesse caso, igual a 0,80, que corresponde ao conceito de função distância do produto de Shephard ⁷ (Fried; Lovell; Schmidt, 2008).

A escolha da orientação produto reconhece que a finalidade primordial das instituições públicas voltadas para a educação não é minimizar custos. Todavia, não podemos ignorar o fato de que os gastos em educação concorrem com outros de importância equivalente, e.g., gastos em saúde, em um contexto de restrição orçamentária. Dessa forma, quando não houver mais a necessidade de se ampliar o atendimento, tornar-se-á necessária a redução de custos. Nada impede, portanto, que se aborde o problema em questão desses dois ângulos. Porém, como já esclarecido, optamos nesse trabalho pela orientação a produto.

⁷ $D_0(x, y) = \min \left\{ \lambda : \left(\frac{y}{\lambda} \right) \in P(x) \right\}$, sendo $P(x) = \{y : (x, y) \in T\}$ a tecnologia de produção. (4)

3.2 2^o Estágio - O modelo DEA-SFA

A abordagem DEA-SFA consiste na estimação de uma fronteira estocástica no seguinte formato:

$$S_{ni}^+ = f_n(z_n; \beta_n) + v_{ni} - u_{ni}, \quad n = 1, \dots, N, \quad i = 1, \dots, I, \quad u_{ni} \geq 0 \quad (3)$$

Em que $S_{ni}^+ = Y\lambda - y_{ni}$ é o *slack* total relativo ao produto “n” e à DMU “i”, equivalente ao hiato ou à insuficiência total de cada produto de dada DMU; z_{ni} são as variáveis ambientais; β_i é o vetor de parâmetros a ser estimado por Máxima Verossimilhança e a expressão $v_{ni} - u_{ni}$ corresponde ao erro composto por uma parte estocástica e simétrica e outra parte assimétrica e associada às ineficiências gerenciais, respectivamente. As suposições relativas aos componentes do erro são: $v_{ni} \sim N(0, \sigma_{vn}^2)$ e $u_{ni} \sim N^+(\mu_n, \sigma_{un}^2)$, ou seja, que seu componente aleatório seja distribuído normalmente, com média zero e variância σ_{vn}^2 , e que o componente ligado às ineficiências gerenciais seja distribuído conforme uma Normal Truncada, com média μ_n e variância σ_{un}^2 . Assume-se ainda que cada parte do erro é distribuída independentemente da outra e que o erro composto não está correlacionado às variáveis independentes. Aplicada a formulação SFA aos *slacks* totais de cada variável de produto, não só teremos a tripartição da medida de ineficiência obtida no primeiro estágio, como também as direções e magnitudes dos efeitos atribuídos às variáveis ambientais sobre tal medida. Note-se, por fim, que a variância de cada componente do erro pode ser tanto homocedástica como heterocedástica. A heterocedasticidade está geralmente relacionada ao tamanho das unidades em análise. Para se considerar esse aspecto, variáveis relacionadas ao tamanho das unidades foram incluídas na estimação de cada uma dessas variâncias, da seguinte maneira: $\sigma_{vn}^2 = \exp(w_n \delta)$ e/ ou $\sigma_{un}^2 = \exp(w_n \delta)$, em que w_n são as referidas variáveis de tamanho. Não levar em conta esse aspecto pode levar à superestimação do intercepto da fronteira e a vieses nos demais coeficientes (Caudill; Ford, 1993).

Ainda no segundo estágio, o passo seguinte consiste no ajuste dos valores originais das variáveis de produto a partir das estimações geradas pela SFA, cuja intenção é eliminar as diferenças das DMUs no que se refere ao ambiente e aos fatores aleatórios, a fim de compará-las apenas no que tange à capacidade gerencial de cada uma. A proposta de Fried *et al.* (2002) para tanto é penalizar as unidades muito favorecidas pelos fatores ambientais e aleatórios, diminuindo

seus produtos no montante equivalente a esses fatores que as favorecem, ou compensar as unidades menos favorecidas na direção oposta. A segunda opção foi a escolhida por uma questão de conveniência, já que a segunda poderia gerar valores negativos para algumas unidades, o que não é compatível com o método DEA–BCC, a ser utilizada no passo seguinte. O referido ajuste é o que se segue:

$$y_{ni}^a = y_{ni} + [z_{ni} \hat{\beta}_n - \min \{z_{ni} \hat{\beta}_n\}] + [\hat{v}_{ni} - \min \{\hat{v}_{ni}\}] \quad (5)$$

$$n = 1, \dots, N, \quad i = 1, \dots, I$$

Em que y_{ni} refere-se ao valor original da variável do produto “n” da DMU “i” e o mesmo símbolo, com o sobrescrito “a”, representa o valor do produto em questão ajustado. Os demais símbolos seguem a mesma notação da fronteira apresentada no início dessa seção. Tone e Tsuitsui (2009) argumentaram que a adição ou subtração de uma constante aos valores originais dos produtos (operação chamada de translação), ao mudar a amplitude dos dados originais, pode causar grande viés nos escores da DEA, já que os modelos DEA-BCC produto (insumo) orientados não apresentam a propriedade da Invariância à Translação nos valores dos produtos (insumos), como nos explicam Cooper, Seiford e Tone (2007, p. 97). Desse modo, a fim de corrigir esse problema, os autores propõem o seguinte ajuste:

$$y_{ni}^{AA} = \frac{y_{imax} - y_{imin}}{y_{imax}^A - y_{imin}^A} (y_{ni}^A - y_{imin}^A) + y_{imin} \quad (6)$$

Em que: $y_{ni}^A = y_{ni} + z_{ni} \hat{\beta}_n + \hat{v}_{ni}$, $y_{imax} = \max \{y_{ni}\}$, $y_{imin} = \min \{y_{ni}\}$,
 $y_{imax}^A = \max \{y_{ni}^A\}$ e $y_{imin}^A = \min \{y_{ni}^A\}$

A correção acima proposta, por manter a amplitude dos dados originais, não introduz o viés já comentado. Por esse motivo, optamos por essa forma de ajustar os dados.

3.3 3^o Estágio – DEA-BCC com dados ajustados

Nessa última parte calculamos o mesmo modelo DEA utilizado inicialmente, porém, com os dados ajustados conforme descrito na seção anterior. Por fim, comparamos as medidas de eficiência obtidas no primeiro estágio com aquelas obtidas no terceiro. Constatada estatisticamente a diferença entre a medida de eficiência do

primeiro e a do terceiro estágio, teremos um indicativo de que os fatores ambientais, além daqueles considerados estocásticos, influenciam a eficiência das unidades avaliadas.

4. Dados

Os dados utilizados neste trabalho foram extraídos dos Microdados do Censo Escolar 2007 e daqueles relativos à Prova Brasil desse mesmo ano, disponíveis no site do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Em 2007, foram coletadas informações de 198.507 escolas em atividade, total que engloba instituições públicas, privadas, urbanas e rurais. Contudo, o presente trabalho se ocupará de um subgrupo das 78.138 escolas públicas e urbanas ativas do país, que são aquelas avaliadas pela Prova Brasil 2007.

A Prova Brasil é parte integrante do Sistema de Avaliação da Educação Básica (SAEB). Tal avaliação é realizada a cada dois anos, desde 2005, por meio da aplicação de provas de português e matemática aos alunos da 4ª série e da 8ª série e/ou do 5º e do 9º ano do Ensino Fundamental Regular, das escolas que possuam pelo menos vinte alunos matriculados na série avaliada e cujos resultados podem ser comparados ao longo dos anos, graças à metodologia de elaboração dessa avaliação, a chamada Teoria de Resposta ao Item (TRI).

Os resultados dos alunos nas avaliações variam de 0 a 350 pontos na prova de Língua Portuguesa e, de 0 a 425 pontos na prova de Matemática, tanto para a 4ª como para a 8ª série. A Prova Brasil 2007 foi aplicada a todas as escolas públicas urbanas com turmas de 4ª e/ou 8ª séries e cujos gestores tinham aderido à Prova Brasil. Nesse conjunto há escolas que ofertam apenas 4ª ou 8ª série. A fim de comparar unidades mais homogêneas, decidimos por incluir apenas aquelas que oferecem 4ª e 8ª séries. Dessa decisão, nosso conjunto diminuiu de cerca de 48 mil escolas para pouco mais de 16 mil. A última restrição aplicada ao conjunto a de estudar as escolas das regiões Nordeste e Sudeste do Brasil, culminou na redução do grupo de escolas a ser estudado para quase 10 mil escolas. Finalmente, desse conjunto, terminamos com 8.155 escolas, devido às perdas de observações que ocorreram em função de informações ausentes e incompletas por parte de algumas unidades.

4.1 Variáveis de insumo e produto

No tocante às variáveis de produto, vem persistindo ao longo do tempo nessa literatura o uso de notas em exames padronizados, como medidas de resultado. Por essa razão, as variáveis escolhidas como produtos/resultados são as notas medianas de matemática da Prova Brasil de cada escola para a 4ª e para a 8ª série e o número total de alunos matriculados em cada uma delas. Essa escolha teve como critério a disponibilidade dos dados e a compatibilidade com a literatura. A decisão pela mediana das notas de matemática, ao invés da média dessa variável, deu-se pela tentativa de se captar, em alguma medida, aspectos distribucionais do desempenho escolar dentro de cada escola, aspectos esses ocultados pela média. O que é de extrema importância, já que devemos considerar como boa escola aquela que atenua os efeitos do processo chamado pelos sociólogos de “reprodução social” (Riordan, 2004, p. 20), especialmente para os que ocupam a base da hierarquia social.

Cabe explicar o porquê do uso das notas de matemática apenas. Convém salientar que os resultados dos alunos em português e matemática estão relacionados de maneira significativa, como era de se esperar, apresentando os seguintes coeficientes de correlação: 64,06% e 58,61% para os alunos de 4ª e 8ª série, respectivamente. A escolha pela nota de matemática se deu, no entanto, por dois motivos: (a) a disciplina em questão é tipicamente escolar, em contraposição ao aprendizado da linguagem, que se dá nos mais diversos ambientes (Franco *et al.*, 2007); e (b) os alunos apresentaram pior rendimento em matemática.⁸ Em decorrência dessa constatação, consideramos que esse resultado merece maior atenção, já que precisa ser melhorado com maior urgência.

