

## EFICIÊNCIA DAS ESCOLAS PÚBLICAS ESTADUAIS DE MINAS GERAIS

Victor Maia Senna Delgado\*  
Ana Flávia Machado\*\*

O presente trabalho desenvolve o método semiparamétrico denominado Análise Envolvória de Dados – Data Envelopment Analysis (DEA) – em dois estágios para detectar a fronteira de eficiência das escolas públicas estaduais de Minas Gerais nos níveis fundamental e médio. A primeira etapa desse processo consiste em calcular a eficiência pela DEA e a segunda em comparar os resultados de eficiência por meio de uma regressão com variáveis de condições socioeconômicas familiares, de infra-estrutura e dotação das escolas. As bases censitárias do Sistema Mineiro de Avaliação da Educação Pública (Simave), que aplica os exames de proficiência de matemática e português para 4ª e 8ª séries do fundamental e 3ª série do ensino médio, e do Sistema Informacional de Custo Aluno (Sica), além do Censo Educacional do Ministério da Educação/Instituto Nacional de Estudos e Pesquisas Educacionais (MEC/Inep) de 2003 permitiram aplicar uma das primeiras análises de eficiência por escola desenvolvida para o ensino básico do país. Os resultados encontrados sugerem que uma complementaridade dos insumos, dentro e fora da escola, possibilita o seu melhor desempenho. Escolas localizadas nas mesorregiões do estado onde há mais abundância de recursos educacionais possuem chance maior de serem mais eficientes e prestarem um ensino de melhor qualidade. Porém, existem bons exemplos de desempenho em regiões mais carentes e, em termos gerais, os resultados do produto educacional do estado podem melhorar bastante, caso se consiga um maior nível de eficiência para as escolas estaduais.

### 1 INTRODUÇÃO

Garantir qualidade e eficiência é uma das preocupações recentes da política educacional do país. Nas últimas décadas, o Brasil conquistou algumas melhorias nos indicadores do seu quadro educacional como o aumento da escolaridade média da população, a diminuição da evasão escolar e do trabalho infantil e uma cobertura maior do ensino fundamental, hoje bastante próximo do objetivo da universalização.

Apesar da favorável evolução dos indicadores, os exames internacionais e a confrontação no mercado de trabalho mostram que a formação do nosso estudante está aquém da prevista quando comparada com a de outros países em desenvolvimento. O impacto de uma qualidade menor implica que um ano a mais no ensino fundamental no Brasil corresponde a um tempo menor de formação nos outros países.

No âmbito dessa constatação, emergem questões referentes à eficiência da provisão de serviços em educação. A alocação de recursos faz parte de um dos

\* Mestre em Economia pelo Cedeplar/UFMG.

\*\*Professora do Cedeplar/UFMG.

desafios com o qual uma sociedade deve lidar em seu dia-a-dia. Por ser tão importante a distribuição de recursos, é preciso que seu uso se dê de forma a promover o máximo de benefício social possível. E, no caso do sistema educacional, essas restrições são reforçadas, uma vez que educação formal não é um bem qualquer.

Este artigo busca avaliar a eficiência nas escolas públicas estaduais mineiras, empregando o método não-paramétrico de Análise Envoltória de Dados – Data Envelopment Analysis (DEA). Esse método permite apontar as escolas que se desempenham melhor em termos de custo-oportunidade, identificando as que oferecem aprendizado maior aos alunos, dados os recursos disponíveis. A escolha de Minas Gerais se justifica pelo fato de o estado refletir o contexto nacional, retratando a diversidade do processo de desenvolvimento socioeconômico brasileiro.<sup>1</sup> Uma outra razão é a disponibilidade de base de dados. Têm-se, em caráter censitário, tanto uma base de dados de custo-aluno de escolas públicas estaduais quanto uma base de avaliação do ensino, ambas desenvolvidas pela Secretaria de Estado e Educação (SEE).

Na seção a seguir, revisam-se a literatura econômica sobre educação e o seu emprego na análise de eficiência por meio da análise envoltória. Na terceira parte, descrevem-se o método DEA-clássico e o recente método DEA-*bootstrap*. Na quarta seção, apresentam-se as três bases de dados principais empregadas: Sistema Mineiro de Avaliação da Educação Pública (Simave); Sistema Informacional de Custo Aluno (Sica) e censo escolar; e a construção das variáveis do modelo. Na quinta parte, analisam-se os resultados dos modelos de eficiência. Na última seção, tecem-se alguns comentários finais.

## 2 DESEMPENHO EDUCACIONAL E EFICIÊNCIA NAS ESCOLAS

A eficiência econômica pretende alcançar o produto máximo, dado determinado volume de recursos. Ou então, estabelecida uma meta para o produto, como conseguiu-la com um gasto mínimo. No âmbito da educação, a eficiência está associada à qualidade do ensino, uma vez que esse atributo permite às crianças, e jovens já adultos, serem mais produtivos e, quiçá, socialmente integrados. Nesse contexto, a eficiência da educação incorpora componente intergeracional ao contrário do estabelecido na definição estrita.

Na linguagem da eficiência econômica, a construção teórica importante é a função de produção ou fronteira de eficiência na educação, formulada pela primeira vez por Coleman *et al.* (1966). Com o objetivo de encontrar evidências de que o

---

1. Minas Gerais é reconhecido na literatura como um caso ilustrativo das desigualdades sociorregionais presentes no país (FUNDAÇÃO JOÃO PINHEIRO, 2000; BDMG, 2002). O estado serve de exemplo para modelos de regressão ou análise de dados, sendo considerada a pesquisa em seu território como uma instrutiva e importante primeira etapa de uma posterior aplicação ao caso brasileiro. Dessa forma, a limitação geográfica do território não constituirá limitação às implicações obtidas neste trabalho.

desempenho dos alunos negros aquém do esperado era devido a uma insuficiência de insumos em suas escolas, Coleman *et al.* (1966) agrupam os dados de educação em uma função de produção para a escola:

$$y_i = f(x_i, z_i, d_i) \quad (1)$$

onde  $y_i$  é o rendimento dos alunos da escola  $i$ ;  $x_i$  são os insumos educacionais supervisionados pela direção da escola;  $z_i$  são características individuais dos alunos e das famílias dos alunos daquela escola; e  $d_i$  são variáveis de ‘dotação’.

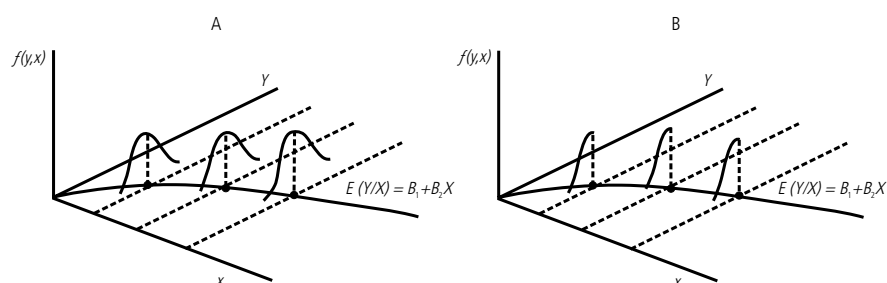
Os resultados do relatório de Coleman e o uso criativo de uma função de produção escolar incentivaram a pesquisa sobre a questão dos insumos educacionais. Hanushek (1986) e Hanushek e Luque (2002) empregam exames de proficiência em uma amostra de escolas norte-americanas e em uma análise entre países, respectivamente, contrastando os resultados das provas com insumos do tipo professor/aluno, salários docentes ou tamanho das turmas. Em ambos os trabalhos, as conclusões são de que os insumos analisados não surtem efeitos sobre o desempenho educacional, tanto nos países desenvolvidos quanto nos em desenvolvimento.

No entanto, usar a função de produção para medir eficiência ou ineficiência requer uma interpretação diferente sobre a teoria. Os trabalhos, até aqui mencionados, consideram a função de produção como um ajuste ideal para os pontos médios da distribuição da variável dependente. A partir da equação (1), os modelos que tomam o  $y$  como produto, por exemplo, pressupõem que o ajustamento correto da função passa pela média condicionada  $E(Y|X)$ . Ocorre, então, que o nível de produto observado pode estar tanto acima quanto abaixo da função, como se observa na figura 1A.

Implicitamente, o que os modelos postos dessa forma supõem é que, em média, as observações são eficientes, ou seja, situam-se sobre a fronteira. Para

FIGURA 1

#### Estimação da fronteira condicionada com erros normais e erros normais truncados



análise microeconômica de firmas ou de unidades gestoras do governo (como é o caso das escolas), é razoável supor que uma ineficiência pode permanecer por um determinado período de tempo. Porém, para se incorporar ineficiência a uma função de produção, são necessárias algumas alterações nos pressupostos estatísticos aplicados à função a ser estimada.

É possível visualizar como a função descrita na equação (1) incorpora ineficiência econômica: basta apenas acrescentar um  $\theta_i$ , índice de eficiência, multiplicado à função. Como o índice de eficiência varia de 0 a 1, temos a possibilidade de incorporar ineficiência a uma função de produção. O valor de  $\theta_i$  inferior à unidade indica que a observação produz aquém do que é possível.

$$y_i = \theta_i f(x_i, z_i, d_i) \quad (2)$$

A partir desse ponto, o problema passa a ser como obter a estimação da função sem considerar os pontos estimados como eficientes na média. Uma maneira paramétrica de se conseguir isso é estabelecer a distribuição dos erros como uma normal truncada de erros não-positivos, método conhecido como *fronteiras estocásticas*. O usual é decompor o erro em duas partes: a dos erros normais ( $v_i$ ) e a dos erros não-negativos ( $u_i$ ). Uma equação para estimação da fronteira pode ser especificada da seguinte forma:

$$\ln(y_i) = x_i \beta + v_i - u_i \quad (3)$$

onde  $\ln(y_i)$  é o produto logaritimizado;  $x_i$  as variáveis para os *inputs*;  $v_i$  os erros aleatórios de acordo com uma distribuição normal; e  $u_i$  erros não-positivos especificados a partir de uma normal truncada ou de uma exponencial. A eficiência ( $\theta_i$ ) é obtida por meio de  $\exp(-u_i)$ . A evolução dos estudos em fronteiras estocásticas acompanha os trabalhos de Aigner e Chu (1968), Aigner, Lovell e Schmidt (1977) e Greene (1993) e a intuição pode ser acompanhada pela figura 1B já mostrada.

Outra maneira de obtenção da fronteira é a estimação não-paramétrica, que não utiliza erros aleatórios. A construção é feita por uma fronteira envoltória de dados (DEA), desenvolvida por Charnes, Cooper e Rhodes (1978). Os autores argumentam que a DEA pode ser um procedimento bastante útil para a análise da eficiência no setor público, uma vez que, ao gestor público, interessa a hierarquização de unidades de análise por determinado conjunto de insumos/ produtos e não a análise dos determinantes dessa hierarquização. A partir de então, seguiu-se uma vasta literatura aplicada a esse setor, a começar com os próprios

autores Charnes, Cooper e Rhodes (1981), que fizeram uma das primeiras aplicações ao caso da educação.

McCarty e Yaisawarng (1993) utilizam dois modelos de medição da eficiência para escolas no distrito de New Jersey, Estados Unidos. No primeiro modelo é desenvolvido o procedimento em dois estágios, onde são identificadas variáveis sob controle das escolas: proficiência, gastos, relação professor/aluno e variáveis que não estão sob a influência da direção como as variáveis socioeconômicas para as quais a política educacional *per se*, a princípio, não tem controle. As primeiras variáveis fazem parte da medição de eficiência obtida pela DEA, as demais são regredidas em um modelo *tobit*. O segundo modelo adotado por esses autores incorpora todas as variáveis dentro da função estimada pela DEA.

Recentemente, dois trabalhos de aplicação da DEA à eficiência da educação se destacam: Wilson (2005) e Afonso e Aubyn (2005). Em comum, os dois estudos apresentam as novas incorporações do método DEA para tornar os índices de eficiência mais robustos. As aplicações utilizam dados de proficiência do Program for International Student Assessment (PISA)<sup>2</sup> – e variáveis não-discrecionárias como *status* socioeconômico e escolaridade dos pais. Wilson incorpora todas as variáveis em uma só estimação da DEA, enquanto Afonso e Aubyn desenvolvem o método de dois estágios da maneira antiga e com as novas incorporações.