Por fim, decidiu-se por utilizar o número total de alunos, ao invés daquele restrito à Educação Fundamental (EF), para que fosse mantida a compatibilidade com o insumo número de funcionários, a ser comentado a seguir, que engloba, além do número de professores, o

⁸ Cabe destacar que, em termos absolutos, a média das notas de matemática, tanto nacional como das duas regiões estudadas, é maior que a da prova de português. Todavia, como as escalas dos testes de cada disciplina são diferentes, conforme já comentado neste trabalho, ao dividirmos as notas pelo máximo da escala de cada disciplina, enxergamos que o rendimento dos alunos em matemática é pior que o verificado em português. Para termos uma ideia, em âmbito nacional os alunos atingem em média 49% e 65% dos pontos da prova de português, resultado esse relativo à 4ª e à 8ª série, respectivamente, enquanto na prova de matemática eles acertam em média 45% e 57%.

pessoal administrativo das escolas, que não poderia ser apropriado para cada nível, modalidade e etapa de ensino.

Em relação às variáveis de insumo, dispomos apenas o número de funcionários da escola,⁹ que compreende não só os ligados às funções docentes como também aqueles ligados às atividades administrativas da escola. Em que pese o fato de ser apenas uma variável, ela representa parte significativa do orçamento de cada escola para o caso do Brasil (Farenzena, 2005). Outra variável considerada como insumo é a Taxa de Distorção Série Idade (TDI) da EF, que é o percentual de alunos do ensino fundamental de cada escola em idade superior em dois anos à considerada adequada para cada série. Essa taxa reflete um resultado inerente à atividade escolar, mas que não é desejado, sendo, pois, tratado como insumo (Thanassoulis; Portela; Despić, 2008, p. 301). A Tabela 01 apresenta as estatísticas descritivas para insumos e produtos.

Tabela 01 - Estatísticas descritivas das variáveis de produtos/resultados e insumos

Produtos/Resultados					
Variáveis	Nº de obs	Média	Desvio Padrão	Mínimo	Máximo
Nota mediana de matemática - 4ª Série	8.155	183,70	20,81	120,36	289,24
Nota mediana de matemática - 8ª Série	8.155	234,10	20,84	170,83	357,37
Número de alunos	8.155	873	479	70	4.092
Insumos					
Variáveis	Nº de obs	Média	Desvio Padrão	Mínimo	Máximo
Número de funcionários da escola	8.155	64	31	12	345
Taxa de Distorção Série-Idade (TDI) do Ensino Fundamental (%)	8.155	31,78	16,69	0,50	94,20

Fonte: As medianas das notas são cálculos próprios com base nos dados da Prova Brasil 2007 (INEP, 2007b), o número de alunos e de funcionários foram extraídos do Censo Escolar (INEP, 2007a) e a TDI, do site do INEP (INEP, 2011).

A respeito dos dados extremos, utilizamos o método *jackstrap* de detecção de *outliers* (Sampaio de Sousa; Štosić, 2005) para inspecionar a base de dados. No entanto, as medidas de influência (alavancagens) calculadas por este método revelaram-se pouco importantes para a grande maioria das observações. Das 8.155 escolas, com dados consistentes, apenas 320 apresentaram alavancagens não

⁹ No Censo Escolar há variáveis relativas ao número de computadores por escola. Porém, como há muitos valores *missing*, utilizar tais variáveis implicaria a perda de muitas observações. Há o mesmo problema para outras variáveis relacionadas a insumos.

nulas – menos de 4% das observações. O exame desses resultados mostrou que as observações com maiores alavancagens correspondem às escolas com valores de insumos e produtos realistas, que são, portanto, exemplos de melhores práticas. Diante disso, não foi necessária a eliminação de nenhuma observação, sendo todas as escolas selecionadas com probabilidade 1.¹⁰

5. Resultados

O Gráfico 01 apresenta o histograma dos escores obtidos no primeiro estágio do modelo. A medida de Farrell equivalente ao escore médio apresentado é 1,32 ($1 \div 0,76$), indicando que, em média, as escolas devem aumentar seus resultados, mantidas as proporções entre eles, em 32% para se tornarem plenamente eficientes. A ineficiência máxima, equivalente ao escore mínimo, é de 78% ($1 \div 0,56 = 1,78$).

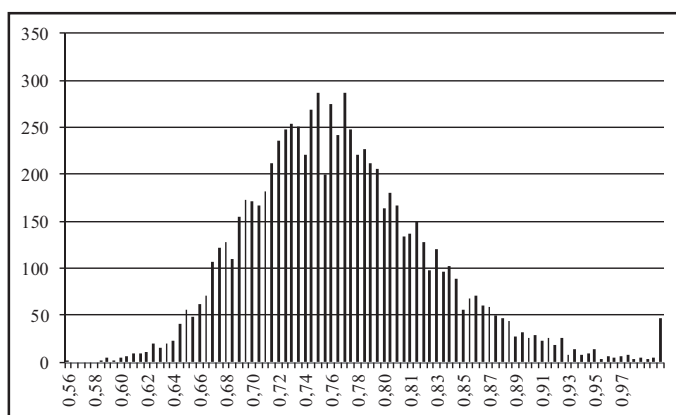


Gráfico 1- Histograma relativo aos escores gerados no 1º Estágio

O conceito de *slack* total, como apresentado na seção desse trabalho relativa à metodologia, refere-se à diferença entre as metas calculadas pelo algoritmo DEA para as variáveis de produto e seus valores originais. A Tabela 02 apresenta as estatísticas descritivas para os *slacks* totais.

¹⁰ Maiores detalhes sobre os resultados da aplicação da metodologia *Jackstrap* neste trabalho poderão ser obtidos juntos aos autores.

Tabela 02 - Estatísticas descritivas relativas aos *slacks* das variáveis de produto

Variáveis	Nº de obs	Média	Desvio Padrão	Mínimo	Máximo
Slack nota mediana matemática 4ª série	8.155	58,94	17,02	0,00	113,63
Slack nota mediana matemática 8ª série	8.155	74,47	21,97	0,00	136,52
Slack número de alunos	8.155	282	145	0	1.143

Fonte: Elaboração própria com base nos resultados gerados pelo software DEA-Solver-PRO 8.0 (SAITECH, 2012).

5.1 Resultados do 2^o Estágio

Nessa seção serão apresentadas e comentadas as estimações dos parâmetros das três fronteiras estocásticas, além das variáveis de produto ajustadas, que serão utilizadas no último estágio. As três fronteiras de eficiência, cujas variáveis dependentes são os *slacks* dos três produtos utilizados na fronteira DEA, computada no primeiro estágio, a saber, nota de matemática da 4ª série, nota de matemática da 8ª série e número de alunos. Como as variáveis dependentes de cada uma das fronteiras são insuficiências relativas a cada um dos resultados esperados das escolas, devem ser lidas como ineficiências. Por essa razão, os coeficientes das variáveis independentes que apresentam sinal negativo mostram que a variável independente em questão tem efeito redutor sobre a ineficiência, analogamente para o caso contrário. Como não temos informações *a priori* sobre a forma funcional da *fronteira*, assumimos linearidade, como é o uso corrente nesse caso. A forma funcional adotada foi a seguinte:

$$S_{ni} = \beta_0 + \sum_{i=1}^{35} \beta_{ni} z_{ni} + v_{ni} + u_{ni} \quad (7)$$

Os parâmetros e variáveis da expressão acima foram previamente definidos na seção 3 deste trabalho. Note-se que as variáveis ambientais z são aquelas ligadas aos fatores sobre os quais as escolas possuem pouco ou nenhum controle, tais como: composição étnica e Nível Socioeconômico (NSE) de seus alunos, composição étnica e qualificação de seus docentes, tamanho das escolas, esfera administrativa a que pertencem (estadual, municipal ou federal), características dos seus diretores e dos municípios em que se encontram, dentre outras.

Essas serão as variáveis independentes que serão utilizadas nas fronteiras estocásticas, cujas variáveis dependentes serão os *slacks* totais de cada uma das três variáveis de produto. Abaixo apresentaremos as estatísticas descritivas dessas variáveis. Elas se dividem em três grupos: proporções que variam de 0 a 1; variáveis *dummies* e as variáveis que expressam alguma quantidade (número de salas, número de turmas, etc.). A única que foge a esse padrão é a variável *info*, variável discreta que assume valores 0, 1 ou 2, indicando, respectivamente, se a escola não possui laboratório de informática, nem outros computadores; se a escola possui um ou outro; ou se a escola possui os dois.

As variáveis de proporção têm seus rótulos iniciados pela letra “p”, seguidas de um ou dois números ou de uma expressão que indicam a que a etapa ou a que modalidade de ensino se referem, e.g., 4^a série, 1^a a 4^a série, ensino fundamental, seguida por uma palavra que se presta a descrever resumidamente a variável. As *dummies* têm os rótulos iniciados pela letra “d”. As demais variáveis foram rotuladas com palavras que buscam descrevê-las de forma sucinta. A Tabela 03 apresenta as estatísticas descritivas para as variáveis ambientais usadas na fronteira da 4^a série. Os resultados das estimações dos parâmetros relativos à fronteira cuja variável dependente é o *slack* total referente à nota mediana de matemática da 4^a série de cada escola encontram-se na Tabela 04.