No Brasil, são poucos os estudos sobre eficiência utilizando a DEA e, ainda em menor número, os que se aplicam à educação. Porém, podem-se destacar os trabalhos de Marinho, Resende e Façanha (1997) e Façanha e Marinho (1999, 2001), aplicados às instituições do ensino superior brasileiro; de Sampaio de Sousa e Ramos (1999), trabalho que se concentra na eficiência dos gastos públicos municipais em geral; Gasparini e Ramos (2003), sobre efetividade e eficiência, por estados, no ensino médio brasileiro; e Faria e Januzzi (2006), sobre eficiência de gastos na área de educação e saúde dos municípios do Rio de Janeiro.

### 3 MÉTODO DE ANÁLISE ENVOLTÓRIA DE DADOS (DEA)

Os modelos de medição de eficiência, utilizando a DEA, ganharam novo fôlego a partir da segunda metade da década de 1990 e, principalmente nos anos 2000, com as incorporações de Gibels *et al.* (1999), Kneip, Simar e Wilson (2003) e Simar e Wilson (1998, 2002 e 2007). No entanto, seu referencial teórico inicial surge em Debreu (1951). A partir desse marco teórico, Farrel (1957) constrói o novo método não-paramétrico de mensuração da eficiência por programação linear.

2. Exame aplicado entre os 28 países membros da Organização para Cooperação e Desenvolvimento Econômico (OCDE) mais 14 países não-membros: Albânia, Argentina, Brasil, Bulgária, Chile, Hong-Kong (China), Indonésia, Israel, Letônia, Liechtenstein, Macedônia, Peru, Rússia e Tailândia no ano 2000.

A vantagem dos modelos DEA não-paramétricos é sua flexibilidade. Tais modelos assumem poucas hipóteses sobre o comportamento das variáveis e por isso não implicam nenhuma forma funcional *a priori* para a fronteira de educação. Captam a melhor prática existente da organização produtiva e criam um referencial para as instituições analisadas. Estão bem fundamentados teoricamente, com base apenas nos axiomas mais fracos da teoria econômica, e podem aplicar mais de um produto ao mesmo tempo em uma estimação. As desvantagens dos modelos advêm também de sua não-parametricidade: a convergência é lenta. Para pequenas amostras, esse problema pode ser um fator limitador, já que os indicadores que serão obtidos poderão ser inconsistentes; também por esse motivo os modelos possuem um limite do número de variáveis a ser incluído um pouco mais estreito.

A abordagem moderna da DEA define um *conjunto de possibilidades de produção*  $P$ :

$$P = \{(x, y) \mid x \text{ pode produzir } y\}; P \subset \mathfrak{R}^{S+M} \quad (4)$$

onde  $x$  e  $y$  fazem parte de dois conjuntos de vetores de variáveis observadas que são independentes e identicamente distribuídas (*iid*). Em nosso caso,  $x$  é o vetor de *inputs*  $x_{iM} = (x_{i1}, \dots, x_{iM})$ , o subscrito  $i$  identifica a unidade de análise, são  $i = 1, 2, \dots, N$  observações e o subscrito  $M$ , o número de *inputs* diferentes,  $M = 1, 2, \dots, M$  *inputs*;  $y$  é o vetor de *outputs*,  $y_{iS} = (y_{i1}, \dots, y_{iS})$  e  $S$  identifica o número de *outputs*,  $S = 1, 2, \dots, S$ . Dessa forma, temos que  $P$  define a tecnologia educacional e a partir dessa tecnologia obtemos a função de produção estabelecida na equação (1) da seção anterior:  $y = f(x)$ , porém, aqui representada em termos vetoriais e ainda sem as *environment variables*.

O conjunto de variáveis observadas é definido como  $L_N = \{(x_i, y_i, z_i, d_i)\}$  onde além de  $x$  e  $y$ , temos os vetores com as variáveis socioeconômicas das famílias  $z_i$ , e o vetor das variáveis de *dotação*,  $d_i$ . De  $P$  surge um processo gerador de dados que delimita uma *fronteira de eficiência* a ser captada pela DEA ou pelo método de fronteiras estocásticas.<sup>3</sup> O método não-paramétrico da DEA estabelece que a fronteira será construída somente com os pontos que atingiram o máximo de produto, dado determinado nível de insumos ou com o mínimo de *inputs* para dado nível de *outputs*. Ou seja, a fronteira é delimitada pela melhor prática no esquema de

3. Na linguagem estatística, todos os dados observados na realidade possuem por trás um invisível processo gerador de dados *Data Generating Process* (DGP). Portanto, o DGP é um processo gerador de dados conhecido como a distribuição uniforme, normal, exponencial e os demais processos.

produção em voga. A construção de tal fronteira exige um método de programação linear que trabalha com o seguinte processo:

$$\theta_i = \theta_i(x_i, y_i | P) \equiv \max\{\theta_i | (x_i, \theta y_i) \in P, \theta_i > 0\} \quad (5)$$

$$P(x_i, y_i | C, D) = \left\{ (x_i, y_i) \mid \theta_i y_i \leq \sum_{i=1}^n \lambda_i y_i, \right.$$

sujeito a :

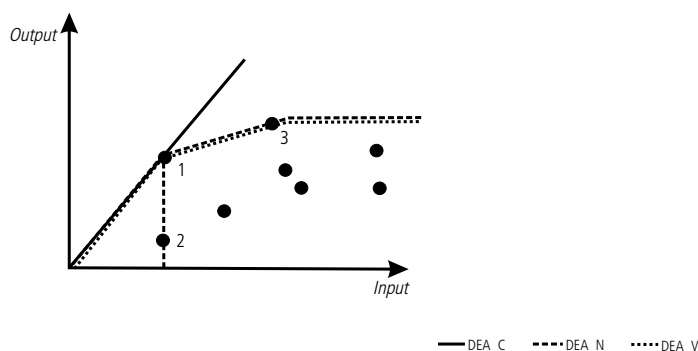
$$x_j \geq \sum_{i=1}^n \lambda_i x_i, \lambda \in \mathfrak{R}^n, y \in \mathfrak{R}^s, x_i \in \mathfrak{R}^m \} \quad (6)$$

A equação (5) revela um procedimento de maximização onde  $\theta_i$  é o *índice de eficiência-Farrel*, medido de 0 a 1, sendo a unidade o indicador de eficiência máxima, situação onde a observação está sobre a fronteira. A eficiência pode recair sobre os insumos no que é chamada de *orientada para o insumo*,  $\theta_i x_p$ , ou sobre os produtos, *orientada para o produto*,  $\theta_i y_p$ , forma que foi exposta anteriormente.

A equação (6) reúne as restrições para a maximização presente em (5). Há alguns novos elementos na equação (6), pois, ao se definir a fronteira de produção é preciso estabelecer duas propriedades microeconômicas. A primeira é a de retornos constantes de escala, indicada por  $C$ , e a segunda é a livre disponibilidade de insumos,  $D$ . O  $\lambda_i$  é um vetor de intensidade,  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N) \in \mathfrak{R}_+^N$ , os  $\lambda$ s denotam os pesos que possibilitam a construção de uma fronteira convexa, outro pressuposto microeconômico, o da possibilidade de combinação convexa de fatores.

Assim, por meio de programação linear, é possível construir as fronteiras de eficiência representadas na figura 2. Cada uma das fronteiras indicadas se assenta em um tipo de rendimentos de escala e na livre disponibilidade de insumos e convexidade de fatores. A partir de alterações na equação de restrição (6), é possível obter dois outros índices além do de rendimentos constantes (DEA-C): o índice de rendimentos não-crescentes, DEA-N, e o índice de rendimentos variáveis, DEA-V. Observa-se que o envelope da DEA-V é o mais maleável, pois envolve os dados a uma distância menor do que as outras duas medidas. Isso se refletirá no índice Farrel de eficiência que, por construção, possui média maior que a DEA-C. A literatura sobre eficiência educacional não estabelece qual retorno de escala seria mais apropriado para a fronteira de eficiência educacional. Charnes, Cooper e Rhodes (1981) empregam retornos constantes. Fare, Grosskopf e Weber (1989) e McCarty e

FIGURA 2  
Construção das fronteiras de eficiência DEA



Yaisawarng (1993) empregam retornos variáveis. Wilson (2005) estabelece uma comparação entre os métodos DEA e *Free Disposal Hull* (FDH) (DEA livre de convexidade). Neste trabalho, escolheu-se a DEA-C, mas as três formas de medição são úteis para a obtenção dos rendimentos de escala das observações eficientes e são aplicadas para cada uma das séries.<sup>4</sup> As fronteiras estão representadas no  $\mathcal{R}^2$  (um produto, um insumo) por simplificação.

Os pontos representam as observações  $(x_j, y_j)$  e no caso da DEA-C, a fronteira é sempre uma reta, um plano ou hiperplano. A hipótese de retornos constantes pode ser alterada para a possibilidade de retornos não-crescentes, *N*, onde uma nova restrição  $\sum \lambda_i \leq 1$  não permite expansões radiais por um escalar maior do que 1, ou alterada para a possibilidade de rendimentos variáveis de escala, *V*, obtidos por meio da adição da restrição  $\sum \lambda_i = 1$ . A soma dos escalares está presa à unidade, indicando que a fronteira é construída aos pares de pontos eficientes. Na DEA-V os pontos se reportam sempre a um segmento de fronteira ( $\lambda_i + \lambda_j = 1$ , onde  $i \neq j$ ).

A fronteira DEA-V possui ao menos um ponto em comum com a fronteira DEA-C, como o ponto 1 exemplificado na figura 2. Isso indica que, quando  $\theta_{iDEA-V} = \theta_{iDEA-C}$  tem-se rendimentos constantes de escala (RCE). Se não são iguais, cabe verificar duas outras possibilidades: se  $\theta_{iDEA-V} = \theta_{iDEA-N}$ , variável e não-crescente, o que indica rendimentos decrescentes de escala (RDE); ou se  $\theta_{iDEA-V} \neq \theta_{iDEA-C} \neq \theta_{iDEA-N}$ , caso onde resta apenas a possibilidade de retornos crescentes (RCC).

A construção dos índices é feita radialmente a partir da origem e em relação à fronteira. Trata-se da distância radial entre o ponto observado e a fronteira de

4. Pode-se obter os rendimentos das observações não-eficientes a partir de sua projeção na fronteira. A escala é obtida comparando-se os índices das observações com o nível de produto necessário para deixarem de ser ineficientes.



eficiência. Como observado, duas formas de construção são possíveis: a eficiência *orientada para o produto* e a *orientada para o insumo*. Sob retornos constantes de escala, as duas modalidades fornecem os mesmos resultados nos índices de eficiência, e resultados diferentes sob retornos não-crescentes ou variáveis.<sup>5</sup>

Neste trabalho, utilizamos o índice de eficiência medido pelo critério de Sheppard (1970) orientado pelo produto e denotado pelo  $\delta_i$ . Não há muita diferença entre o *Sheppard-eficiência* e o *Farrel-eficiência*, apenas a construção: um caracteriza-se por ser a inversa do outro índice.<sup>6</sup>

A real técnica de produção  $P$  é observada por meio de uma realização estocástica denominada  $\hat{P}$ . Para se obter uma distribuição dos dados realizados em  $\hat{P}$  que se aproxime do  $P$ -real é empregado o *bootstrap* proposto em Simar e Wilson (1998) que elimina o viés dado pelo último termo da equação (7) a seguir:

$$\hat{\delta}_i = \delta_i + \vartheta_p \left( \frac{n-2}{m+s+1} \right) \quad (7)$$

O  $\hat{\delta}_i$  é o índice observado, o  $\delta_i$  é o índice real desconhecido. O viés dado por  $\vartheta_p$  se aproxima de 0 quando  $n \rightarrow \infty$  e a velocidade da convergência diminuem conforme aumenta a dimensão ( $m + s$ ). No entanto, para se conseguir estimar o tamanho do viés, é preciso conhecer a distribuição dos índices de eficiência  $\hat{\delta}_i$ , resultado impossível de se obter analiticamente no caso de um conjunto multivariado onde  $s + m > 2$ .<sup>7</sup> O *bootstrap* é uma maneira de aproximar assintoticamente a distribuição dos estimadores. A partir da técnica proposta, é possível se obter uma série de estimativas mais confiáveis para os índices de eficiência, assim como construir intervalos de confiança, impossíveis de serem obtidos sem a nova técnica.

5. A rigor, o conjunto de eficientes é DEA-V  $\supseteq$  DEA-N  $\supseteq$  DEA-C, ou seja, não há eficiente no DEA-C que não seja no DEA-V, porém é comum a DEA-V possuir um pouco mais de observações eficientes, algo demonstrado pela figura 2. A ordem pode se alterar um pouco de acordo com a escala, porém a fronteira não se altera com o tipo de orientação. Coelli, Rao e Battese (1998) ressaltam que a escolha entre *orientação para o insumo* e *orientação para o produto* não é crucial, pois apenas mudam o valor medido em cada um dos índices ineficientes.

6. A adoção do *Sheppard-eficiência* em *orientada para o produto* baseia-se no fato de que nessa forma de medição podemos comparar o índice de eficiência a uma medida de capacidade que varia de 0% a 100%. Dessa forma, uma observação eficiente ( $\delta_i = 1$ ) está produzindo 100% da sua capacidade presumida. Observe-se que, por ser a sua inversa, para se obter o mesmo efeito de interpretação com *orientação para o insumo*, ter-se-ia que usar a *Farrel-eficiência*.

7. Gibels et al. (1999) demonstram a convergência para o caso de uma função estimada no  $R^2$ .

### 3.1 O novo método de dois estágios

O método de dois estágios emprega duas etapas de estimação: a primeira é a mensuração não-paramétrica da DEA abordada na seção anterior; e a segunda é a estimação dos parâmetros das variáveis que podem influir nos índices de eficiência obtidos no primeiro estágio, com o uso do método dos Mínimos Quadrados Ordinários (MQO) ou outra técnica.<sup>8</sup> A regressão mais simples dos modelos de dois estágios envolve o índice de eficiência  $\delta_i$  e as variáveis do ambiente, destacadas aqui como  $z_i$ :

$$\delta_i = \beta z_i + \varepsilon_i \quad (8)$$

O  $\beta_i$  é um vetor de parâmetros empregados para captar a influência de  $z_i$  sobre o índice estimado. Muitas variáveis externas ao controle gerencial da escola impactam em sua eficiência. É praticamente impossível incorporar todas ao modelo DEA. Em primeiro lugar, porque a presença de muitas variáveis dificultaria a convergência dos estimadores e, em segundo, porque optamos por deixar, no primeiro modelo, apenas variáveis gerenciais-pedagógicas que estão sob o controle das escolas ou das superintendências de ensino. Os  $\varepsilon_i$ s são erros da regressão, onde  $E(\varepsilon_i) = 0$  e  $E(\varepsilon_i^2) = \sigma^2$ .

Não há erros aleatórios na obtenção da DEA; sendo assim, não é difícil perceber que os índices de eficiência estão todos de alguma forma *correlacionados* entre si. Uma observação ou é eficiente, ou se reporta a, no mínimo, duas outras observações que formam a fronteira. A correlação entre os  $\delta_i$ s nos impede de regressar uma equação simples da forma exposta em (8), os  $\varepsilon_i$ s denotariam a autocorrelação e a estimação seria menos estatisticamente eficiente. É necessária uma correção, e a pesquisa identificava esse problema: o *tobit* não se tratava de uma correção perfeita. Os avanços foram se dando até chegar na idéia de *bootstrap* da fronteira não-paramétrica.

Em síntese, a idéia da aplicação dos modelos *bootstrap* é obter maior robustez nos métodos não-paramétricos. Observar o comportamento da fronteira não apenas com a base de dados inicial, mas capturar, por meio desta, sua distribuição provável e conseguir, dessa maneira, a consistência dos estimadores (retirando-lhes o viés), os intervalos de confiança e a aplicação dos testes de hipóteses. Quando se obtém uma mesma fronteira várias vezes por *bootstrap*, retira-se a correlação existente nos  $\delta_i$ s, exclui-se, também, a parte aleatória do erro que pode ocasionar uma ineficiência.<sup>9</sup>

8. Por conta da combinação das duas técnicas, costuma-se denominar semiparamétrico o modelo em dois estágios.

9. Como se pode imaginar, dados sobre variáveis econômicas apresentam discrepância estatística (erro). Nesses casos, uma observação pode ser declarada eficiente ou ineficiente apenas pela ocorrência de tal discrepância. O uso da DEA-*bootstrap* evidencia esse problema. Dada a distribuição do DGP descrita, os estimadores de eficiência-*bootstrap* passam a ter um componente aleatório e não tão determinístico para sua distribuição.

O índice de eficiência observado,  $\hat{\delta}_i$ , é dado pela esperança do índice,  $E(\hat{\delta}_i)$ , mais um componente de erro aleatório,  $u_i$ :

$$\hat{\delta}_i = E(\hat{\delta}_i) + u_i \quad (9)$$

onde  $E(u_i) = 0$ , definindo que assintoticamente o estimador é consistente. Porém, com a convergência lenta a correlação de  $\mu_i$  não desaparece rápido o suficiente para aproximações-padrão de inferência. Tem-se, então, que definir o viés de  $\hat{\delta}_i$  dado pela diferença entre a esperança do estimador observado e o verdadeiro:

$$viés(\hat{\delta}_i) \equiv E(\hat{\delta}_i) - \delta_i \quad (10)$$

Para se descobrir o parâmetro verdadeiro, manipula-se a equação (10):

$$\delta_i = E(\hat{\delta}_i) - viés(\hat{\delta}_i) \quad (11)$$

Por (9) também tem-se que  $E(\hat{\delta}_i) = \hat{\delta}_i - u_i$ , que substituindo em (11) fica:

$$\delta_i = \hat{\delta}_i - viés(\hat{\delta}_i) - u_i \quad (12)$$

A parte do  $viés(\hat{\delta}_i) - u_i$  é negligível assintoticamente; lembrando (8) tem-se:

$$\delta_i = \hat{\delta}_i = \beta z_i + \varepsilon_i \quad (13)$$

Mostrando que a eficiência verdadeira pode ser estimada nos dois estágios, desde que o viés seja considerado no estimador observado  $\hat{\delta}_i$ . A regressão de dois estágios original (com o *tobit*) desconsiderava o  $viés(\hat{\delta}_i)$  e o erro  $u_i$ . O uso do *bootstrap* procura uma distribuição para os indicadores e estima o viés. Cada obser-

vação permanece, portanto, apenas com a parte que confere ineficiência devido ao mau uso dos fatores e não mais à atribuição de discrepância estatística.

A equação (13) é reescrita da seguinte maneira:

$$\delta_i^* = \beta z_i + \varepsilon_i \quad (14)$$

O índice de eficiência  $\delta^*$  é o *Sheppard-Eficiência* obtido por *bootstrap*, cujas propriedades assintóticas foram tratadas na seção anterior. Ao utilizá-lo, elimina-se a correlação dos erros com as variáveis  $z_i$  e com  $\delta^*$ , assintoticamente a convergência será mais rápida.<sup>10</sup> Simar e Wilson (2007, p. 41) fornecem os passos para se obter a regressão citada anteriormente na prática:

Algoritmo #1 Simar e Wilson:

[1] Usando-se os dados originais em  $L_N = \{(x_p, y_p, z_p, d_i)\}$ , calcule-se  $\hat{\delta}_i = \hat{\delta}(x_i, y_i | \hat{P})$ ,  $\forall i = 1, \dots, n$  usando-se a DEA.

[2] Usar o método de Máxima Verossimilhança (MV) para se obter estimações de  $\hat{\beta}$  e do  $\beta$ -verdadeiro, assim como uma estimativa de  $\hat{\sigma}_\varepsilon$  e do  $\sigma_\varepsilon$ -verdadeiro na regressão truncada de  $\delta_i$  em  $z_i$  em  $\hat{\delta}_i = \beta z_i + \varepsilon_i \geq 1$ .

[3] Os próximos três passos são um *loop B* vezes para se obter um conjunto de estimadores *bootstrap*:  $A = \left\{ \left( \hat{\beta}_b^*, \hat{\sigma}_\varepsilon^* \right) \right\}_{b=1}^B$ .

[3.1] Para cada  $i = 1, \dots, n$  obtenha-se os erros  $\varepsilon_i$  a partir de  $N(0, \sigma_\varepsilon)$  truncada à esquerda em  $(1 - z_i \hat{\beta})$ .

[3.2] Novamente para cada  $i = 1, \dots, n$  calcule-se  $\hat{\delta}_i^* = z_i \hat{\beta} + \varepsilon_i$ .

[3.3] Use-se MV para se estimar a regressão truncada de  $\hat{\delta}_i^*$  em  $z_i$  obtendo-se as estimativas  $(\hat{\beta}, \hat{\sigma}_\varepsilon)$ .

[4] Usem-se os valores *bootstrap* de  $A$  e as estimativas originais de  $\hat{\beta}, \hat{\sigma}_\varepsilon$  para se construir os intervalos de confiança de  $\beta$  e  $\sigma_\varepsilon$ .

10. O *bootstrap* serve para eliminar o viés. Através da regressão proposta em Simar e Wilson (2007, p. 40), anteriormente representada, obtém-se uma estimação mais eficiente.

Além das considerações sobre o estimador DEA já colocadas, por ser determinístico, o método está sujeito a alterações bruscas da eficiência devido a observações *outliers*, as chamadas *supereficientes*. A ocorrência desses casos desloca toda a fronteira e prejudica a análise das demais observações. Neste trabalho, recorre-se aos métodos de correção de *outliers* propostos por Sampaio de Sousa e Stosic (2005) e Cazals, Florens e Simar (2002).

O método de Sampaio-Stosic é obtido através do cálculo de influência que cada observação produz em todas as demais. Para captar a influência de cada observação, os autores utilizam um estimador  $l_j$  (*leverage*) especificado na equação (15) a seguir:

$$l_j = \sqrt{\frac{\sum_{k=1; k \neq j}^k (\delta_{kj}^* - \delta_k)^2}{K-1}} \quad (15)$$

O  $k$  é o conjunto de unidades de análise de 1 até  $K$ .<sup>11</sup> O  $\delta_k$  denota o conjunto de índices de eficiência calculados para todas as DMUs  $\{\delta_{kj} \mid k = 1, \dots, K\}$ ; o  $\delta_{kj}^*$  denota esse conjunto de eficiência com a alavancagem, isto é, sem a DMU  $j$ ,  $\{\delta_{kj}^* \mid k = 1, \dots, K, k \neq j\}$ . Como observam os autores, a idéia básica é a de que se espera grande *leverage* das observações *outliers*. Se  $l_j$  destoar muito da média global tem-se a suspeita de ser um *outlier*. Quando a observação retirada está dentro da fronteira delimitada por  $P$ , tem-se  $(\delta_{kj}^* - \delta_k) = 0$ , o que denota uma observação não-influente,  $l_j = 0$ . No caso extremo de uma observação superinfluyente, sua retirada faz com que todas as outras subam para 1, assim  $\sum (\delta_{kj}^* - \delta_k) = K - 1$ , fazendo com que  $l_j = 1$ . Portanto,  $0 \leq l_j \leq 1$ .

O segundo teste de detecção de *outliers* empregado paralelamente foi proposto por Cazals, Florens e Simar (2002), nomeado por CFS. O teste CFS usa um engenhoso princípio de *trimming parameter*. A idéia é construir fronteiras mais internas já que, na presença de *outliers*, a primeira fronteira a ser encontrada estará muito distante das outras informações.

11. Aqui ocorre a mudança da notação utilizada por Sampaio-Stosic para evitar confusão entre a parte de cálculo de eficiência com a do cálculo dos *leverages*.

#### 4 FONTE DE DADOS E TRATAMENTO DAS VARIÁVEIS

O emprego de dados educacionais sobre o desempenho e custo por aluno em Minas Gerais é um dos pontos relevantes deste artigo. Neste estudo, estão combinadas três fontes de dados. Os dados do Simave permitem a abordagem de aspectos ligados à qualidade do ensino. A base do Sica, por sua vez, permite tratar de aspectos associados à eficiência. O Censo Educacional do Ministério de Educação e Cultura/Instituto Brasileiro de Geografia e Estatística (MEC/IBGE) possibilita construir variáveis de infra-estrutura e oferta de serviços educacionais. Essas três bases visam compor as variáveis de *produtos*, *insumos*, *dotação* e *variáveis socioeconômicas* dos alunos matriculados em determinada escola. É importante ressaltar que a unidade de análise é a escola. Desse modo, as informações sobre alunos se referem à média das características dos mesmos no estabelecimento de ensino.