Tabela 03 - Estatísticas descritivas das variáveis ambientais contínuas utilizadas na fronteira relativa aos *slacks* das notas da 4ª série

Variável	Descrição	Nº de Obs.	Média	Desvio Padrão	Mínimo	Máximo
p4_participacao	Proporção dos alunos da 4ª série que participaram da Prova Brasil	8.155	0,84	0,14	0,10	1,00
p4_pais_sep	Proporção dos alunos da 4ª série cujos pais não vivem juntos	8.155	0,39	0,13	0,00	1,00
p4_reuniao_sempre	Proporção dos alunos da 4ª série cujos pais vão às reuniões sempre	8.155	0,62	0,13	0,00	1,00
p4_reuniao_alguma	Proporção dos alunos da 4ª série cujos pais vão às reuniões algumas vezes	8.155	0,30	0,11	0,00	1,00
p4_estante_livros	Proporção dos alunos da 4ª série que possuem o equivalente a uma estante de livros em casa (21 a 100 livros)	8.155	0,17	0,08	0,00	1,00
p4_trab_dom_ate_2h	Proporção dos alunos da 4ª série que realizam até 2h de trabalhos domésticos	8.155	0,80	0,08	0,00	1,00
p4_trab_fora	Proporção dos alunos da 4ª série que trabalham fora	8.155	0,16	0,09	0,00	0,91
p4_correcao_sempre	Proporção dos alunos da 4ª série que declararam que os professores sempre corrigem os deveres de matemática	8.155	0,81	0,12	0,00	1,00
p4_correcao_alguma	Proporção dos alunos da 4ª série que declararam que os professores corrigem os deveres de matemática de vez em quando	8.155	0,17	0,11	0,00	1,00
n_salas	Número de salas existentes em cada escola	8.155	13	6	2	134
tam_medio_turmas_4	Tamanho médio das turmas de 4ª série	8.155	29	7	2	54
p_abandono_1_4_séries	Proporção dos alunos de 1ª à 4ª série que abandonam a escola	8.155	0,04	0,05	0,00	0,64
p4_a	Proporção dos alunos da 4ª série pertencentes à classe A	8.155	0,01	0,03	0,00	1,00
p4_b	Proporção dos alunos da 4ª série pertencentes à classe B	8.155	0,21	0,17	0,00	1,00
p4_cd	Proporção dos alunos da 4ª série pertencentes à classe C e D	8.155	0,74	0,17	0,00	1,00
p1_4_alun_fem	Proporção dos alunos do sexo feminino nas turmas de 1ª à 4ª série	8.155	0,47	0,04	0,06	0,68
p1_4_alun_brancos	Proporção dos alunos da 1ª à 4ª série que se declararam brancos	8.155	0,15	0,20	0,00	0,95
p1_4_alun_pretos	Proporção dos alunos da 1ª à 4ª série que se declararam pretos	8.155	0,02	0,04	0,00	0,56
p1_4_alun_pardos	Proporção dos alunos da 1ª à 4ª série que se declararam pardos	8.155	0,20	0,19	0,00	0,99
p1_4_doc_14_fem	Proporção dos docentes de 1ª à 4ª série do sexo feminino	8.155	0,93	0,10	0,00	1,00
p1_4_doc_brancos	Proporção dos docentes da 1ª à 4ª série que se declaram brancos	8.155	0,26	0,32	0,00	1,00
p1_4_doc_pretos	Proporção dos docentes da 1ª à 4ª série que se declaram pretos	8.155	0,02	0,06	0,00	1,00
p1_4_doc_pardos	Proporção dos docentes da 1ª à 4ª série que se declaram pardos	8.155	0,15	0,20	0,00	1,00
p1_4_doc_nivel_sup	Proporção de docentes que possuem nível de educação superior nas turmas de 1ª à 4ª série	8.155	0,65	0,29	0,00	1,00

Fonte: Cálculos próprios com base nos dados disponíveis no site do INEP (INEP, 2011) e nos microdados do Censo Escolar e da Prova Brasil 2007 (INEP, 2007a, 2007b). Todas as proporções foram calculadas para cada escola da amostra, tomando como base apenas as respostas válidas.

Tabela 04 - Fronteira relativa aos *slacks* da nota mediana de matemática da 4ª Série

Variáveis	Coefficientes	Valores Coeficientes	Erro Padrão	Razão t	p-valor	Intervalo de Confiança 95%	
constante	β_0	111,83	7,20	15,54	0,00	97,73	126,00
p4_participacao	β_1	-4,50	1,35	-3,32	0,00	-7,15	-1,82
p4_pais_sep	β_2	8,75	1,26	6,94	0,00	6,28	11,24
p4_reunião_sempre	β_3	-13,00	2,78	-4,68	0,00	-18,44	-7,54
p4_reunião_alguna	β_4	-8,73	3,14	-2,78	0,01	-14,88	-2,62
p4_estante_livros	β_5	-9,97	2,15	-4,64	0,00	-14,19	-5,81
p4_trab_dom_ate_2h	β_6	-18,19	2,01	-9,07	0,00	-22,12	-14,40
p4_trab_fora	β_7	8,14	2,07	3,93	0,00	4,08	12,11
p4_correcao_sempre	β_8	-21,37	3,94	-5,42	0,00	-29,10	-13,63
p4_correcao_alguna	β_9	-14,23	4,43	-3,22	0,00	-22,91	-5,55
n_salas	β_{10}	0,61	0,11	5,59	0,00	0,40	0,83
(n_salas)2	β_{11}	-0,03	0,00	-7,68	0,00	-0,04	-0,02
(n_salas)3	β_{12}	0,00	0,00	7,74	0,00	0,00	0,00
tam_médio_turmas_4	β_{13}	1,71	0,48	3,57	0,00	0,77	2,66
(tam_médio_turmas_4)2	β_{14}	-0,06	0,02	-3,59	0,00	-0,09	-0,03
(tam_médio_turmas_4)3	β_{15}	0,00	0,00	3,08	0,00	0,00	0,00
p_abandono_1_4_séries	β_{16}	19,48	3,68	5,29	0,00	12,26	26,57
p4_a	β_{17}	-29,81	5,38	-5,54	0,00	-40,36	-19,37
p4_b	β_{18}	-9,70	2,35	-4,13	0,00	-14,30	-5,12
p4_cd	β_{19}	-0,73	2,26	-0,32	0,75	-5,16	3,64
p1_4_alun_fem	β_{20}	-6,60	3,71	-1,78	0,08	-13,86	0,52
p1_4_alun_branços	β_{21}	-15,45	1,27	-12,21	0,00	-17,93	-13,01
p1_4_alun_pretos	β_{22}	5,80	4,63	1,25	0,21	-3,27	14,78
p1_4_alun_pardos	β_{23}	6,99	1,11	6,28	0,00	4,81	9,21
p1_4_doc_fem	β_{24}	-3,45	1,64	-2,11	0,04	-6,66	-0,20
p1_4_doc_branços	β_{25}	-1,98	0,71	-2,80	0,01	-3,36	-0,60
p1_4_doc_pretos	β_{26}	4,83	2,60	1,86	0,06	-0,26	9,91
p1_4_doc_pardos	β_{27}	-1,57	0,88	-1,79	0,07	-3,29	0,13
p1_4_doc_nivel_sup	β_{28}	-3,37	0,59	-5,72	0,00	-4,52	-2,19
d_fed	β_{29}	-10,71	7,09	-1,51	0,13	-24,59	3,18
d_est	β_{30}	-1,59	0,37	-4,32	0,00	-2,31	-0,87
d_dir_ded_exc	β_{31}	-1,19	0,34	-3,51	0,00	-1,86	-0,52
d_dir_eleito	β_{32}	0,72	0,45	1,60	0,11	-0,16	1,59
d_dir_sel_eleito	β_{33}	-5,05	0,46	-11,06	0,00	-5,94	-4,17
info	β_{34}	0,41	0,27	1,50	0,13	-0,13	0,94
d_infra_otima	β_{35}	-3,39	0,47	-7,23	0,00	-4,31	-2,48
Outros Parâmetros da Fronteira							
$\ln(\sigma_v^2)$		4,98	0,03	159,09	0,00	4,92	5,04
$\ln(\sigma_u^2)$		4,20	0,08	52,67	0,00	4,04	4,35
σ_v		12,08	0,19			11,71	12,45
σ_u		8,15	0,32			7,54	8,81
$\sigma^2 = \sigma_v^2 + \sigma_u^2$		212,28	3,76			204,90	219,65
$\lambda = \sigma_u / \sigma_v$		0,67	0,48			-0,26	1,61

Demais Informações

Foram necessárias duas iterações para a convergência da fronteira.
 O termo u, que designa a ineficiência, tem distribuição exponencial.

Fonte: As estimativas acima foram realizadas no software STATA SE/11.2 (STATA CORP, 2009b) por meio do comando Frontier.

Passemos aos comentários acerca dos coeficientes. A primeira variável, proporção de alunos da 4ª série que participaram da Prova Brasil (p4_participação) foi incluída com o propósito de verificar se há indícios de algum comportamento estratégico por parte das escolas para melhorar seu desempenho na referida avaliação, tal como incentivar apenas os bons alunos a participar. Uma ou outra escola pode até ter adotado esse tipo de estratégia, mas, no agregado, esse comportamento não parece ter sido predominante, já que o coeficiente em questão, apesar de sua pequena monta, mostrou-se estatisticamente significativo e apresentou sinal negativo, sugerindo que quanto maior a participação menor a ineficiência. Se todos os alunos de dada escola participaram da Prova Brasil, a variável em questão atinge valor igual a 1 e isso representa uma diminuição na ineficiência em 4,50 pontos. Dizemos pontos porque o *slack* da nota de matemática da 4ª série está expresso nessa unidade.

Os coeficientes β_2 a β_7 , β_{17} a β_{19} e β_{21} a β_{23} , referentes às variáveis que descrevem o contexto socioeconômico dos alunos, são estatisticamente significativas, se não individualmente, pelo menos em conjunto, como veremos adiante, e apresentaram os sinais esperados. Na escolha dessas variáveis buscamos contemplar, além do que naturalmente entendemos como elementos definidores do NSE (e.g., variáveis *proxies* para renda e escolaridade dos pais), o conceito de capital social, destacado por Riordan (2004), que trata da relação dos adultos de uma família, sobretudo dos pais, com suas crianças. Tal conceito possui duas principais dimensões, uma estrutural e outra funcional. A primeira diz respeito à presença física dos membros da família e a segunda diz respeito à intensidade do relacionamento entre pais e filhos.

O ponto aqui é entender que não basta a existência de renda e de capital humano nas famílias, se não há presença e envolvimento dos pais nas atividades escolares de seus filhos. O coeficiente β_2 , que busca captar a estrutura familiar, é positivo, indicando que quanto maior a proporção de alunos na 4ª série cujos pais não vivem juntos (p4_pais_sep), maior é a ineficiência da escola. Os coeficientes relacionados às variáveis ligadas à frequência com que os pais vão às reuniões escolares (p4_reunião_sempre e p4_reunião_alguma), ao fato dos alunos da 4ª série realizarem até duas horas de trabalhos domésticos (p4_trab_dom_ate_2h) e ao fato de trabalharem fora (p4_trab_fora); buscaram capturar o comentado envolvimento dos pais com os filhos.