O Simave é um sistema de avaliação que tem como base o Programa de Avaliação da Rede Pública da Educação Básica (Proeb). A cada ano, os alunos das turmas da 4ª e da 8ª séries da rede pública estadual, assim como os alunos da 3ª série do ensino médio, realizam provas em uma das disciplinas básicas do currículo. Dessa forma, em 2002, os estudantes fizeram prova de língua portuguesa e, em 2003, de matemática. A comparação das notas é possível pelo critério de Resposta ao Item, método que permite a identificação do desempenho dos alunos por item específico e possibilita a comparação das notas entre os diversos anos do ciclo educacional. Assim, podemos comparar a evolução das médias por série avaliada.<sup>12</sup>

A utilização de duas provas em vez de uma é justificada pela avaliação de que dois exames são mais representativos do que um. Credita-se isso ao fato de se evitar, com os dois exames, oscilações muito bruscas de média a que estão sujeitas as escolas com turmas muito pequenas.

A análise da distribuição das notas revela o caráter normal e bastante simétrico desse tipo de informação. Não foram detectados *outliers* e as distribuições não requisitaram nenhuma transformação. A tabela 1 a seguir resume as principais informações por disciplina e série.<sup>13</sup>

Tão importantes quanto os exames de proficiência são as informações do questionário socioeconômico das famílias aplicado pelo Simave. A tabela 2 resume as variáveis utilizadas na análise por escolas. Tais variáveis se apresentam na proporção de alunos da escola.

12. Sobre o critério de Resposta ao Item, ver Soares e Pereira (2002).

13. Os dados originais das provas do Simave estão desagregados por alunos. Como o objetivo é apresentar informações por escolas, os resultados são reportados pela média. Portanto, a distribuição na tabela apresentada é a distribuição de médias por escola e não dos alunos.

TABELA 1

**Resumo dos resultados das provas por anos de ensino**

	Média	Desvio	Min.	Max.
4ª série				
Matemática	183,73	25,22	106,72	284,93
Português	186,72	22,86	53,25	249,86
8ª série				
Matemática	237,49	20,24	164,00	315,56
Português	239,85	13,90	164,06	296,77
3ª série				
Matemática	270,07	20,05	218,13	340,88
Português	269,92	12,12	208,84	331,81

Fontes: Simave de 2002 e 2003.

TABELA 2

**Variáveis do questionário do Simave**

Variáveis socioeconômicas de alunos e de sua família (segundo estágio)

cor	1- Brancos e amarelos, 0 pretos e pardos
esc_res_fem	1- Responsável do sexo feminino tem 8ª série ou mais, 0 caso contrário
livros	1- Mais do que 20 livros em casa, 0 caso contrário
computador	1- Computador na residência, 0 caso contrário
ler_liv_inf	1- Se o aluno leu livros de histórias infantis, 0 caso contrário
ler_jornais	1- Se o aluno leu jornais durante o ano, 0 caso contrário
ntrab_fora	1- Não trabalha fora de casa, 0 caso trabalhe
ntrab_doméstico	1- Aluno não exerce tarefas domésticas, 0 caso gaste algum tempo
devermat	1- Se faz sempre o dever de matemática, 0 caso contrário
reprovado	1- Não reprovou de ano nenhuma vez, 0 caso tenha reprovado
Noite	1- Aluno estuda à noite, 0 caso contrário

Fonte: Simave de 2003.

A segunda fonte de dados dessa pesquisa é o Sica que fornece informações de gasto por aluno para cada escola estadual através da coleta de informações sobre o financiamento dos vários níveis de ensino: infantil, fundamental e médio. O sistema

foi desenvolvido pela Superintendência de Planejamento da Secretaria do Estado de Minas Gerais em 1997 e faz parte do Sistema de Informações sobre Orçamentos Públicos em Educação (Siope).<sup>14</sup>

O Sica reúne o gasto por aluno dentro das divisões dos chamados custos *dentro* e custos *fora* da escola. Os custos *dentro* são compostos em *diretos*, custos com profissionais da educação e convênios, e *indiretos*, no qual entram os custos com o pessoal administrativo e as despesas gerais. Os custos *fora* são aqueles que não estão direcionados ao custeio do pessoal ativo e tampouco às despesas escolares. Neste trabalho utiliza-se apenas o custo dentro da escola, visto que se enquadra melhor à análise e se refere a gastos diretamente ligados à nossa unidade de observação.

Por fim, a terceira fonte de dados é o Censo Escolar de 2003, realizado desde 1998 pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep/MEC). As variáveis do censo estão separadas, neste artigo, em duas categorias: as agrupadas e as não-agrupadas. A primeira categoria engloba os modelos DEA no primeiro estágio; são informações que se apresentaram mais úteis de maneira agrupada. A segunda categoria é a das variáveis não-agrupadas, utilizadas na

TABELA 3

**Variáveis agrupadas e não-agrupadas do censo escolar**

Variáveis agrupadas (1º grupo):

<i>qdocente</i>	Número de professores com 3º grau – licenciatura completa para 4ª e 8ª séries, e 3ª série (ensino médio), respectivamente
<i>Salas</i>	Número de salas de aula de toda a escola
<i>Infra</i>	Índice de 0 a 5 que verifica a presença de sanitário; energia pública; água, esgoto e coleta periódica do lixo
<i>Vcm</i>	Variável obtida por <i>Principal Component Analysis</i> para captar concomitantemente presença de aparelhos de "vídeo", "computadores" e recursos de "multimídia"

Variáveis não-agrupadas (2º grupo):

<i>Urbana</i>	1 – Se escola se situa em área urbana, 0 caso contrário
<i>matriculas</i>	Número de alunos matriculados separado por período: 1ª a 4ª série, 5ª a 8ª série e ensino médio
<i>Meso</i>	<i>Dummy</i> para mesorregião do estado
<i>Fnde</i>	1 – Escolas que recebem o Fundo Nacional de Desenvolvimento da Educação, 0 caso contrário
<i>mere_esc</i>	1 – Escola oferece alimentação aos alunos

Fonte: Censo Escolar MEC/Inep de 2003.

14. Sistema recém-implementado pelo Inep: <<http://www.siope.inep.gov.br/apresentacao.do>>.



maneira em que estão disponíveis no censo e empregadas na etapa do segundo estágio, com exceção das matrículas, que fazem parte dos *produtos* do primeiro estágio.<sup>15</sup>

Os dados do Simave nos mostram que, para as escolas do ensino fundamental, a minoria é composta por alunos brancos, 41% na 4ª série, 45% na 8ª série, alternando para uma pequena maioria, (51%) na terceira série do ensino médio. Há, também, nesta última série do ensino médio, a presença significativa de escolas onde os alunos estudam à noite, 69%. Quase todos os indicadores socioeconômicos se destacam negativamente, com a notável exceção da proporção de leitura de livros (e a leitura de jornais, para o caso específico da 3ª série do ensino médio). Nas escolas do ensino fundamental, os indicadores positivos são a grande parcela de alunos que não trabalha fora de casa, 84%; e 64% para 4ª e 8ª séries, respectivamente, além do número de alunos que faz o dever de matemática (71% e 55%, respectivamente, na 4ª e 8ª séries, e que não são reprovados; 73% na 4ª série e 67% na 8ª série). A tabela 4 reúne informações das escolas públicas estaduais mineiras.

No que tange à variável de custo, após a retirada das escolas *outliers*, a variável *cdentro* apresenta ainda uma cauda direita bastante acentuada, entretanto, as análises

TABELA 4

**Resumo das informações do questionário do Simave para todas as séries**

Variável	4ª série	8ª série	3ª série
cor	0,41	0,45	0,51
esc_resp_fem	0,31	0,19	0,20
livros	0,22	0,24	0,26
computador	0,13	0,11	0,11
Ler_livro_inf	0,78	0,70	0,71
Ler_jornais	0,44	0,45	0,62
ntrab_fora	0,84	0,64	0,47
ntrab_domestico	0,28	0,26	0,25
devermat	0,71	0,55	0,42
nreprovado	0,73	0,67	0,63
noite	-	-	0,69

Fonte: Simave de 2003.

15. Os termos *agrupadas* e *não-agrupadas* a que nos referimos dizem respeito à maneira como os dados estavam disponíveis no censo escolar. *Agrupadas* se referem a dados que se constituíram em mais de uma variável no censo e foram reunidas em uma só categoria. *Não-agrupadas* são apresentadas de acordo com a sua disposição no censo e representam apenas uma variável.

demonstraram que não seria preciso um tratamento maior do que o já dado. A seguir temos a apresentação na tabela 5 para essa variável.

No censo escolar, a variável *condições de infra-estrutura (Infra)* provém da soma de cinco variáveis *dummy*: *sani\_den*, observa se o sanitário está dentro da escola; *ener\_pub*, capta se a escola é ligada à rede pública de energia elétrica; *agua\_pub*, verifica se a escola possui sistema de água da rede pública; *esg\_pub*, se a escola é ligada ao esgotamento da rede pública; e *lixo\_col*, se o destino do lixo é coleta periódica da prefeitura. Na tabela 6, nota-se que algumas escolas não possuem nenhuma condição de infra-estrutura, sendo que é crescente, ao longo de cada série, o número das que apresentam algumas ou todas, além disso, observa-se que as condições de infra-estrutura também evoluem com o avanço das séries.

A variável *VCM* de vídeo, computador e multimídia, construída pelo método de PCA, combina variáveis que isoladas não se apresentavam muito significativas, possuíam pouca correlação com os resultados das provas, porém eram bastante correlacionadas entre si. *Apvídeo* é o número de aparelhos relacionados à televisão, uma variável quantitativa obtida através da soma entre *vvideo*, *vtv* e *parabol*, quantidade de videocassetes, televisores e antenas parabólicas, respectivamente. *Comp&Imp* é a variável que indica o número de aparelhos de informática que a escola possui. São agregados, computadores (qualquer tipo) e impressoras.

TABELA 5  
Tabela com resumo dos custos dentro da série

	Média	Desvio	Min.	Max.
4ª série	56,49	23,38	9,06	329,33
8ª série	52,33	17,72	4,11	180,60
3ª série	41,02	17,20	1,34	169,85

Fonte: Sica de 2005.

TABELA 6  
Tabela com resumo para variável de infra-estrutura

	Frequência (%)						Média do índice
	0	1	2	3	4	5	
4ª série	0,53	3,46	9,85	8,18	13,79	64,20	4,24
8ª série	0,08	1,22	5,29	6,40	14,93	72,07	4,51
3ª série	0,00	0,14	1,62	4,30	13,88	80,06	4,72

Fonte: Censo Escolar MEC/IBGE de 2003. Variável *Infra* construída, ver tabela 3.

Obs.: O indicador 0 denota a completa ausência dos itens avaliados; 5 é o total.

*Multimídia* capta os recursos dentro da escola disponíveis aos alunos, pois engloba a soma das variáveis originais: *biblioteca*, *videoteca*, *salvivid* (sala de TV e vídeo), *sala\_lei* (sala de leitura) e *lab\_info* (laboratório de informática).<sup>16</sup>

Do grupo das variáveis desagregadas destacam-se *matrículas*, variável que é utilizada como indicador de *produto*. Para o caso das matrículas, é preciso um cuidado semelhante ao que foi abordado na parte de custos devido à dispersão considerável dos valores dessa variável, distribuição assimétrica à direita, com o mínimo de cinco alunos em uma pequena escola do primeiro ciclo até um máximo de 4.078 na maior escola em termos de matrícula no ensino médio. As médias de matrículas se elevam ao longo das séries e o desvio-padrão aumenta consideravelmente, o que pode ser observado na tabela 7. As informações estão apresentadas por ciclo: matrículas de 1ª a 4ª série foram utilizadas para a DEA 4ª série, matrículas de 5ª a 8ª série, para a DEA 8ª e *matric\_médio* é o total das 1ª, 2ª e 3ª séries do ensino médio, empregada na análise da 3ª série.

À medida que se avança no ciclo escolar, verificam-se a maior urbanização, a menor necessidade de auxílio do Fundo Nacional de Desenvolvimento da Educação (FNDE), ligeira queda na merenda escolar e melhoria geral nos indicadores de infra-estrutura e *VCM* (*TemTV*, *TemPC*, *TemMult*, que indicam apenas se a escola tem ou não os aparelhos e recursos que constam em *VCM*).

Nos três bancos de dados disponíveis, essas variáveis apresentadas foram as escolhidas. Para a DEA no primeiro estágio, são três variáveis de produto: proficiência em matemática, português e matrículas por ciclo. No caso de insumos, são cinco variáveis: custo-aluno, quantidade de professores com formação superior, número de salas, indicador de infra-estrutura e a *VCM*. Para os acontecimentos e características que afetam o resultado, mas não estão sob o controle nem das escolas nem da secretaria, é que se compõe o método de dois estágios que conta com a

TABELA 7

**Resumo do número de matrículas por ciclos do ensino**

	Média	Desvio	Min.	Max.
Matric_1 a 4 (4ª série)	247,64	193,38	5	1.675
Matric_5 a 8 (8ª série)	419,98	252,27	8	1.406
Matric_médio (3ª série)	524,38	411,25	45	4.078

Fonte: Censo Escolar MEC/IBGE de 2003.