Quanto maior a proporção de pais que participam das reuniões escolares, menor a ineficiência da escola, com um efeito mais forte quando os pais vão sempre às reuniões. Da mesma forma ocorre quanto maior a proporção de alunos da 4ª série que realizam até duas horas de trabalho doméstico. Isso pode indicar que crianças que realizam alguns trabalhos domésticos estão bem integradas às suas famílias, pois participam da rotina de suas casas, além de estarem exercitando a disciplina, característica fundamental para o sucesso nos estudos. Já a proporção de alunos da 4ª série que declararam trabalhar fora, pode indicar não só a pobreza da família a que esses alunos pertencem, como o descaso dos pais em relação ao cotidiano de seus filhos, ou até mesmo, uma situação de exploração dos filhos pelos pais. Como esperado, tal variável aumenta a ineficiência da escola quanto maior a proporção observada.

O coeficiente da proporção de alunos que dizem possuir o equivalente a uma estante de livros em casa (*p4_estante_livros*), que pode ser considerada uma medida de capital humano, também mostrou efeito redutor sobre a ineficiência corroborando estudos anteriores (Delgado; Machado, 2007). As variáveis relativas à proporção de alunos que pertencem a cada classe social foram construídas com base no Critério de Classificação Econômica Brasil da Associação Brasileira de Empresas (ABEP, 2009), que atribui uma pontuação para cada tipo e quantidade de bens que uma família possui (geladeira, carro, televisor, etc.) e também para o grau de escolaridade de seu chefe. O critério da ABEP define 8 classes diferentes: A1, A2, B1, B2, C1, C2, D e E. Contudo, para facilitar a exposição, agregamos as duas primeiras classes sob o rótulo de A, as duas seguintes, da mesma maneira, delimitando a classe B e as três, seguintes sob a denominação CD. Os resultados dessas variáveis se mostraram como o esperado, quanto maior a proporção de alunos em classes mais altas, menor a ineficiência da escola. É interessante notar a grande magnitude do coeficiente relativo ao pertencimento à classe mais alta e a queda de seu valor à medida que passamos às classes mais baixas. Sobre a escolaridade do chefe de família, optamos por utilizar o grau de escolaridade da mãe, pelo fato de haver um volume maior de respostas nesse item em comparação com a mesma resposta dada em relação ao pai.

As características raciais dos alunos, como são autodeclaradas, podem refletir outros aspectos tais como nível socioeconômico e até mesmo autoestima dos alunos, por isso são de difícil interpretação.

A inclusão dessas características, amplamente utilizadas na estimação das funções de produção educacional, facilita a comparação do nosso trabalho com outros estudos. Nossos resultados sugerem que escolas cuja proporção de alunos que se autodeclaram brancos é maior, a ineficiência é menor.

As variáveis relativas ao gênero dos alunos e professores (e às suas características) e à taxa de abandono são proporções que consideram alunos da 1ª a 4ª série. A utilização dessa agregação reflete melhor o contexto da escola. Acerca do gênero dos alunos, esperava-se resultado contrário ao verificado, já que quando se relaciona desempenho em matemática com gênero, geralmente os meninos se saem melhor (Menezes-Filho; Ribeiro, 2009). Em relação às características dos professores, no que se refere à raça, aplica-se a mesma discussão feita para os alunos no que tange à dificuldade em se entender o que exatamente a autodeclaração da raça expressa.

Com respeito à qualificação dos professores e às suas atitudes (frequência da correção das tarefas de matemática), os sinais dos efeitos sobre a ineficiência se mostraram como esperado: quanto mais professores qualificados e que corrigem as tarefas de casa, menor a ineficiência. Sobre o gênero dos docentes, não havia expectativa a respeito do efeito dessa variável. Em nossas estimativas observamos um pequeno decréscimo na ineficiência para escolas com maior proporção de professoras.

No que se refere ao tamanho das salas e das escolas, as evidências empíricas não são conclusivas e há argumentos tanto pró como contra escolas e salas pequenas. Riordan (2004) expõe argumentos e resultados de diversos trabalhos para os dois lados dessa disputa. A favor das escolas menores o autor elenca as razões apontadas para o sucesso dessas escolas, tais como, possuir maior controle sobre o comportamento dos alunos, evitando abandono e comportamentos desviantes em seu corpo discente; além do que, tais escolas são capazes de oferecer um atendimento mais personalizado, dentre outras.

Por outro lado, há argumentos no sentido contrário que afirmam que, por uma questão de economia de escala, as escolas maiores são capazes de oferecer uma melhor infraestrutura e uma maior diversidade de atividades.

¹⁴ See Dixit and Pindyck (1994), p. 142.

A evidência mais forte em favor da existência de um tamanho ótimo de escola é verificada quando relacionamos tamanho das escolas com eficiência econômica, já que, em muitos casos, constatou-se a existência de uma curva de custo médio por aluno em formato de “U”, o que significa que esses custos decrescem à medida que a escola admite mais alunos, porém, isso só ocorre até certo ponto, quando o custo médio por aluno atinge seu valor mínimo, a partir daí a admissão de mais alunos implicará custos médios crescentes (Slate; Jones, 2005). Para testar essa tese, incluímos a variável número de salas (n_{salas}) em uma especificação polinomial até o terceiro grau. Daí, obtivemos que a ineficiência é decrescente para escolas que possuem entre 13 e 88 salas, o que equivale, em termos de número de alunos, ao intervalo entre 377 e 2.552 alunos, considerando uma média geral de 29 alunos por turma. Das 8.155 escolas em estudo, 3.452 escolas se encontram nessa faixa.

Raciocínio análogo àquele feito para o tamanho das escolas foi tecido para o tamanho das turmas. Especificamente para as turmas de ensino regular da 4ª série, encontramos que escolas com turmas de tamanho médio entre 20 e 47 alunos estão no intervalo de ineficiência cadente. Grande parte das escolas se encontra nesse intervalo (7.540). Os argumentos pró e contra salas pequenas são bem parecidos àqueles empregados para justificar o tamanho das escolas. Ademais, os custos de se reduzir o tamanho das salas são muitos altos diante da fraca evidência de que tal redução será benéfica na maioria dos casos (Riordan, 2004; Hanushek, 1998). Hanushek (1998) argumenta ainda haver evidências econométricas de que o impacto positivo de professores qualificados pode mais que compensar a pequena melhoria no aproveitamento dos alunos em turmas reduzidas, quando tal efeito é detectado.

Sobre a dependência administrativa das escolas, como já esperado, as escolas federais mostraram-se muito mais eficientes que as demais. E as escolas estaduais mostraram-se pouco mais eficientes que as municipais. Já em termos das características dos diretores, a literatura empírica sugere que o papel do diretor é fundamental para o desempenho da escola (Witziers *et al.*, 2003). Em nosso estudo, testamos várias características dos diretores; a única significativa foi a que se refere à forma de seleção. Escolas que apresentam diretores com dedicação exclusiva e que foram selecionados e eleitos, apresentam menor ineficiência, com destaque, em termos de magnitude,

para essa última variável. Esse resultado não é surpreendente; diretores bem preparados e cuja liderança na escola foi chancelada por meio de eleição terão melhores condições de conduzi-la para obter melhores resultados.

Uma maior proporção de alunos da 1^a à 4^a série que abandonam a escola aumenta a ineficiência das escolas, o que era esperado. Hanushek, Lavy e Hitomi (2006), em estudo feito com dados do Egito, também encontraram resultados nesse sentido, de que as taxas de abandono são bastante sensíveis à qualidade das escolas.

As variáveis *d_infra_otima* e *info* tentam captar o impacto da infraestrutura, tanto básica como relativa aos equipamentos que as escolas possuem, sobre seus níveis de eficiência. A primeira foi escolhida porque resume várias características físicas da escola. Por ocasião da Prova Brasil, pesquisadores preenchem um questionário que cobre desde a existência de estruturas e equipamentos, seu estado de conservação, passando por questões de segurança. Com base nessas respostas, o pesquisador responde à questão “Como você avalia esta escola, considerando sua infraestrutura e condições gerais de funcionamento?” marcando uma, entre cinco alternativas que variam de péssima a ótima. Nossa variável se refere à última alternativa. Seu coeficiente, como esperado, apresentou sinal negativo corroborando estudos anteriores que mostram que características ligadas à infraestrutura básica da escola, tais como qualidade dos telhados, paredes, piso, com mesas e cadeiras, além da existência de biblioteca, conduzem ao melhor rendimento dos alunos (Glewwe *et al.*, 2012). A variável *info*, que se refere ao fato da escola ter laboratório de informática e/ou outros computadores, não parece afetar a eficiência das escolas. Esse resultado é compatível com estudos anteriores que atestam que o efeito desse tipo de equipamento sobre os resultados das escolas não é claro. 18 das 26 estimativas analisadas por Glewwe *et al.* (2012) não apresentaram significância estatística e, em termos de sinais, são distribuídas quase que igualmente. O fato de a variável *info* não influenciar os resultados educacionais deve-se, provavelmente, ao fato de que, em muitos casos, esses laboratórios não passam de uma sala com finalidade múltipla em que os computadores são guardados. Embora declarados como laboratório, eles não desenvolvem, de forma sistemática, atividades que efetivamente permitam aos alunos um melhor aprendizado na área de informática.

Raciocínio similar aplica-se às bibliotecas, fazendo com que o poder discriminatório dessa variável seja praticamente nulo.

A Tabela 05 apresenta os testes estatísticos relativos à especificação da fronteira.

Tabela 05 - Testes Razão de Verossimilhança relativos à fronteira relativa aos *slacks* da mediana das notas de matemática da 4ª Série

H_0	$\ln LR_{H_0}$	$\ln LR_{H_A}$	Graus de Liberdade	Estatística de Teste $-2*(\ln LR_{H_0} - \ln LR_{H_A})$	p-valor
$\sigma_u=0$	-33.419,41	-33.338,43	1	161,95	0,00
u segue distribuição seminormal	-33.353,86	-33.338,43	1	30,85	0,00
$\beta_1 = \dots = \beta_{35}=0$	-34.566,06	-33.338,43	35	2.455,25	0,00
$\beta_{17} = \beta_{18} = \beta_{19} = 0$	-33.391,50	-33.338,43	3	106,14	0,00
$\beta_{20} = \beta_{21} = \beta_{22} = \beta_{23} = 0$	-33.418,44	-33.338,43	4	160,01	0,00
$\beta_{24} = \dots = \beta_{28} = 0$	-33.364,50	-33.338,43	5	52,13	0,00
$\beta_{29} = \beta_{30} = 0$	-33.348,58	-33.338,43	2	20,29	0,00
$\beta_{31} = \beta_{32} = \beta_{33} = 0$	-33.421,44	-33.338,43	3	166,01	0,00
$\beta_{34} = \beta_{35} = 0$	-33.364,65	-33.338,43	2	52,44	0,00

Obs: Cálculos realizados com base nos resultados gerados pelo STATA/SE 11.2 (STATACORP, 2009b). Convém comentar que o primeiro teste desta tabela obedece à uma distribuição Qui-Quadrado modificada, que considera o fato do parâmetro testado poder assumir apenas valores maiores que zero. Tal modificação na estatística de teste tem como consequência a redução do p-valor à metade do que seria usualmente calculado, conforme (GUTIERREZ; CARTER; DRUKKER, 2001).

O primeiro teste constante dessa tabela confirma a significância estatística da variância ligada ao termo ineficiência, u , apontando para o bom ajuste dos dados a essa forma de fronteira. O segundo teste prestou-se à escolha entre duas hipóteses de distribuições possíveis para a variável u , conduzindo-nos à rejeição da distribuição seminormal em favor da exponencial. Os demais testes permitem verificar a significância conjunta dos parâmetros. Conforme podemos ver, todos os parâmetros testados foram validados.

Passaremos agora à análise dos resultados da fronteira relativa aos *slacks* da nota mediana de matemática da 8ª série (Tabelas 06 a 08). A Tabela 06 apresenta as estatísticas descritivas para as variáveis utilizadas na fronteira, a Tabela 07, os coeficientes estimados e a 08, os testes de Razão de Verossimilhança relativos à fronteira em questão.

Tabela 06 - Estatísticas descritivas das variáveis ambientais contínuas utilizadas na fronteira relativa aos *slacks* das notas da 8ª série

Variável	Descrição	Nº de Obs.	Média	Desvio Padrão	Mínimo	Máximo
p8_participacao	Proporção dos alunos da 8ª série que participaram da Prova Brasil	8.155	0,77	0,15	0,08	1,00
p8_pais_sep	Proporção dos alunos da 8ª série cujos pais não vivem juntos	8.155	0,38	0,13	0,00	1,00
p8_reuniao_sempre	Proporção dos alunos da 8ª série cujos pais vão às reuniões sempre	8.155	0,59	0,15	0,00	1,00
p8_reuniao_alguma	Proporção dos alunos da 8ª série cujos pais vão às reuniões algumas vezes	8.155	0,33	0,12	0,00	0,90
p8_estante_livros	Proporção dos alunos da 8ª série que possuem o equivalente a uma estante de livros em casa (21 a 100 livros)	8.155	0,19	0,09	0,00	0,69
p8_trab_dom_ate_2h	Proporção dos alunos da 8ª série que realizam até 2h de trabalhos domésticos	8.155	0,77	0,10	0,14	1,00
p8_trab_fora	Proporção dos alunos da 8ª série que trabalham fora	8.155	0,22	0,13	0,00	0,94
p8_correcao_sempre	Proporção dos alunos da 8ª série que declaram que os professores sempre corrigem os deveres de matemática	8.155	0,85	0,12	0,17	1,00
p8_correcao_alguma	Proporção dos alunos da 8ª série que declaram que os professores corrigem os deveres de matemática de vez em quando	8.155	0,12	0,10	0,00	0,83
n_salas	Número de salas existentes em cada escola	8.155	13	6	2	134
tam_medio_turmas_8	Tamanho médio das turmas de 8ª série	8.155	30	8	2	54
p_abandono_5_8_séries	Proporção dos alunos de 5ª à 8ª série que abandonam a escola	8.155	0,08	0,09	0,00	0,70
p8_a	Proporção dos alunos da 8ª série pertencentes à classe A	8.155	0,01	0,02	0,00	0,50
p8_b	Proporção dos alunos da 8ª série pertencentes à classe B	8.155	0,16	0,14	0,00	0,84
p8_cd	Proporção dos alunos da 8ª série pertencentes à classe C e D	8.155	0,79	0,14	0,00	1,00
p5_8_alun_fem	Proporção dos alunos do sexo feminino nas turmas de 5ª à 8ª série	8.155	0,51	0,04	0,15	0,76
p5_8_alun_brancos	Proporção dos alunos da 5ª à 8ª série que se declararam brancos	8.155	0,13	0,19	0,00	0,96
p5_8_alun_pretos	Proporção dos alunos da 5ª à 8ª série que se declararam pretos	8.155	0,02	0,04	0,00	0,57
p5_8_alun_pardos	Proporção dos alunos da 5ª à 8ª série que se declararam pardos	8.155	0,17	0,19	0,00	0,99
p5_8_doc_fem	Proporção dos docentes de 5ª à 8ª série do sexo feminino	8.155	0,74	0,15	0,00	1,00
p5_8_doc_brancos	Proporção dos docentes da 5ª à 8ª série que se declaram brancos	8.155	0,29	0,31	0,00	1,00
p5_8_doc_pretos	Proporção dos docentes da 5ª à 8ª série que se declaram pretos	8.155	0,02	0,05	0,00	1,00
p5_8_doc_pardos	Proporção dos docentes da 5ª à 8ª série que se declaram pardos	8.155	0,16	0,18	0,00	1,00
p5_8_doc_nivel_sup	Proporção de docentes que possuem nível de educação superior nas turmas de 5ª à 8ª série	8.155	0,87	0,20	0,00	1,00

Fonte: Cálculos próprios com base nos dados disponíveis no site do INEP (INEP, 2011) e nos microdados do Censo Escolar e da Prova Brasil 2007 (INEP, 2007a, 2007b). Todas as proporções foram calculadas para cada escola da amostra, tomando como base apenas as respostas válidas.

Tabela 07 - Fronteira relativa aos *slacks* da nota mediana de matemática da 8ª Série

Variáveis	Coefficientes	Valores	Coefficientes	Erro Padrão	Razão t	p-valor	Intervalo de Confiança 95%	
constante	β_0	122,46		8,64	14,18	0,00	105,53	139,39
p8_participacao	β_1	-4,94		1,60	-3,09	0,00	-8,07	-1,81
p8_pais_sep	β_2	21,26		1,75	12,13	0,00	17,82	24,70
p8_reuniao_sempre	β_3	-3,16		3,23	-0,98	0,33	-9,50	3,17
p8_reuniao_alguma	β_4	2,98		3,79	0,78	0,43	-4,46	10,41
p8_estante_livros	β_5	-5,11		2,52	-2,03	0,04	-10,05	-0,17
p8_trab_dom_ate_2h	β_6	-2,30		2,31	-1,00	0,32	-6,82	2,22
p8_trab_fora	β_7	-4,06		1,87	-2,17	0,03	-7,73	-0,40
p8_correcao_sempre	β_8	-14,74		4,68	-3,15	0,00	-23,92	-5,56
p8_correcao_alguma	β_9	-6,67		5,92	-1,13	0,26	-18,28	4,94
n_salas	β_{10}	0,95		0,14	6,98	0,00	0,68	1,22
(n_salas)2	β_{11}	-0,04		0,00	-9,04	0,00	-0,05	-0,03
(n_salas)3	β_{12}	0,00		0,00	9,01	0,00	0,00	0,00
tam_medio_turmas_8	β_{13}	1,91		0,49	3,87	0,00	0,94	2,88
(tam_medio_turmas_8)2	β_{14}	-0,07		0,02	-4,03	0,00	-0,10	-0,04
(tam_medio_turmas_8)3	β_{15}	0,00		0,00	3,49	0,00	0,00	0,00
p_abandono_5_8_séries	β_{16}	6,28		3,20	1,96	0,05	0,01	12,55
p8_a	β_{17}	-96,36		11,61	-8,30	0,00	-119,13	-73,60
p8_b	β_{18}	-49,88		3,45	-14,46	0,00	-56,64	-43,12
p8_cd	β_{19}	-16,80		3,10	-5,42	0,00	-22,87	-10,73
p5_8_alun_fem	β_{20}	-11,08		4,86	-2,28	0,02	-20,61	-1,55
p5_8_alun_branços	β_{21}	-16,05		1,65	-9,74	0,00	-19,28	-12,82
p5_8_alun_pretos	β_{22}	14,13		5,96	2,37	0,02	2,44	25,82
p5_8_alun_pardos	β_{23}	4,64		1,41	3,30	0,00	1,88	7,39
p5_8_doc_fem	β_{24}	-4,44		1,41	-3,15	0,00	-7,20	-1,68
p5_8_doc_branços	β_{25}	-3,75		0,93	-4,02	0,00	-5,58	-1,92
p5_8_doc_pretos	β_{26}	4,36		4,19	1,04	0,30	-3,84	12,57
p5_8_doc_pardos	β_{27}	-0,36		1,24	-0,29	0,77	-2,80	2,08
p5_8_doc_nivel_sup	β_{28}	-5,83		1,14	-5,10	0,00	-8,07	-3,59
d_fed	β_{29}	-4,08		9,23	-0,44	0,66	-22,16	14,01
d_est	β_{30}	-0,25		0,47	-0,53	0,60	-1,17	0,67
d_dir_ded_exc	β_{31}	-1,58		0,43	-3,70	0,00	-2,41	-0,74
d_dir_eleito	β_{32}	-0,55		0,56	-0,98	0,33	-1,66	0,56
d_dir_sel_eleito	β_{33}	-7,83		0,57	-13,83	0,00	-8,94	-6,72
info	β_{34}	0,34		0,34	0,97	0,33	-0,34	1,01
d_infra_otima	β_{35}	-3,68		0,59	-6,28	0,00	-4,82	-2,53
Outros Parâmetros da Fronteira								
$\ln(\sigma_v^2)$		5,34		0,03	159,21	0,00	5,27	5,40
$\ln(\sigma_u^2)$		4,92		0,07	73,66	0,00	4,79	5,05
σ_v		14,41		0,24			13,94	14,89
σ_u		11,70		0,39			10,96	12,49
$\sigma^2 = \sigma_v^2 + \sigma_u^2$		344,46		6,50			331,72	357,20
$\lambda = \sigma_u / \sigma_v$		0,81		0,59			-0,34	1,96
Demais Informações								

Foram necessárias três iterações para a convergência da fronteira.