16. Os valores do PCA foram transformados de modo a não apresentarem valores negativos que impossibilitam os programas computacionais de calcularem a fronteira. Para *VCM* adotamos a soma de um valor que seja suficiente para deslocar toda a distribuição da variável, obtendo apenas valores positivos. As transformações não distorcem a fronteira, ver Ali e Seiford (1990).

TABELA 8

**Resumo das informações do censo escolar: variáveis não-agrupadas – todas as séries**

Variável	4ª série	8ª série	3º ano
urbana	0,79	0,89	0,99
fnde	0,81	0,79	0,76
merenda	0,96	0,95	0,94
TemTV	0,95	0,98	0,99
TemPC	0,58	0,68	0,74
TemMult	0,83	0,89	0,94
sani_den	0,94	0,97	0,98
ener_pub	0,98	1,00	1,00
agua_pub	0,83	0,91	0,96
esg_pub	0,71	0,78	0,85
lixo_col	0,77	0,86	0,93

Fonte: Censo Escolar MEC/IBGE de 2003.

presença da média das características dos alunos e condições a que as escolas estão sujeitas, podendo algumas delas, inclusive, serem alteradas por meio de políticas públicas. O segundo estágio do modelo inclui as variáveis de *características socioeconômicas* dos alunos: *dummy* urbana (controle), cor, escolaridade da mãe, livros em casa, internet em casa, ler livros ou jornais, não trabalhar fora de casa, não desempenhar trabalho doméstico, realizar o dever de casa, não ser reprovado e, para o caso do ensino médio, estudar à noite. As variáveis de *dotação* são: urbana (controle), merenda escolar, auxílio FNDE, presença de computadores e impressoras, recursos multimídia, sanitários, se a escola está ligada à rede pública de energia, ao abastecimento geral de água e à rede de esgoto e se possui coleta de lixo. Cada um desses conjuntos de variáveis mencionadas, características socioeconômicas e dotação, compõe um tipo diferente de regressão. O regressando será o índice de eficiência (invertido) obtido no primeiro estágio.

## 5 RESULTADOS

O modelo de dois estágios, empregado nesta seção, implementa todas as técnicas recentes enunciadas nas seções anteriores. No primeiro estágio, os resultados são descritos para as 12 mesorregiões de Minas Gerais em cada uma das séries. No segundo estágio, estendemos o alcance dos principais resultados encontrados relativos à eficiência da escola estadual mineira.

### 5.1 Primeiro estágio

Neste primeiro estágio, o índice DEA-eficiência resume as variáveis e revela aspectos que seriam difíceis de serem captados por uma análise exaustiva da relação insumo-produto solucionados um a um. Assim, regiões que, a princípio, poderiam se sair bem na ordenação por notas não se apresentam como eficientes, quando considerados os insumos, enquanto outras que não revelavam desempenho excepcional acerca da proficiência se saem bem, considerando-se o baixo nível de insumos que possuem. A configuração de eficiência no Estado de Minas Gerais é descrita por intermédio de suas 12 mesorregiões retratadas no mapa a seguir. Observa-se uma concentração das eficientes em torno da mesorregião metropolitana.<sup>17</sup>

MAPA

#### Minas Gerais: eficiência média por mesorregião



Para se entender melhor as propriedades da eficiência por escolas, e a construção de suas médias por mesorregiões, analisa-se a eficiência para cada uma das séries. A tabela 9 fornece a média de produtos e insumos para cada uma das 12 regiões do estado no nível da 4ª série. Constata-se que, em geral, em Minas Gerais os custos por aluno aumentam à medida que se deslocam para o sul do estado. As regiões desenvolvidas do centro-sul possuem médias maiores, tanto para o nível de insumos como para o de produtos, enquanto nas mesorregiões pobres do norte, o nível de insumos e produtos é precário.

Percebe-se, portanto, que os resultados dos produtos estão de alguma forma relacionados diretamente com o nível de insumos, mas a determinação de eficiência considera a *melhor prática* dentro de uma escala de rendimentos, ou seja, eficiência é o resultado de um bom balanceamento entre insumos e produtos; muitos insumos

17. A mesorregião Sul/Sudoeste poderia ser incluída nas seis primeiras, em detrimento da mesorregião do Jequitinhonha, por exemplo, mas a motivação aqui é apresentar um padrão geral da eficiência e não uma distribuição regional da eficiência.

TABELA 9

**Média dos produtos e insumos por mesorregião para a 4ª série**

Código	Mesorregião	Escolas	Produtos			Insumos				
			<i>Pmate</i>	<i>Pport</i>	<i>Alunos</i>	<i>Custo</i>	<i>Qdoc</i>	<i>Salas</i>	<i>Infra</i>	<i>Vcm</i>
310001	Noroeste de Minas	54	187,43	188,94	274,91	57,01	4,89	21,09	4,48	20,35
310002	Norte de Minas	411	165,86	167,49	195,89	43,47	2,15	15,74	3,20	19,74
310003	Jequitinhonha	191	171,89	175,24	218,36	49,80	2,34	16,23	3,72	19,45
310004	Vale do Mucuri	98	164,25	176,36	223,64	58,49	3,57	18,02	4,06	19,61
310005	Triângulo Mineiro	192	194,11	198,61	292,97	66,44	7,44	23,03	4,77	20,76
310006	Central Mineira	55	190,02	191,64	199,91	63,34	3,38	19,29	4,55	20,26
310007	Metropolitana	527	188,63	192,56	340,25	55,64	5,17	23,21	4,70	19,80
310008	Vale do Rio Doce	303	180,28	183,17	188,15	62,72	3,43	15,88	4,05	20,05
310009	Oeste de Minas	90	204,56	206,15	221,52	78,72	5,92	20,10	4,58	20,50
310010	Sul/Sudoeste de Minas	183	200,12	197,40	281,28	66,16	6,19	21,26	4,82	20,66
310011	Campo das Vertentes	61	201,63	199,92	223,57	75,29	5,44	18,85	4,31	20,00
310012	Zona da Mata	293	188,75	193,28	208,58	64,97	4,78	18,34	4,53	19,79
310013	Minas Gerais	2.458	183,73	186,72	247,64	58,41	4,36	19,21	4,24	19,98

Fontes: Elaboração própria a partir das bases do Sica de 2003, Simave de 2003 e Censo Escolar 2003.

e pouco produto denotam ineficiência; muito produto para pouco insumo evidenciam evento improvável, grande chance de a observação ser *outlier*.

Pode-se dividir a ordenação entre as seis primeiras e seis últimas mesorregiões. O critério para a classificação é a média das eficiências. Logicamente, o uso de médias esconde muitas nuances dentro de cada uma das mesorregiões. A ordenação de eficiência é sensível ao tipo de índice adotado. No que se refere ao método de construção, pode variar de forma significativa.

O *ranking* aqui empregado é a *DEA bootstrap* (*DEA\_c\**) proposto por Simar e Wilson (1998). A eficiência é corrigida por *smooth-bootstrap* de maneira que se obtêm o intervalo de confiança para o índice e uma distribuição mais confiável após a aplicação da técnica. Outra ordenação foi obtida a partir do estimador *jackstrap* de Sampaio de Sousa e Stosic (2005), denominado *DEA\_ss*. O índice Sampaio-Stosic detecta *outliers* e confecciona novos índices após descartar as observações superinfluentes. Após a exclusão dos *outliers*, a mesorregião que mais se

beneficiou foi a Norte de Minas, passando para o primeiro grupo. As escolas *outliers* prejudicavam a comparação da região como um todo.<sup>18</sup>

A terceira ordenação construída foi a de número de escolas eficientes por total de escolas pertencentes à mesorregião. A região metropolitana (RM) é a que possui o maior número de escolas eficientes; entretanto, é também a que possui maior fração da população e mais escolas, por isso o quociente  $n^{\circ}$  *eficientes/total de escolas* apresenta um resultado relativo. Observa-se que a Zona da Mata é a que mais perde posições em relação ao *rank1*. O contrário ocorre com a mesorregião norte de Minas, que possui uma razão elevada de escolas eficientes sobre o total de escolas.

A última coluna da tabela 10 resume o número de escolas eficientes com retornos decrescentes de escala (Decresc), útil para observar a incidência de rendimentos decrescentes em cada mesorregião. O número de eficientes varia de acordo com o uso da DEA-C, DEA-N ou DEA-V. Por construção, há sempre mais

TABELA 10

**Ordenação das mesorregiões para a 4ª série**

Código	Mesorregião	Escolas	DEA_c*	Rank1	DEA_ss	Rank2	quocient	Rank3	Decresc
310009	Oeste de Minas	90	0,654	1	0,757	4	0,033	6	2
310007	Metropolitana	527	0,650	2	0,767	1	0,047	4	6
310011	Campo das Vertentes	61	0,650	3	0,747	3	0,082	1	1
310012	Zona da Mata	293	0,647	4	0,747	6	0,017	11	1
310006	Central Mineira	55	0,646	5	0,744	7	0,055	2	3
310003	Jequitinhonha	191	0,641	6	0,755	5	0,016	12	0
310001	Noroeste de Minas	54	0,629	7	0,731	10	0,019	10	0
310010	Sul/Sudoeste de Minas	183	0,629	8	0,732	9	0,027	7	3
310008	Vale do Rio Doce	303	0,623	9	0,734	8	0,040	5	6
310002	Norte de Minas	411	0,623	10	0,760	2	0,054	3	4
310005	Triângulo Mineiro	192	0,621	11	0,724	11	0,021	8	2
310004	Vale do Mucuri	98	0,607	12	0,701	12	0,020	9	0
310013	Minas Gerais	2.458	0,635		0,748		0,037		28

Fonte: Construção dos índices de eficiência por elaboração própria a partir de variáveis do 1º estágio.

18. Embora qualquer escola eficiente do estado possa servir de comparação para todas as outras, independentemente da região específica, a proximidade geográfica traz semelhança nos *inputs* e *outputs*. Com isso, a localização mais provável de influência de um *outlier* ocorre nas escolas do seu entorno.

escolas eficientes na DEA-V do que na DEA-C – a primeira é mais flexível, aborda qualquer tipo de variação da escala. Neste trabalho, o DEA-C é rejeitado apenas para a 3ª série do ensino médio. No entanto, a construção dos três índices é importante para se captarem as escolas por grupo de retornos de escala. A tabela 11, apresenta essas informações em maiores detalhes.

Os dados descritos na tabela 11 confirmam a intuição inicial: as regiões que possuem melhores níveis de insumo, melhores condições de infra-estrutura e condições socioeconômicas acima da média do estado apresentam mais escolas operando com retornos *decrecentes de escala*, enquanto as regiões mais carentes têm maior tendência a apresentar proporções mais elevadas de escolas com rendimentos *constantes e crescentes*. A RM e a região oeste de Minas, por serem mais populosas e apresentarem as maiores escolas, possuem um dos maiores níveis de escolas com rendimentos decrescentes, ao passo que as regiões do Jequitinhonha e norte de Minas, por serem o oposto das primeiras regiões, possuem muitas escolas com rendimentos crescentes.

Para a 8ª série (tabela 12), temos poucas modificações em relação ao quadro de eficiência apresentado anteriormente. Cabe ressaltar que é a série com maior número de escolas na esfera estadual e que o ganho nas médias das notas é significativo (mais de 50 pontos). A média de eficiência é maior, encontrando-se as escolas

TABELA 11  
**Distribuição das escalas de rendimento por mesorregião**

Código	Mesorregião	Escolas	Constante (%)	Decrescente (%)	Crescente (%)
310001	Noroeste de Minas	54	0,00	98,15	1,85
310002	Norte de Minas	411	5,35	89,54	5,11
310003	Jequitinhonha	191	5,26	89,47	5,26
310004	Vale do Mucuri	98	5,10	91,84	3,06
310005	Triângulo Mineiro	192	2,09	97,38	0,52
310006	Central Mineira	55	3,57	91,07	5,36
310007	Metropolitana	527	3,99	94,87	1,14
310008	Vale do Rio Doce	303	3,63	92,74	3,63
310009	Oeste de Minas	90	2,30	97,70	0,00
310010	Sul/Sudoeste de Minas	183	2,69	96,77	0,54
310011	Campo das Vertentes	61	5,00	91,67	3,33
310012	Zona da Mata	293	4,41	94,58	1,02
310013	Minas Gerais	2.458	3,13	94,34	2,52

Fonte: Construção dos retornos de escala por elaboração própria.