O termo u , que designa a ineficiência, tem distribuição exponencial.

Fonte: As estimativas acima foram realizadas no software STATA SE/11.2 (STATA CORP, 2009b) por meio do comando Frontier.

Nessa estimativa, a maioria dos coeficientes manteve o sinal da primeira fronteira. Os que não seguiram esse padrão foram os relativos às seguintes variáveis: proporção de pais de alunos da 8ª série que

frequentam algumas reuniões (p8_reunião_alguma), ao percentual de alunos dessa mesma série que trabalham fora (p8_trab_fora) e ao fato dos diretores serem eleitos apenas (d_dir_eleito).

A primeira inversão pode ser lida como a perda da importância desse aspecto, quando falamos de crianças mais velhas. Especialmente se a referida inversão for interpretada levando-se em conta tanto a redução da magnitude do coeficiente relativo à outra variável que trata da frequência dos pais às reuniões escolares (p8_reunião_sempre), como a perda de significância estatística de cada um desses coeficientes. A segunda inversão pode indicar que adolescentes que trabalham fora não têm rendimento prejudicado por isso, uma vez que os que chegaram até a 8ª série são os mais persistentes, já acumularam conhecimento e que, por todos esses motivos, conseguem ir bem na escola, apesar de trabalharem fora.

Outras variáveis que merecem destaque são as que tiveram seus coeficientes muito aumentados ou diminuídos em termos absolutos, mantidos os sinais. Na primeira categoria estão as variáveis relativas ao fato dos pais morarem separados (p8_pais_sep), às classes sociais a que os alunos pertencem (p8_a, p8_b, p8_cd), aos alunos da 5ª à 8ª série que se declaram negros (p5_8_alun_pretos), aos docentes que se declaram brancos (p5_8_doc_branco) e aos que possuem nível superior (p5_8_doc_nivel_sup), além daquela variável referente ao fato dos diretores serem selecionados e eleitos (d_dir_sel_eleito). Esperava-se que esse grupo de variáveis, sobretudo àquelas relativas aos alunos, exercesse efeitos de menor intensidade que aquela verificada no caso dos alunos mais novos, já que, intuitivamente, crianças mais velhas, ou adolescentes, seriam menos suscetíveis ao meio que vivem. Diante disso, não sabemos se a intensificação dos impactos de algumas variáveis se deve à existência de um efeito cumulativo do meio sobre a criança ou decorre da melhor qualidade das respostas dadas nos questionários por crianças maiores ou adolescentes.

Na categoria das variáveis que tiveram seus efeitos reduzidos em termos absolutos, mantidos os sinais, destacamos as proporções de: alunos que realizam até duas horas de trabalhos domésticos (p8_trab_dom_ate_2h), alunos que afirmam que os professores corrigem sempre as tarefas de matemática (p8_correcao_sempre), alunos que afirmam que os professores corrigem algumas vezes as tarefas de matemática (p8_correcao_alguma), alunos da 5ª à 8ª série que aban-

donaram a escola (p_abandono_5_8_séries), alunos da 5ª à 8ª série que se declaram pardos (p5_8_alun_pardos), docentes da 5ª à 8ª série que se declaram pardos (p5_8_doc_pardos). Sobre os tamanhos tanto de salas como das escolas, encontramos que: para salas entre 20 e 50 alunos e para escolas que tenham entre 14 e 88 salas, a ineficiência é decrescente, resultados bem parecidos com aqueles encontrados na primeira fronteira. Finalmente, a Tabela 08 apresenta os testes estatísticos relativos à especificação da fronteira.

Tabela 08 - Testes Razão de Verossimilhança relativos à fronteira relativa ao *slack* da mediana das notas de matemática da 8ª Série

H ₀	lnLR _{H0}	lnLR _{HA}	Graus de Liberdade	Estatística de Teste -2*(lnLR _{H0} -lnLR _{HA})	p-valor
$\sigma_{\epsilon_i}=0$	-35.367,59	-35.234,67	1	265,83	0,00
u segue distribuição seminormal	-35.245,51	-35.234,67	1	21,67	0,00
$\beta_1 = \dots = \beta_{35} = 0$	-36.647,51	-35.234,67	35	2.825,67	0,00
$\beta_3 = \dots = \beta_7 = 0$	-35.245,78	-35.234,67	5	22,22	0,00
$\beta_8 = \beta_9 = 0$	-35.252,50	-35.234,67	2	35,66	0,00
$\beta_{24} = \dots = \beta_{28} = 0$	-35.263,76	-35.234,67	5	58,18	0,00
$\beta_{29} = \dots = \beta_{33} = 0$	-35.362,62	-35.234,67	5	255,90	0,00
$\beta_{34} = \beta_{35} = 0$	-35.254,14	-35.234,67	2	38,93	0,00

Obs: Cálculos realizados com base nos resultados gerados pelo STATA/SE 11.2 (STACORP, 2009b). Convém comentar que o primeiro teste desta tabela obedece à uma distribuição Qui-Quadrado modificada, que considera o fato do parâmetro testado poder assumir apenas valores maiores que zero. Tal modificação na estatística de teste tem como consequência a redução do p-valor à metade do que seria usualmente calculado, conforme (GUTIERREZ; CARTER; DRUKKER, 2001).

Trataremos agora da fronteira relativa aos *slacks* do número de alunos. As Tabelas 09 e 10 apresentam as estatísticas descritivas para as variáveis utilizadas na fronteira relativa aos *slacks* referentes ao número de alunos. E a Tabela 11 mostra os coeficientes estimados para essa fronteira.

Nessa fronteira procuramos inserir variáveis que pudessem explicar porque as escolas não atendem à quantidade de alunos que poderiam atender. Levamos em conta os seguintes aspectos que podem influenciar tanto a demanda como a oferta por vagas: o tamanho do município em que a escola se encontra, se tais municípios atraem população, qual a composição da população em termos de faixa etária,

qual o tamanho da escola, se ela é de qualidade, bem administrada, dotada de boa infraestrutura e equipamentos e se oferece segurança aos alunos.

Tabela 09 - Estatísticas descritivas das variáveis ambientais contínuas utilizadas na fronteira relativa ao *slack* de alunos

Variável	Descrição	Nº de Obs.	Média	Desvio Padrão	Mínimo	Máximo
p60_mais	Proporção da população de cada município com idade superior a 60 anos	8.155	0,11	0,02	0,03	0,25
n_salas	Número de salas existentes na escola	8.155	13	6	2	134
p_abandono_fund	Proporção dos alunos de cada escola que abandonaram o ensino fundamental	8.155	0,06	0,07	0,00	0,47
n_turmas	Número de turmas em cada escola	8.155	29	13	7	136

Fonte: Cálculos próprios com base nos dados do Censo Demográfico 2010 (IBGE, 2011b), para a primeira variável; do Censo Escolar 2007 (INEP, 2007a), para a segunda e a quarta; e disponíveis no site do INEP (INEP, 2011), para a terceira.

Tabela 10 - Estatísticas descritivas das variáveis ambientais discretas utilizadas exclusivamente na fronteira relativa ao *slack* de alunos

Variáveis	Descrição	Distribuição das observações de acordo com os valores assumidos pelas variáveis		
		Total	0	1
d_0_20mil	<i>Dummy=1</i> se a escola está em um município com até 20 mil habitantes	8.155	6.276	1.879
d_rm	<i>Dummy=1</i> se a escola está em um município integrante de alguma região metropolitana	8.155	5.073	3.082
d_depredação	<i>Dummy=1</i> se a escola apresenta sinais de depredação (vidros, portas e janelas quebrados, lâmpadas estouradas etc.)	8.155	5.432	2.723
d_med_segurança	<i>Dummy=1</i> se a escola adota medidas de segurança para proteger os alunos nas suas imediações	8.155	5.411	2.744

Fonte: Cálculos próprios com base nos dados do Censo Demográfico 2010 (IBGE, 2011b), para a primeira variável; do site do IBGE (IBGE, 2011a), para a segunda, e nos microdados da Prova Brasil 2007 (INEP, 2007b), para as duas últimas.

Tabela 11 - Fronteira relativa aos *slacks* referentes ao número de alunos

Variáveis	Coefficientes	Valores Coeficientes	Erro Padrão	Razão t	p-valor	Intervalo de Confiança 95%	
constante	β_0	137,38	9,05	15,17	0,00	119,64	155,13
d_0_20mil	β_1	9,96	3,28	3,04	0,00	3,53	16,39
d_rm	β_2	45,20	2,84	15,92	0,00	39,64	50,76
p60_mais	β_3	-558,90	55,09	-10,15	0,00	-666,87	-450,94
n_salas	β_4	22,73	0,82	27,56	0,00	21,11	24,35
(n_salas)2	β_5	-0,48	0,03	-17,66	0,00	-0,53	-0,43
(n_salas)3	β_6	0,00	0,00	14,25	0,00	0,00	0,00
p_abandono_fund	β_7	519,54	19,43	26,74	0,00	481,47	557,62
d_fed	β_8	-237,82	48,41	-4,91	0,00	-332,70	-142,95
d_est	β_9	-6,65	2,72	-2,44	0,02	-11,99	-1,32
d_dir_ded_exc	β_{10}	-17,75	2,61	-6,81	0,00	-22,86	-12,64
d_dir_eleito	β_{11}	-3,60	3,39	-1,06	0,29	-10,25	3,04
d_dir_sel_eleito	β_{12}	-33,28	3,35	-9,92	0,00	-39,85	-26,70
info	β_{13}	8,98	2,08	4,33	0,00	4,92	13,05
d_infra_otima	β_{14}	-25,00	3,64	-6,87	0,00	-32,12	-17,87
d_depredação	β_{15}	12,70	2,62	4,84	0,00	7,55	17,84
d_med_segurança	β_{16}	-12,17	2,62	-4,64	0,00	-17,30	-7,03
Outros Parâmetros da Fronteira							
$\ln(\sigma_v^2)$							
constante		9,32	0,02	570,60	0,00	9,28	9,35
$\ln(\sigma_u^2)$							
n_turmas		-0,15	0,00	-60,11	0,00	-0,16	-0,15
constante		11,13	0,05	241,66	0,00	11,04	11,22
σ_v		105,43	0,86			103,76	107,13
Demais Informações							

Foram necessárias 241 iterações para a convergência da fronteira.