TABELA 12  
**Ordenação das mesorregiões para a 8ª série**

Código	Mesorregião	Escolas	DEA_c*	Rank1	DEA_ss	Rank2	quocient	Rank3	Decresc
310003	Jequitinhonha	163	0,766	1	0,964	2	0,043	4	2
310010	Sul/Sudoeste de Minas	259	0,765	2	0,942	5	0,039	6	8
310011	Campo das Vertentes	70	0,762	3	0,903	8	0,043	5	1
310009	Oeste de Minas	117	0,759	4	0,902	10	0,026	10	1
310012	Zona da Mata	292	0,758	5	0,880	7	0,034	8	4
310006	Central Mineira	62	0,758	6	0,867	3	0,032	9	1
310002	Norte de Minas	347	0,756	7	0,866	1	0,069	1	3
310008	Vale do Rio Doce	307	0,755	8	0,864	6	0,046	3	5
310007	Metropolitana	557	0,753	9	0,858	4	0,050	2	14
310005	Triângulo Mineiro	226	0,743	10	0,853	9	0,035	7	4
310004	Vale do Mucuri	71	0,738	11	0,825	11	0,014	11	0
310001	Noroeste de Minas	60	0,737	12	0,819	12	0,000	12	0
310013	Minas Gerais	2.531	0,755		0,844		0,043		43

Fonte: Construção dos índices de eficiência por elaboração própria a partir de variáveis do 1º estágio.

mais perto da fronteira nesta série. Com exceção de algumas mudanças de posição um pouco mais drásticas para a mesorregião sul/sudoeste que subiu para o segundo lugar no *rank1* e a metropolitana, que caiu para nono, a ordenação da 8ª série corresponde à da 4ª série. Para entender a mudança na posição relativa da metropolitana, é preciso observar o *rank2*. Nessa segunda classificação, a região volta ao grupo dos seis primeiros, porque, na 8ª série, a RM possui relativamente muitas escolas eficientes (*rank3*): são 14 escolas e todas elas na parte decrescente dos ganhos de escala (*decresc*).<sup>19</sup>

Entre os primeiros lugares, a mesorregião de Campo das Vertentes merece novamente destaque, continuou na terceira posição. A oeste de Minas sai da primeira posição, mas fica em 4º lugar, continuando no grupo dos seis primeiros. Sobre as variáveis de Campo das Vertentes e oeste de Minas ocorre um efeito de “transbordamento”, pois os indicadores dessas duas mesorregiões, muitas vezes, não são tão melhores quanto os da metropolitana, mas são mais uniformes. Por isso, apesar de não terem os picos das notas, seus resultados são eficientes. Observe-se que, no *rank2*, essas duas regiões não se saem tão bem, isso se deve ao fato de a ordenação *DEA\_ss* ser menos suave.

19. A correção do *bootstrap* confere um intervalo maior para os eficientes trazendo sua correção para baixo. O modelo Sampaio-Stosic (*DEA\_ss*) mostra, também, uma queda de posições para essa região mas não tão grande quanto a do *bootstrap*.

Por outro lado, o desempenho é mais fraco, em termos de eficiência, na mesorregião do Vale do Mucuri, novamente em 11º lugar, no noroeste de Minas e no Triângulo Mineiro. As duas primeiras apresentam resultados precários em termos de produtos; a última possui um nível de insumos elevado, mas resultados fracos, ainda não condizentes com sua quantidade de insumos.

Na 3ª série, conforme tabela 13, há algumas mudanças mais importantes. A primeira diz respeito à alteração do índice DEA, o principal a ser utilizado. Sob o teste de rendimento de escala de Banker (1993) e Simar e Wilson (2002), rejeita-se a hipótese de rendimentos constantes para a função de produção educacional, indicando que a análise ideal é feita pelo índice de retornos variáveis DEA-V. Esse fenômeno está relacionado à segunda mudança principal para a 3ª série: o sistema de ensino médio é bem mais complexo que o do fundamental, uma vez que a média de eficiência DEA-C é extremamente baixa (0,534). Portanto, conjectura-se necessária uma abordagem mais ampla. O uso da DEA-V satisfaz em parte esse anseio. Por outro lado, o uso de mais indicadores de produtos e insumos seria bem-vindo para tratar dessa complexidade maior, no entanto, reforçamos que a base da 3ª série possui menos escolas (1,419) e é mais heterogênea do que as anteriores.

Um terceiro ponto que distingue a 3ª série é a apresentação de uma elevada aderência entre escolas declaradas eficientes pela Secretaria de Estado de Educação

TABELA 13  
Ordenação das mesorregiões para a 3ª série

Código	Mesorregião	Escolas	DEA_v*	Rank1	DEA_ss	Rank2	quocient	Rank3	Decresc
310006	Central Mineira	39	0,897	1	0,920	1	0,077	2	2
310010	Sul/Sudoeste de Minas	169	0,895	2	0,914	3	0,047	7	8
310009	Oeste de Minas	74	0,892	3	0,914	4	0,014	11	1
310011	Campo das Vertentes	53	0,889	4	0,906	8	0,038	8	2
310012	Zona da Mata	184	0,888	5	0,910	7	0,049	5	8
310005	Triângulo Mineiro	136	0,883	6	0,902	9	0,051	4	6
310007	Metropolitana	352	0,878	7	0,912	5	0,063	3	16
310003	Jequitinhonha	62	0,877	8	0,912	6	0,048	6	2
310008	Vale do Rio Doce	146	0,873	9	0,896	11	0,027	10	3
310004	Vale do Mucuri	32	0,868	10	0,898	10	0,031	9	1
310002	Norte de Minas	138	0,867	11	0,917	2	0,101	1	3
310001	Noroeste de Minas	34	0,864	12	0,885	12	0,000	12	0
310013	Minas Gerais	1.419	0,903		0,903		0,052		52

Fonte: Construção dos índices de eficiência por elaboração própria a partir de variáveis do 1º estágio.

de Minas Gerais (SEE/MG) e a identificação de eficiência pelo índice Sampaio-Stosic. A correlação de uma ordenação com a outra é de 0,17, considerada elevada, já que, das 250 eficientes da lista DEA, mais de 20% (68 escolas) eram também consideradas eficientes pela secretaria, resultado obtido por outro método.

Quanto à ordenação, percebe-se semelhança com a da 8ª série. Em parte, essa evidência é verificada por conta de muitas escolas fornecerem simultaneamente o segundo ciclo do fundamental e o ensino médio, embora não necessariamente a classificação de uma mesma escola em diferentes séries seja idêntica, ou mesmo próxima; depende do comportamento de todas as oito variáveis do modelo. A central mineira se destaca positivamente, pulando para a primeira posição nas duas primeiras ordenações. Sul/sudoeste, Campo das Vertentes e Zona da Mata continuam se situando bem. O Triângulo Mineiro é a novidade, aparecendo pela primeira vez no grupo dos seis primeiros, e a RM e o Jequitinhonha não estão mais no primeiro grupo, mas ocupam a 5ª e a 6ª posições no *rank2*, enquanto o Triângulo Mineiro cai para o 9º lugar nessa ordenação.

As regiões de fraco desempenho se repetem: Vale do Rio Doce e Mucuri, norte e noroeste de Minas. Para o caso específico do norte de Minas cabe uma interessante ressalva: em todas as tabelas de classificação, apesar de a mesorregião estar situada entre as últimas posições para o *rank1*, ela se sai bem nos *ranks* 2 e 3. Isso ocorre porque há algumas escolas boas na região norte mas, apesar disso, a grande maioria é de escolas pobres, com fraco desempenho.

## 5.2 Segundo estágio

No estágio anterior, calculam-se os índices de eficiência levando-se em conta três variáveis de *output* mais cinco variáveis de insumos. No segundo estágio, é realizada uma regressão dos índices de eficiência obtidos para cada escola, considerando-se as variáveis das condições físicas e de localização e as variáveis de dotação. Essa etapa permite identificar as variáveis que afetam o nível de eficiência, porém, não estão sob o controle da gerência das escolas.

Foram realizadas seis regressões pelo modelo de dois estágios, o índice de eficiência empregado precisou ser alterado para o inverso do índice anterior ( $1/\delta_i$ ). O intervalo para essa transformação se situa entre de 1 e  $\infty$ . O valor 1 continua afinando a eficiência, porém, a escala percorre sentido inverso: quanto maior o índice, menor a eficiência. A interpretação para os coeficientes das regressões ocorre, portanto, da seguinte maneira: se o parâmetro é positivo o resultado se afasta da eficiência, se negativo, aproxima-se. As regressões (1) e (2) na tabela 14 são estimações das normais truncadas obtidas por máxima verossimilhança (MV), já que as variáveis regredidas possuem o limite inferior de valor igual a 1. As duas primeiras regressões seguem o algoritmo#1 proposto em Simar e Wilson (2007) e utilizado também em Afonso e Aubyn (2005).

A primeira regressão utiliza as informações do questionário Simave como variáveis primárias. As informações se referem à turma de alunos por escolas e, na regressão, estão devidamente controladas para escolas urbanas ou rurais. Na segunda regressão, as variáveis são os indicadores de dotação presentes nos dados do censo escolar, onde também há correção urbano-rural. Essas duas primeiras regressões servem de guia para as quatro demais, pois são mais rigorosas na construção dos intervalos dos testes de hipóteses e rejeitam a hipótese nula com menor frequência. As regressões (3) e (4), na tabela 14, são regressões normais-truncadas por MV – no entanto, sem seguir os passos do algoritmo Simar e Wilson (2007). Nas regressões (5) e (6) temos o modelo *tobit* na forma usual de regressão de dois estágios. A diferença na regressão (6) é a utilização da eficiência não-corrigida no regressando.<sup>20</sup> Na 4ª série, com exceção de *ntrab\_fora*, todas as variáveis significativas apresentam a mesma direção das regressões (1) e (2).

Os sinais das variáveis significativas se apresentam na direção esperada. A surpresa é *urbana*, o coeficiente das regressões (1), (3), (5) e (6) é positivo, indicando que se a escola é urbana, essa tende a ser uma escola menos eficiente em probabilidade.<sup>21</sup> Esse coeficiente pode ser interpretado em termos de eficiência técnica. As escolas rurais possuem, em média, menos recursos do que as localizadas nas cidades, entretanto, muitas delas não são seriamente reprimidas em seus resultados por conta disso. No quesito das notas, as rurais não perdem tanto quanto seria de se esperar, quando se leva em conta que possuem relativamente menos insumos, ou, por outro lado, as urbanas não estão fazendo jus às melhores condições e não se saem relativamente melhor do que as rurais. A última interpretação ganha reforço, porque, ao se controlar as regressões pelas condições de infra-estrutura, menores no meio rural (regressões 2 e 4), o coeficiente negativo passa a indicar que as escolas urbanas apresentam melhor desempenho.

Outra variável de destaque é a escolaridade da responsável feminina (*esc\_resp\_fem*). É, em termos absolutos, o maior coeficiente e confirma toda a literatura que aponta que as condições familiares importam para o rendimento das crianças. A análise de eficiência indica que, das escolas declaradas eficientes, há maior proporção de mães de alunos que possuem acima do ensino fundamental completo. As escolas eficientes absorvem filhos de pais com maior escolaridade, ou, pais mais instruídos optam, com maior frequência, por matricular seus filhos em escolas eficientes.

Ter livros em casa (*livros*) aponta melhores condições socioeconômicas dos alunos. As eficientes possuem alunos com maior acervo em sua biblioteca particular.

20. Mesmo usando uma regressão em sua versão mais simples, os dados parecem apontar para a direção esperada.

21. Se nos deslocarmos do meio rural para o urbano, maior será a proporção de escolas ineficientes nesse meio em relação às rurais.

Embora não significativa em (1), o sinal dessa variável é esperado e assinala-se como significativa a 1% em todas as outras regressões, como se verifica, também, para a 3ª série. Não se pode dizer o mesmo para o caso dos alunos que não trabalham fora (*ntrab\_fora*); apesar de, nas demais regressões, o coeficiente se apresentar significativo e na direção esperada pela intuição, as informações não se confirmam na primeira regressão.

A indicação sobre se os alunos na sala de aula lêem (*ler\_livro\_inf*) é também um interessante ponto afirmativo de eficiência. As escolas eficientes podem ser aquelas que conseguem estimular mais a leitura, ou então, aquelas que atraem alunos com maior interesse por leitura. De igual maneira para a realização do dever de matemática (*devermat*).

O coeficiente de *nreprovado* é positivo, o que indica que as escolas que menos reprovam são as ineficientes. Tal resultado reflete a relação custo/benefício entre fluxo escolar e qualidade do ensino. Os programas de progressão continuada reduzem a taxa de reprovação, mas, ao mesmo tempo, afetam negativamente o desempenho dos alunos nas disciplinas, uma vez que alunos com ainda pouca habilidade em determinado conteúdo são introduzidos em novos conteúdos cujo aprendizado depende do conteúdo anterior.

Quanto a algumas variáveis de infra-estrutura, os resultados parecem corroborar interpretações de que muitas das práticas pedagógicas empregadas nas escolas não conseguem aproveitar muito bem os recursos disponíveis (OLIVEIRA; SCHWARTZMAN, 2002). O número de aparelhos de TV, *apvideo*, evidencia essa questão: várias das escolas ineficientes possuem televisão, em Minas Gerais. TV não é um recurso escasso para as escolas, tanto que um número considerável de escolas da 4ª série possui dez ou mais televisores (120 escolas).<sup>22</sup> O resultado aponta que as escolas eficientes não têm tantos televisores. O mesmo ocorre para *multimid*, ou seja, presença de biblioteca, videoteca, sala de TV e vídeo, sala de leitura e laboratório de informática não contribuem para o desempenho escolar.

Todavia, observa-se o contrário para o caso dos computadores (*comp\_imp*). Na 4ª série são 1.024 escolas sem computador, destas, 82 (8%) foram declaradas eficientes; das 629 com 10 ou mais computadores, 13 (18,7%) eram eficientes. Os computadores nas escolas contribuem para sua eficiência. A explicação pode ser a seguinte: a técnica pedagógica vigente consegue utilizar o computador como uma ferramenta na melhoria do ensino, enquanto os demais recursos parecem não estar contribuindo tanto, o que pode sugerir subutilização ou uma utilização não proveitosa para as notas dos alunos em português e matemática.

22. Treze das 49 escolas sem nenhum televisor foram definidas como eficientes, e parte, por se saírem bem, mesmo com a completa ausência de um dos recursos, o que se revela em um baixo *VCM*, na maioria dos casos, vêm acompanhado de baixa infra-estrutura. No entanto, 7 daquelas 13 são na verdade *outliers*.

TABELA 14

**Resultados das regressões de dois estágios para a 4ª série**

	Regressão (1)	Regressão (2)	Regressão (3)	Regressão (4)	Regressão (5) ( <i>tobit1</i> )	Regressão (6) ( <i>tobit2</i> )
constante	1,238*** (0,05)	0,253** (0,12)	1,847*** (0,05)	1,554 (0,03)	1,841*** (0,05)	1,584*** (0,05)
urbana	0,136*** (0,05)	-0,216*** (0,01)	0,084*** (0,01)	-0,075*** (0,02)	0,081*** (0,01)	0,164*** (0,01)
fnde	0,071 (0,09)		0,035*** (0,01)		0,033*** (0,01)	0,041*** (0,01)
mere_esc	0,001 (0,10)		0,002 (0,02)		0,002 (0,02)	0,074*** (0,02)
cor	0,025 (0,13)		-0,013 (0,03)		-0,012 (0,03)	-0,018 (0,03)
esc_resp_fem	-0,385** (0,16)		-0,145*** (0,04)		-0,139*** (0,03)	-0,100*** (0,03)
livros	-0,189 (0,18)		-0,127*** (0,04)		-0,121*** (0,04)	-0,125*** (0,04)
computador	-0,143 (0,12)		-0,088* (0,05)		-0,083* (0,04)	-0,170*** (0,05)
Ler_livro_inf	-0,041 (0,10)		-0,104*** (0,03)		-0,100*** (0,03)	-0,090 (0,03)
Ler_jornais	-0,088 (0,13)		-0,056** (0,03)		-0,054** (0,03)	-0,052** (0,03)
ntrab_fora	0,092 (0,12)		-0,135*** (0,04)		-0,130*** (0,04)	-0,122*** (0,04)
ntrab_domestico	0,053 (0,11)		-0,006 (0,03)		-0,005 (0,03)	-0,005 (0,03)
devermat	-0,087 (0,11)		-0,102*** (0,03)		-0,098*** (0,03)	-0,103*** (0,03)
nreprovado	0,137*** (0,00)		0,060** (0,03)		0,058** (0,03)	0,028 (0,03)
apvideo		0,120*** (0,01)		0,045*** (0,00)		

(continua)

(continuação)

	Regressão (1)	Regressão (2)	Regressão (3)	Regressão (4)	Regressão (5) (tobit1)	Regressão (6) (tobit2)
comp_imp		-0,054 (0,04)		-0,016*** (0,00)		
multimid		0,223 (0,15)		0,064*** (0,00)		
sani_den		-0,098 (0,10)		-0,023 (0,02)		
ener_pub				-0,101*** (0,03)		
agua_pub				0,019 (0,02)		
esg_pub		-0,037 (0,12)		-0,015 (0,01)		
lixo_col		0,052*** (0,00)		0,008 (0,01)		
sigma_hat	0,599	0,899	0,226	0,191	0,222	0,223

Fonte: Elaboração própria a partir das variáveis escolhidas para o 2º estágio.

\* Significativo a 10%; \*\* significativo a 5%; \*\*\* significativo a 1%.

Para o caso da 8ª série (tabela 15), duas novas variáveis passaram a ser estatisticamente significativas: merenda escolar (*mere\_esc*) e cor. A primeira mostra que as escolas que possuem merenda escolar são, em média, menos eficientes. Essa variável indica uma relativa carência das escolas, poucas dispensam a merenda escolar (menos de 5%). A segunda variável se apresenta negativa, indicando que quanto maior a proporção de alunos brancos, maior a eficiência, fator explicado também pela disparidade das condições familiares. Destacamos, novamente, a importância da escolaridade da responsável feminina (*esc\_resp\_fem*); de se ter minibiблиотека em casa (*livros*); da maior proporção de alunos que lê com frequência (*ler\_livro*); de não se trabalhar fora (*ntrab\_fora*); e de realizar o dever de casa (*dever\_mat*). A não-reprovação (*nreprovado*) continua indicando escolas mais fracas e, novamente, significativa apenas na regressão (1). A análise de 'apvideo' é ratificada assim como a importância do computador é reforçada para a 8ª série, sendo significativa a 1% no modelo mais rigoroso (2). As demais características seguem as considerações feitas para 4ª série.

No modelo de dois estágios da 3ª série, cujos resultados são apresentados na tabela 16, o índice de eficiência empregado no *regressando* é o DEA-V. Poucas variáveis das duas primeiras regressões conseguiram atingir a convergência da MV.

TABELA 15

**Resultados das regressões de dois estágios para a 8ª série**

	Regressão (1)	Regressão (2)	Regressão (3)	Regressão (4)	Regressão (5) ( <i>tobit1</i> )	Regressão (6) ( <i>tobit2</i> )
constante	0,444*** (0,08)	1,202*** (0,01)	1,396*** (0,03)	1,268*** (0,04)	1,397 (0,03)	1,263*** (0,03)
urbana	0,370*** (0,05)	0,019*** (0,00)	0,097*** (0,01)	0,006 (0,01)	0,094*** (0,01)	0,126*** (0,01)
fnde	0,034 (0,11)		0,009 (0,01)		0,009 (0,01)	0,012 (0,01)
mere_esc	0,168 (0,14)		0,044*** (0,01)		0,043*** (0,01)	0,082*** (0,01)
cor	-0,164 (0,21)		-0,076*** (0,02)		-0,074*** (0,02)	-0,045*** (0,02)
esc_resp_fem	-0,414* (0,21)		-0,099*** (0,03)		-0,096*** (0,03)	-0,098*** (0,03)
livros	-0,191 (0,30)		-0,074*** (0,03)		-0,073*** (0,03)	-0,057** (0,03)
internet	0,093 (0,15)		0,025 (0,04)		0,025 (0,03)	-0,005 (0,04)
ler_livro	-0,135 (0,14)		-0,111*** (0,02)		-0,108*** (0,02)	-0,095*** (0,02)
ler_jornais			0,016 (0,02)		0,015 (0,02)	-0,011 (0,02)
ntrab_fora	-0,062 (0,14)		-0,049*** (0,02)		-0,048*** (0,02)	-0,043** (0,02)
ntrab_domestico			-0,012 (0,03)		-0,012 (0,03)	-0,005 (0,03)
devermat	-0,014 (0,16)		-0,044*** (0,02)		-0,043*** (0,02)	-0,054*** (0,02)
nreprovado	0,121*** (0,00)		-0,001 (0,02)		-0,001 (0,02)	-0,012 (0,02)
apvideo		0,011*** (0,00)		0,011*** (0,00)		
comp_imp		-0,008*** (0,00)		-0,008*** (0,00)		
multimid		0,027*** (0,01)		0,027*** (0,00)		
sani_den				0,021* (0,01)		
ener_pub				-0,098*** (0,04)		
agua_pub				0,034*** (0,01)		
esg_pub		0,032*** (0,01)		0,028*** (0,01)		
lixo_col		0,048*** (0,00)		0,042*** (0,01)		
sigma_hat	0,525	0,104	0,112	0,104	0,111	0,123

Fonte: Elaboração própria a partir das variáveis escolhidas para o 2º estágio.

\* Significativo a 10%; \*\* significativo a 5%; \*\*\* significativo a 1%.



No modelo (1) temos somente *urbana*, que deixou de ser significativa (quase totalidade das escolas do ensino médio são urbanas) e *esc\_resp\_fem* que continua significativa a 1%, indicando que a maior escolaridade dos pais contribui para a eficiência dos alunos e da escola. Para os modelos menos rigorosos como o *tobit*, *urbana* volta a ser significativa. *Cor* apresenta os mesmos resultados da 8ª série e se confirma novamente a importância da leitura de livros. Os aparelhos de vídeo e TV confirmam as tendências anteriores. Computadores deixam de ser significativos em (2) e (4).

TABELA 16

**Resultados das regressões de dois estágios para a 3ª série**

	Regressão (1)	Regressão (2)	Regressão (3)	Regressão (4)	Regressão (5) ( <i>tobit1</i> )	Regressão (6) ( <i>tobit2</i> )
constante	0,829*** (0,13)	-2,147 (0,01)	1,192*** (0,02)	1,109*** (0,06)	1,190*** (0,02)	1,127*** (0,02)
urbana	0,111 (0,12)		0,038** (0,02)	-0,010 (0,02)	0,034** (0,01)	0,073*** (0,02)
fnde			0,000 (0,00)		-0,000 (0,00)	0,001 (0,00)
mere_esc			0,012* (0,01)		0,010* (0,01)	0,021*** (0,01)
cor			-0,076*** (0,01)		-0,071*** (0,01)	-0,059*** (0,01)
esc_resp_fem	-0,531*** (0,00)		-0,071*** (0,02)		-0,064*** (0,02)	-0,088*** (0,02)
livros			-0,024 (0,02)		-0,021 (0,02)	-0,013 (0,02)
internet			0,016 (0,02)		0,015 (0,02)	0,004 (0,02)
Ler_livro			-0,053*** (0,01)		-0,049*** (0,01)	-0,047*** (0,01)
Ler_jornais			-0,012 (0,01)		-0,011 (0,01)	-0,013 (0,01)
ntrab_fora			-0,024* (0,01)		-0,021* (0,01)	-0,039*** (0,01)
ntrab_domestico			-0,012 (0,02)		-0,011 (0,02)	-0,005 (0,02)

(continua)

(continuação)

	Regressão (1)	Regressão (2)	Regressão (3)	Regressão (4)	Regressão (5) ( <i>tobit1</i> )	Regressão (6) ( <i>tobit2</i> )
devermat			-0,013 (0,01)		-0,011 (0,01)	-0,022* (0,01)
nreprovado			-0,005 (0,01)		-0,004 (0,01)	-0,008 (0,01)
apvideo				0,004*** (0,00)		
comp_imp		0,012 (0,10)		0,000 (0,00)		
multimid		0,202 (0,24)		0,007*** (0,00)		
sani_den		0,069*** (0,01)		0,001 (0,01)		
ener_pub				-0,076 (0,06)		
agua_pub				0,023** (0,01)		
esg_pub				0,000 (0,00)		
lixo_col				0,014** (0,01)		
noite			0,029*** (0,01)	0,056*** (0,01)	0,027*** (0,00)	0,023*** (0,00)
sigma_hat	0,233	0,633	0,057	0,057	0,054	0,059

Fonte: Elaboração própria a partir das variáveis escolhidas para o 2º estágio.