O termo u , que designa a ineficiência, tem distribuição seminormal.

Fonte: As estimativas acima foram realizadas no software STATA SE/11.2 (STATA CORP, 2009b) por meio do comando Frontier.

As duas primeiras *dummies* constantes da fronteira acima: se escola está em um município com até vinte mil habitantes (d_0_20mil), (IBGE, 2011b) e se está em município considerado região metropolitana (d_rm), (IBGE, 2011a), nos mostram que há ineficiências tanto em cidades grandes como pequenas, proporcionais ao tamanho das cidades.

Escolas localizadas em municípios com maior parcela de idosos, pessoas com mais de sessenta anos ($p60_mais$) (IBGE, 2011b), mostraram ser menos ineficientes. Provavelmente pelo fato de não terem muita demanda para seus serviços. Em relação ao tamanho das escolas (n_salas e seus termos ao quadrado e ao cubo), os resultados obtidos sugerem que as escolas maiores são menos ineficientes. As escolas que possuam entre 32 e 95 salas, ou, entre 928 e 2.755 alunos, considerando a já citada média de 29 alunos por sala, são as que estão no intervalo de ineficiência cadente. Apenas 37 escolas na amostra se enquadram nessa faixa.

Convém considerar que a diferença entre o tamanho ideal de escola verificada entre as duas primeiras fronteiras e essa é justificável, uma vez que as duas primeiras referiam-se às ineficiências ligadas ao rendimento dos alunos, enquanto a última se refere à ineficiência ligada à realidade física das escolas. Acerca da proporção de alunos da EF que abandonam a escola, como esperado, maiores proporções de abandono indicam escolas com maior ineficiência, com destaque para o tamanho do coeficiente. Em termos da dependência administrativa a que a escola pertence e às características do diretor, os resultados são bem similares aos obtidos nas duas primeiras fronteiras. O que se repete para as variáveis d_infra_otima e $info$. As duas últimas variáveis que tratam do estado de conservação da escola (se há sinais de depredação) e da segurança (se os diretores adotam medidas de segurança nas imediações da escola) apresentaram também os efeitos esperados. Escolas com sinais de depredação atraem menos alunos e, portanto, deixam de atingir sua capacidade máxima de atendimento. Por outro lado, escolas que oferecem maiores condições de segurança, atraem mais alunos e por isso reduzem sua ineficiência. Como nessa fronteira estamos lidando com escolas de tamanhos muito diferentes, levamos tal heterogeneidade em conta incluindo a variável número de turmas por escola (n_turmas), que é uma medida do tamanho da escola, como variável explicativa da variância relativa à parte do erro ligada às ineficiências. Como podemos ver na Tabela 12, esse parâmetro, assim como os outros relativos às variâncias, tanto do termo u , como do termo v , são significativos, o que mostra o bom ajuste dos dados ao formato da fronteira. Sobre a hipótese distribucional da parte do erro ligada à ineficiência, não tivemos que escolher, uma vez que a distribuição seminormal foi a única que se ajustou aos dados.

Seguindo o padrão de apresentação das fronteiras, a Tabela 12 apresenta os testes de exclusão das variáveis explicativas, assim como foi feito para as estimações anteriores.

Tabela 12 - Testes Razão de Verossimilhança para a fronteira relativa aos *slacks* referentes ao número de alunos

H ₀	lnLR _{H0}	lnLR _{HA}	Graus de Liberdade	Estatística de Teste -2*(lnLR _{H0} -lnLR _{HA})	p-valor
$\beta_1 = \dots = \beta_{10} = 0$	-51.066,09	-49.745,35	16	2.641,48	0,00
$\beta_{10} = \dots = \beta_{12} = 0$	-49.824,80	-49.745,35	3	158,89	0,00

Obs: Cálculo realizados com base nos resultados gerados pelo STATA/SE 11.2 (STACACORP, 2009b).

5.2 Resultados do 3^o Estágio

Com base na estimação das fronteiras descritas na seção anterior, ajustamos os dados referentes às variáveis de resultado de acordo com (Tone; Tsuitsui, 2009). As variáveis ajustadas foram utilizadas junto com as variáveis de insumo originais em um novo cálculo do algoritmo DEA – BCC, gerando um novo conjunto de escores, que apresentam distribuição, conforme Gráfico 02.

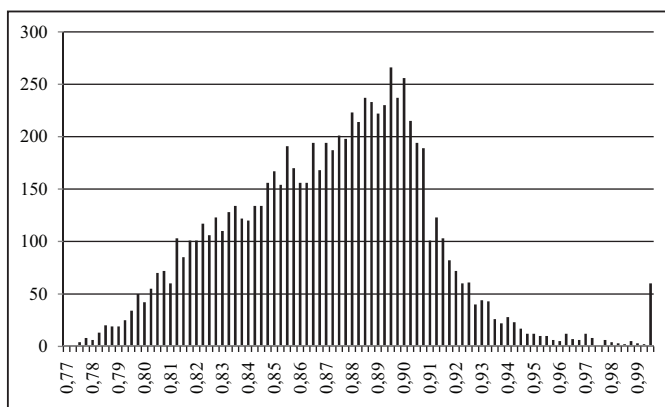


Gráfico 2 - Histograma relativo aos escores gerados no 3^o Estágio

Observa-se que a distribuição dos escores está em torno de um patamar mais alto que aquele verificado no primeiro estágio, sugerindo que os fatores ambientais e aqueles residuais, tratados como aleatórios, têm influência significativa sobre a eficiência de cada escola. Assim como fizemos para os escores do primeiro estágio, é interessante ressaltar que a média da medida de Farrell equivalente ao escore médio apresentado acima é 1,15, indicando que as escolas devem aumentar seus resultados, mantendo-se a proporção entre eles, em média em 15% para lograrem eficiência máxima. É a escola que está em pior situação, tem medida de Farrell de 1,30, com interpretação análoga. Podemos observar ainda a evolução dos escores de eficiência por tamanho de escola, conforme a Tabela 13. Verificamos que os incrementos relativos de eficiência entre o 1º e o último estágio são crescentes no tamanho das escolas até 1.217 alunos.

Tabela 13 - Evolução dos escores médios de eficiência por tamanho das escolas

Intervalos em decis	Intervalos em número de alunos	Número de Escolas no intervalo	Média Eficiência 1º Estágio	Média Eficiência 3º Estágio	Varição absoluta	Varição relativa
Até o 1º	Até 362	818	0,77	0,84	0,07	8,60%
Entre o 1º e o 2º	362 - 472	813	0,75	0,84	0,09	12,57%
Entre o 2º e o 3º	472 - 565	818	0,74	0,85	0,11	14,69%
Entre o 3º e o 4º	565 - 666	822	0,75	0,86	0,11	14,68%
Entre o 4º e o 5º	666 - 779	808	0,75	0,86	0,12	15,53%
Entre o 5º e o 6º	779 - 900	814	0,76	0,87	0,12	15,47%
Entre o 6º e o 7º	900 - 1.040	821	0,76	0,88	0,12	16,07%
Entre o 7º e o 8º	1.040 - 1.217	811	0,76	0,89	0,12	16,37%
Entre o 8º e o 9º	1.217 - 1.521	815	0,77	0,89	0,12	15,91%
Entre o 9º e o 10º	1.521 - 4.092	815	0,80	0,91	0,11	13,75%

Obs: Intervalos fechados à direita. As médias dos escores de eficiência e suas respectivas variações foram calculadas por intervalo.

Com base nos resultados acima, vê-se que os novos escores são, em média, maiores que aqueles obtidos no primeiro estágio. Agora vejamos as alterações ocorridas em termos de ordenação. Calculamos, para tanto, os coeficientes de correlação de Kendall (1938) e de Spearman (1904), além dos testes t para pares e Wilcoxon para dados ordenados. Para os dois primeiros, encontramos os valores estatisticamente significativos de 26,43%¹¹ e 38,51 %, respectivamente.

¹¹ Esse é o valor do τ_a , que não considera os “nós”, que ocorrem quando as observações mantêm sua posição relativa nos dois rankings comparados. O valor de τ_b , que considera esse nós, mostrou-se muito próximo ao primeiro, assumindo valor igual a 26,45%.

Convém comentar que era esperado que o primeiro coeficiente fosse menor que o segundo. Em relação ao teste t para pares, obtivemos que a diferença entre as médias das duas distribuições é estatisticamente significativa. Por fim, para o teste de Wilcoxon, a hipótese nula de que o primeiro ranking é igual ao segundo é fortemente rejeitada. Com base nesses testes, portanto, temos fortes indícios de que os escores ajustados são realmente diferentes dos originais.

Sobre a evolução dos escores médios regionais, a região Nordeste apresentou maiores incrementos, tanto em termos absolutos como relativos. Isso era esperado na medida em que sabemos que essa região enfrenta maiores problemas socioeconômicos que o Sudeste. Como parte desse resultado constatou-se que das 259 escolas que tiveram seus escores reduzidos, do primeiro para o terceiro estágio, 242 estão na região Sudeste. Cabe ainda destacar que apesar da maior evolução do escore médio da região Nordeste, esse ainda assim permaneceu um pouco abaixo do escore médio da outra região. Diante dos resultados, naturalmente surge a curiosidade acerca das práticas presentes nas escolas eficientes que as levam a esse patamar. Contudo, isso foge ao escopo deste trabalho.

6. Conclusões

Neste trabalho, aplicamos o método DEA em três estágios (Fried *et al.*, 2002; Tone; Tsuitsui, 2009), a variáveis extraídas dos microdados, tanto do Censo Escolar como da Prova Brasil 2007 (INEP, 2007a, 2007b), para 8.155 escolas que oferecem EF, localizadas no Sudeste e no Nordeste do Brasil, mediante a aplicação do método DEA-BCC aos dados acima, com orientação para produto. Daí, obtivemos os escores de eficiência para cada escola, além dos *slacks* totais relativos às variáveis de produto, que são o resultado adicional que cada escola deve lograr para se tornar plenamente eficiente. Nessa etapa observamos que as escolas estudadas precisariam aumentar seus resultados em média em 32%, mantidas as proporções entre eles, a fim de atingirem eficiência técnica máxima. No entanto, a utilização desse resultado para cobrar melhor desempenho das unidades escolares poderia fazer com que tais escolas fossem responsabilizadas por resultados que não se encontram sob seu domínio, já que muitas das chamadas variáveis ambientais estão relacionadas a características

dos estudantes e de suas famílias, as quais não estão sob o controle direto da escola. Para corrigir esse tipo de distorção, estimamos, a partir desses resultados iniciais, três fronteiras estocásticas de produção, cujas variáveis dependentes foram os mencionados *slacks* totais, calculados para cada variável de produto. Nessa segunda etapa, verificamos que o NSE dos alunos, assim como as características de suas famílias, influi significativamente nos níveis de eficiência das escolas.