\* Significativo a 10%; \*\* significativo a 5%; \*\*\* significativo a 1%.

A nova variável acrescentada é *noite*, que indica a percentagem de alunos que estudam à noite em determinada escola. Ao contrário das séries anteriores, para a 3ª série, essa variável é uma medição importante já que grande parte dos alunos desse nível (70%) estuda nesse turno. As escolas que ofertam aulas no horário noturno estão mais propensas à ineficiência, o que parece refletir muito as condições dos alunos que estudam à noite, com menor disponibilidade de tempo devido à dupla jornada e à precariedade das condições sociais.

## 6 CONSIDERAÇÕES FINAIS

Os resultados obtidos neste trabalho mostram, de uma maneira geral, que as escolas selecionadas como eficientes provêm ensino de melhor qualidade a um custo relativamente menor do que as outras escolas. Alguns fatores, como a infra-estrutura, desempenham papel importante para propiciar tal resultado. Ademais, enquanto a presença de computador nas escolas contribui para incrementar o indicador de eficiência, equipamentos subutilizados como televisores e outros aparelhos correlatos (parabólica e videocassete), além de alguns outros recursos como bibliotecas, videotecas e laboratórios de ciências, não ocasionam o mesmo resultado.

Outros indicadores importantes de eficiência da escola são as variáveis referentes às condições familiares dos alunos. Citam-se, entre elas, a escolaridade da mãe, o estímulo à leitura, ter livros em casa, não exercer trabalho remunerado e cultivar o hábito do estudo como fatores que contribuem para o melhor desempenho nas provas de matemática e português. Na verdade, tanto a família quanto a escola parecem ser responsáveis pelo bom resultado nas provas. A escola possui um papel importante no estímulo à leitura e no hábito de estudo, assim como o de fornecer o ambiente mínimo para que os alunos se desempenhem bem. A família complementa a escola, já que as mães de maior escolaridade e as famílias com maiores recursos educacionais são, em geral, mais seletivas na escolha da instituição. No sistema público, a escolha não é totalmente livre, está atrelada à proximidade da residência familiar. Dessa forma, as escolas eficientes captam, também, um efeito da vizinhança – as melhores condições dos alunos facilitam a trajetória do aproveitamento dos recursos.

A análise por mesorregiões destaca que as melhores áreas de eficiência são as regiões centrais de Minas: Metropolitana, Campo das Vertentes, oeste de Minas, central e Zona da Mata. Entre as regiões pobres, a do Jequitinhonha é um exemplo positivo e, em algumas análises, também o norte de Minas, por obterem bons resultados mesmo com poucos recursos.

Por fim, cabe ressaltar que esse trabalho pioneiro sobre o tema eficiência na provisão de serviço público em educação no Brasil é mais uma contribuição para a formulação de políticas públicas. A experiência mineira, refletindo as disparidades socioeconômicas do país, evidencia que se faz mister conjugar dois aspectos: maior destinação de assessoria técnica (em alguns casos também financeira) aos municípios localizados em regiões mais pobres, nos quais apresentam a vantagem de operar com rendimentos constantes ou crescentes, e a realocação de recursos naqueles onde a ineficiência não é fruto da escassez de insumos, mas sim de seu mau aproveitamento (por exemplo, nos municípios mais desenvolvidos). Para o primeiro conjunto, o acréscimo marginal nos insumos representará ganhos proporcionais (ou mais que proporcionais) nos resultados, ao passo que, no segundo conjunto, a readequação pode torná-los mais eficientes.

## ABSTRACT

This paper develops the semi-parametrical two-stage Data Envelopment Analysis to detect the efficiency frontier of public Brazilian Minas Gerais state schools. The first stage computes the Data Envelopment Analysis (DEA) – Análise Envolvória de Dados – efficiency, the second one inputs the DEA results to a truncated-regression with background and endowment variables. The new data basis of school proficiency, Sistema Mineiro de Avaliação da Educação Pública (Simave), and cost-measurement, Sistema Informacional de Custo Aluno (Sica), either the Educational Census of Education Ministry 2003, permitted this new application of efficiency measurement for schools. The results shows that input-complementarity *in* and *out* school contribute to greater performance. Best located schools, where the educational inputs are better, are more efficient in probability. However, there are some good examples of efficient schools emerging from poor regions and, in general terms, all stated educational results can improve considerably if the schools get to improve their efficiency.

## REFERÊNCIAS

- AFONSO, A.; AUBYN, M. ST. *Cross-country efficiency of secondary education provision: a semi-parametric analysis with non-discretionary inputs*. Frankfurt: European Central Bank, 2005. 39 p. (Working paper, n. 494).
- AIGNER, D. J.; LOVELL, C. A. K.; SCHMIDT, P. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, v. 6, n.1, p. 21-37, 1977.
- AIGNER, D. J.; CHU, S. F. On estimating the industry production function. *American Economic Review*, v. 58, n. 4, p. 826-839, 1968.
- ALI, A. I.; SEIFORD, L. M. Translation invariance in data envelopment analysis. *Operations Research Letters*, v. 9, n. 6, p. 403-405, Nov. 1990.
- BANKER, R. Maximum likelihood, consistency and data envelopment analysis: a statistical foundation. *Management Science*, v. 39, n. 10, p. 1.265-1.273, Oct. 1993.
- BANKER, R.; CHARNES, A.; COOPER, W. W. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, v. 30, n. 9, p. 1.078-1.092, Sep. 1984.
- BDMG. *Minas Gerais do Século XXI*. Banco de Desenvolvimento de Minas Gerais. Belo Horizonte: Rona Editora, v. 1, cap. 3, 2002.
- BRASIL. Ministério da Educação e Cultura. Instituto Nacional de Estudos e Pesquisas Educacionais. *Censo Educacional 2003: avaliação*. Brasília: Inep, 2003.
- CAZALS, C.; FLORENS, J.; SIMAR, L. Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, v. 106, n. 1, p. 1-25, 2002.
- CHARNES A.; COOPER, W. W.; RHODES, E. Measuring the efficiency of decision making units. *European Journal of Operational Research*, v. 2, n. 6, p. 429-444, 1978.
- \_\_\_\_\_. Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through. *Management Science*, v. 27, n. 6, p. 668-697, June 1981.
- COELLI, T.; RAO, D. S. P.; BATTESE, G. E. *An introduction to efficiency and productivity analysis*. Boston: Kluwer Academic, 1998. 275 p.

- COLEMAN, J. S.; CAMPBELL, E. Q.; HOBSON, C. J.; MCPARTLAND, J.; MOOD, A. M.; WEINFELD, F. D.; YORK, R. L. *Equality of educational opportunity*. Washington, D.C.: U.S. Government Printing Office, 1966.
- DEBREU, G. The coefficient of resource utilization. *Econometrica*, v. 19, n. 3, p. 273-292, July 1951.
- FAÇANHA, L.; MARINHO, A. Instituições federais de ensino superior: modelos de financiamento e o incentivo à eficiência. *Revista Brasileira de Economia*, v. 53, n. 3, p. 357-386, jul./set. 1999.
- \_\_\_\_\_. *Instituições de ensino superior governamentais e particulares: avaliação comparativa de eficiência*. Rio de Janeiro: Ipea, 2001 (Texto para discussão, n. 813).
- FARE, R.; GROSSKOPF, S.; WEBER, W. Measuring school district performance. *Public Finance Quarterly*, v. 17, n. 4, p. 409-420, Oct. 1989.
- FARRELL, M. J. The measurement of productive efficiency. *Journal of the Royal Statistical Society, Series A*, v. 120, n. 3, p. 253-290, 1957.
- FARIA, A. F.; JANUZZI, P. M. *Eficiência dos gastos municipais em saúde e educação: uma investigação através da análise envoltória no Estado do Rio de Janeiro*. Rio de Janeiro: Ipea, 2006 (Prêmio Ipea-Caixa 2006; Concurso de Monografias. Tema 1: Eficiência e Efetividade do Estado no Brasil).
- FUNDAÇÃO JOÃO PINHEIRO. *Atlas de Desenvolvimento Humano*. Pnud e Ipea. 2000. Disponível em: <[http://www.fjp.gov.br/produtos/cees/idh/atlas\\_idh.php](http://www.fjp.gov.br/produtos/cees/idh/atlas_idh.php)>. Acesso em: 03 de julho de 2006.
- GASPARINI, C. E.; RAMOS, F. S. Efetividade e eficiência no ensino médio brasileiro. *Economia Aplicada*, v. 7, n. 2, p. 389-411, 2003.
- GIBELS, I.; MAMMEN, E.; PARK, B. U.; SIMAR, L. On estimation of monotone and concave frontier functions. *Journal of the American Statistical Association*, v. 94, n. 445, p. 220-228, Mar. 1999.
- GREENE, W. The econometric approach to efficiency analysis. In: FRIED, H.; LOVELL, C. A. K.; SCHIMDT, P. (Eds.). *The measurement of productive efficiency: techniques and applications*. Oxford: Oxford University, p. 68-119, 1993.
- HANUSHEK, E. A.; LUQUE, J. *Efficiency and equity in schools around the world*. Cambridge, Mass.: National Bureau of Economic Research, 2002. 42 p. (Working papers, n. 8.949).
- HANUSHEK, E. A. The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature*, v. 24, n. 3, p. 1.141-1.177, Sep. 1986.
- KNEIP, A.; SIMAR, L.; WILSON, P. W. *Asymptotics for DEA estimates in nonparametric frontier models*. 2003. 37 p. (Technical report, n. 0323).
- MARINHO, A.; RESENDE, M.; FAÇANHA, L. O. Brazilian federal universities: relative efficiency evaluation and data envelopment analysis. *Revista Brasileira de Economia*, v. 51, n. 4, p. 489-508, Oct./Dec. 1997.
- MCCARTY, T. A.; YAISAWARNG, S. Technical efficiency in New Jersey School Districts. In: FRIED, H. O.; LOVELL, C. A. K.; SCHMIDT, S. S. (Eds.). *The measurement of productive efficiency: techniques and applications*. Oxford: Oxford University, p. 271-287, 1993.
- OLIVEIRA, J. B. A.; SCHWARTZMAN, S. *A escola vista por dentro*. Belo Horizonte: Alfa Educativa, 2002.
- SAMPAIO DE SOUSA, M. C.; STOSIC, B. Technical efficiency of the Brazilian municipalities: correcting nonparametric frontier measurements for outliers. *Journal of Productivity Analysis*, v. 24, n. 2, p. 157-181, 2005.

SAMPAIO DE SOUSA, M. C.; RAMOS, F. S. Eficiência técnica e retornos de escala na produção de serviços públicos municipais: o caso do Nordeste e do Sudeste brasileiros. *Revista Brasileira de Economia*, Rio de Janeiro, v. 53, n. 4, p. 433-461, 1999.

SHEPPARD, R. W. *Theory of cost and production function*. Princeton, NJ: Princeton University, 1970. 308p.

SIMAR, L. Detecting outliers in frontier models: a simple approach. *Journal of Productivity Analysis*, v. 20, n. 3, p. 391-424, 2003.

SIMAR, L.; WILSON, P. W. Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. *Management Science*, v. 44, n. 1, p. 46-61, Jan. 1998.

\_\_\_\_\_. Non-parametric tests of returns to scale. *European Journal of Operational Research*, v. 139, n. 1, p. 115-132, 2002.

\_\_\_\_\_. Estimation and inference in two-stage, semi-parametric models of production process. *Journal of Econometrics*, v. 136, n. 1, p. 31-64, 2007.

SOARES, T.; PEREIRA, D. Estudo de critérios de adequação para modelos da teoria da resposta ao item (TRI) aplicado ao caso do ensino fundamental da microrregião de Juiz de Fora em 1999. *Educação em Foco*, v. 6, n. 2, p. 91-108, 2002.

WILSON, P. W. *Efficiency in education production among PISA countries with emphasis on transitioning economies*. Texas: University of Texas, 2005. 40 p. Disponível em: <[http://upeg.eerc.kiev.ua/conf/papers/Wilson\\_Paul.pdf](http://upeg.eerc.kiev.ua/conf/papers/Wilson_Paul.pdf)>. Acesso em: 23 de janeiro de 2007.

(Originais recebidos em agosto de 2007. Revistos em outubro de 2007.)