Destaca-se pela magnitude de seu efeito estimado, principalmente para os alunos de 4^a série, o fato de os professores corrigirem as tarefas de casa de matemática frequentemente. Diferentemente das variáveis que refletem as características dos alunos e de suas famílias, o hábito de corrigir as tarefas de matemática é algo passível de modificação por parte das escolas e que, portanto, pode e deve ser estimulado dentro de cada uma delas, já que essa é uma medida simples e de grande efeito que pode ser tomada para melhorar o desempenho dos alunos. Por isso, os diretores das escolas poderiam cobrar mais dos professores esse tipo de atividade. Outras variáveis que mostraram interferir positivamente no desempenho das escolas foram aquelas relacionadas às suas boas condições de infraestrutura, à melhor capacitação dos professores, à escolha dos diretores ser feita por meio de seleção combinada com eleição e ao fato dos diretores se dedicarem exclusivamente às escolas. Por outro lado, a variável que mostrou possuir maiores efeitos no sentido de aumentar a ineficiência das escolas foi a proporção de abandono por parte dos alunos, podendo tal variável ser interpretada de várias maneiras, até mesmo como um indicador da qualidade das escolas, como já comentado.

Nossos resultados sugeriram também a existência de tamanhos ótimos tanto de salas como de escolas. Esse tipo de informação poderá ser útil, por exemplo, na definição dos tamanhos das classes por parte das escolas ou, até mesmo, por parte do poder público. Unidades escolares com turmas muito pequenas poderão decidir aumentá-las até certo tamanho limite, e, desse modo, liberar alguns professores para a realização de outras atividades dentro das escolas. No nosso caso, encontramos que esse tamanho limite está entre 47 e 50 alunos, para classes de 4^a e 8^a séries, respectivamente.

O poder público poderá ainda determinar que as turmas sejam de certo tamanho, o que poderá ensejar menor necessidade de contratação de professores e funcionários.

Referências

- ASSOCIAÇÃO BRASILEIRA DE EMPRESAS DE PESQUISA. *Critério de Classificação Econômica Brasil: Dados com base no Levantamento Socioeconômico 2006 e 2007* – IBOPE. 2009.
- AIGNER, D.; LOVELL, C. A. K.; SCHMIDT, P. Formulation and Estimation is Stochastic Frontier Production Function Models. *Journal of Econometrics*, p. 21-37, 1977.
- BANKER, R. D.; CHARNES, A.; COOPER, W. W. Some Models for estimating technical and scale inefficiencies in Data Envelopment Analysis. *Management Sciences*, v. 30, p. 1078-1092, set. de 1984.
- BANKER, R. D; R.C. MOREY. Efficiency analysis for exogenously fixed inputs and outputs. *Operations Research*, v. 34, p. 513-521, 1986.
- BANKER, R. D.; NATARAJAN, R. Evaluating contextual variables affecting productivity using data envelopment analysis. *Operations Research*, v.56, p. 48-58, 2008.
- BRADLEY, S.; JOHNES, G.; MILLINGTON. The effect of competition on the efficiency of secondary schools in England. *European Journal of Operational Research*, v.135, n. 3, 2001.
- CAUDILL, S. B.; FORD, J.M. Biases in Frontier estimation due to heteroscedasticity. *Economics Letters*, p. 17–20, 1993.
- COOPER, W.; SEIFORD, L.M.; TONE, KAORU. *Data Envelopment Analysis*. 2ª Ed: Editora Springer. 2007.
- CORDERO, J. M.; PEDRAJA, F.; D. SANTÍN. Alternative approaches to include exogenous variables in DEA measures: A Comparison using Monte Carlo. *Computers and Operations Research*, v. 36, p. 2699-2706, 2009.
- DELGADO, V. M. S.; MACHADO, A. F. Eficiência das Escolas Públicas Estaduais de Minas Gerais. *Pesquisa e Planejamento Econômico (PPE)*, p. 427-464, 2007.
- FARENZENA, N. (Org.) Custos e condições de qualidade da educação em escolas públicas: aportes de estudos regionais. Brasília: INEP/MEC, 2005.
- FARRELL, M. J. The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society*, v. 120, n. 03, p. 253-290, 1957.
- FRANCO, C. *et al.* Qualidade e equidade em educação: reconsiderando o significado de “fatores intraescolares”. *Ensaio: avaliação e políticas públicas em educação*, v.15, n. 55, Rio de Janeiro, abr./jun. de 2007.
- FRIED; H. O. *et al.* Accounting for Environmental Effects and Statistical Noise in Data Envelopment Analysis. *Journal of Productivity Analysis*, p.157-174, 2002.
- FRIED; H. O.; LOVELL, C. A. K.; Schmidt, S. S. Efficiency and Productivity. In: FRIED, H. O.; LOVELL, C. A. K.; SCHMIDT; S. S. *The measurement of productive efficiency and productivity growth*. Oxford University Press, 2008.
- GLEWWE, P.; HANUSHEK, E.; HUMPAGE, S.; RAVINA, R. School resources and educational outcomes in developing countries: a review of the literature from 1990 to 2010. Center for International Food and Agricultural Policy, University of Minnesota, Department of Applied Economics, *Working Paper WPI2-1*, jan. de 2012.
- GUTIERREZ, R. G.; CARTER, S.; DRUKKER, D. M. sg160: On boundary-value likelihood-ratio tests. *Stata Technical Bulletin*, n. 60, p. 15–18. Reimpresso em: *Stata Technical Bulletin Reprints*, v. 10, p. 269–273, College Station, TX: Stata Press, 2001.
- HANUSHEK, E. A. The evidence on class size. Occasional paper n. 98-1, 1998.
- HANUSHEK, E. A.; KIMKO, D. D. Schooling, Labor-Force Quality, and the Growth of Nations. *The American Economic Review*, p. 1184-1208, 2000.
- HANUSHEK, E. A.; LAVY, V.; HITOMI, K. D. Do students care about school quality? Determinants of dropout behavior in developing countries. *NBERWP 12737*, 2006.

- IBGE. *rm_atualizada_2008.xls*, 14 de dez. de 2011a. Disponível em: <ftp://geoftp.ibge.gov.br/organizacao_territorial/municipios_por_regioes_metropolitanas/>. Acesso em: 22 de mai. de 2012.
- _____. *tab2.xls*. Brasil.zip. Indicadores Sociais Municipais: uma análise dos resultados do universo do Censo Demográfico. *Censo 2010*, nov. de 2011b. Disponível em: <http://www.ibge.gov.br/home/estatistica/populacao/censo2010/indicadores_sociais_municipais/indicadores_sociais_municipais_tab_zip.shtm>. Acesso em: 27 de fev. de 2012.
- INEP. *Microdados Censo Escolar 2007*. 2007a.
- _____. *Microdados Prova Brasil 2007*. 2007b.
- _____. *Portal INEP*. Disponível em: <http://portal.inep.gov.br/>. Acesso em: 24 de jan. de 2011.
- KENDALL, M. G. A. New Measure of Rank Correlation. *Biometrika*, v. 30, n. 1/2, p. 81-93, jun. de 1938. Disponível em: <http://www.jstor.org/pss/2332226>. Acesso em: 02 de jan. de 2012.
- MENEZES-FILHO, N.; RIBEIRO, F. P. Os determinantes da melhoria do rendimento escolar. In: VELOSO, F. *et al.* (Org.). *Educação Básica no Brasil*. Rio de Janeiro: Elsevier, 2009. cap. 08, p. 171-188.
- OECD. *Education at a Glance 2011: OECD Indicators*. OECD Publishing, 2011.
- RAY, S. C. Resource-Use Efficiency in Public Schools: A Study of Connecticut Data. *Management Science*, v. 37, n. 12, p. 1620-1628, dez. de 1991.
- RIORDAN, C. *Equality and Achievement: An Introduction to the Sociology of education*. 2. ed. Upper Saddle River, New Jersey: Pearson Prentice Hall, 2004.
- RUGGIERO, J. On the measurement of technical efficiency in the public sector. *European Journal of Operational Research*, 553-565, 1996.
- _____. Non-discretionary inputs in data envelopment analysis. *European Journal of Operational Research*, 111, 461-469, 1998.
- SAITECH *DEA-Solver-PRO*: Professional Version 8.0. 2012.
- SAMPAIO DE SOUSA, M. C.; STOSIC, B. D. Technical efficiency of the Brazilian municipalities: correcting nonparametric frontier measurements for outliers. *Journal of Productivity Analysis*, v. 24, p. 155-179, 2005.
- SLATE, J. R.; JONES, C. H. Effects of school size: A review of the literature with recommendations. *Essays in Education*, v. 13, 2005.
- SPEARMAN, C. The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, v. 15, n. 01, p. 72-101, jan. de 1904.
- STATACORP. *Stata Statistical Software*: Release 11. College Station, Texas: Stata Corp LP.2009b.
- THANASSOULIS, E.; PORTELA, M. C. S.; DESPIĆ, O. Data Envelopment Analysis: the mathematical programming approach to efficiency analysis. In: FRIED, H. O.; Lovell, C. A. K.; SCHMIDT, S. S. *The measurement of productive efficiency and productivity growth*. Nova York: Oxford University Press, 2008. cap. 3, p. 251-420.
- TONE, K.; TSUISUI, M. Tuning regression results for use in multi-stage data adjustment approach of DEA. *Journal of the Operations Research*, v.52, n.02, p. 76-85, 2009.
- WILSON, P. W. Efficiency in Education Production among PISA Countries, with Emphasis on Transitioning Economies. Jun. de 2005.
- WITZIERS, B.; R. J. BOSKER; M. L. KRÜGER. Educational Leadership and Student Achievement: The Elusive Search for an Association. *Educational Administration*, v. 39, p. 398-425, ago. de 2003.