

Economias de Escala na Oferta de Serviços Públicos de Saúde: Um Estudo para os Municípios Paulistas

Enlinson Mattos

Fundação Getúlio Vargas (EESP/FGV), Brasil

Fabiana Rocha

Universidade de São Paulo (FEA/USP), Brasil

Lucas Novaes

Doutorando, University of California, Berkeley, EUA

Paulo Arvate

Fundação Getúlio Vargas (EAESP/FGV), Brasil

Verônica Orellano

Fundação Getúlio Vargas (EESP/FGV), Brasil

Resumo

Este artigo avalia os efeitos das variáveis de escala na oferta de serviços públicos de saúde dos municípios paulistas, com o objetivo de investigar os possíveis efeitos do processo de descentralização dos serviços de saúde a partir da consolidação do Sistema Único de Saúde (SUS). Foram construídos indicadores de eficiência dos gastos municipais em saúde pela técnica de *Free Disposable Hull*, tendo sido estimados os efeitos das variáveis de escala não só sobre a eficiência, mas também diretamente sobre os indicadores de desempenho e sobre os gastos *per capita*. Conclui-se que os municípios menores, além de terem maior gasto *per capita* em saúde, oferecem piores condições de acesso ao sistema. Além disso, pequenos municípios não oferecem melhores serviços de prevenção de doenças infecciosas, apesar de terem, a princípio, melhores condições de conhecer as necessidades dos cidadãos. Em consonância com o trabalho de Sampaio de Souza et alii (2005), conclui-se que a excessiva descentralização na gestão pode levar à perda de eficiência, aumentando gastos sem necessariamente melhorar a qualidade.

Palavras-chave: Gastos em Saúde, Economias de Escala, FDH

Classificação JEL: H51, H75, I11

Abstract

Revista Economia

Maio/Agosto 2009

This article evaluates the effect of scale variables in the provision of health services by São Paulo state municipalities in order to investigate possible effects of the decentralization policy that characterize the Sistema Unico de Saude (Public Health System authority-SUS) guidelines. We construct efficiency scores of public health expenditures using the Free Disposable Hull (FDH) technique, and estimate scale effects not only for these scores, but also directly for performance indicators of *per capita* expenditures. We find that smaller cities, aside larger *per capita* public spending, provide inferior access to the health system. Moreover, small municipalities do not possess better infectious diseases prevention system, albeit their assumed closer contact to their citizens' needs. In agreement with de Sampaio de Souza et alii (2005) findings, we conclude that in the health sector excessive managerial decentralization can lead to a loss of efficiency, increasing expenditures with no detectable quality improvement.

1. Introdução

O setor de saúde é um dos mais complexos setores da sociedade. Além de envolver complicadas relações econômicas, mistura diferentes interesses e instituições na sua coordenação. Embora existam opiniões distintas sobre a capacidade e a necessidade do governo intervir na economia, poucos negligenciariam cuidados médicos aos desprovidos de meios. O estado brasileiro não é exceção, sendo a oferta universal de bens de saúde uma de suas obrigações.

O sistema de saúde brasileiro passa por transformações importantes desde a década de sessenta, mas muito do atual arranjo foi desenhado pela Constituição de 1988. O modelo possui as seguintes diretrizes: descentralização, atendimento integral (universalização), equidade e participação social. São metas pouco modestas, ainda mais para um país populoso e disperso, territorialmente grande, desigual e com um estado inchado.

Atualmente, 50% dos gastos públicos em saúde são oriundos do governo federal, enquanto a outra metade é dividida entre estados e municípios. Porém, a política de descentralização da administração vem sendo promovida de forma acentuada, com o objetivo de dar aos municípios maior autonomia gestora. Ainda, como uma maneira de exigir a participação efetiva de estados e principalmente municípios no controle da saúde, a Emenda Constitucional n° 29 de 2000 estabelece um percentual mínimo para a aplicação de recursos próprios na saúde. O estado deve destinar ao menos 12% de sua receita e os municípios 15% de sua receita para a área de saúde. Caso a meta não seja atendida, a lei prevê a possibilidade de intervenção da União nesse estado ou município.

Para atribuir a competência de gasto entre as diferentes esferas de Governo, a literatura de federalismo fiscal parte do princípio de que bens e serviços públicos devem ser ofertados pelo nível de Governo mais próximo da população (Oates

* Recebido em setembro de 2008, aprovado em setembro de 2009.
E-mail address: frocha@usp.br

1972; Alesina e Spolaore 1997; Bolton e Roland 1997; Person e Tabellini 2000; Galasso e Ravallion 2001). Por conhecerem as diferentes preferências da população, caberia então aos municípios a responsabilidade pela provisão de bens públicos. A provisão descentralizada poderia aumentar, então, o bem estar social ao aumentar a variedade de bens e serviços oferecidos a partir do entendimento das diferentes preferências e custos locais.

Por outro lado, a oferta pulverizada de alguns tipos de bens e serviços pode levar a significativas deseconomias de escala. Um argumento a favor da centralização seria, então, o de que a produção de alguns serviços exige escala de produção incompatível com a demanda do município. Por exemplo, a existência de muitos municípios com população pequena faz com que não se justifique a oferta por parte destes de todos os níveis de complexidade dos serviços de saúde.¹

Mendes e Sampaio de Souza (2006) estimam a demanda para o gasto público nos municípios brasileiros em 2000 e concluem que as variáveis de escala afetam negativamente os gastos públicos. Conforme ressaltam os autores, a relação inversa entre a população e a demanda *per capita* dos serviços públicos contraria uma tradição em finanças públicas, de acordo com a qual áreas densamente povoadas acompanham-se de despesas públicas *per capita* elevadas – lei de Brecht. Porém, mais recentemente, uma nova literatura propõe uma base teórica para esse resultado (Reiter e Weichenrieder, 1997), a partir da utilização de funções de congestionamento (*crowding functions*). De acordo com essa abordagem, o impacto da população sobre a despesa *per capita* depende da “tecnologia” de produção dos bens públicos (ou privados, disponibilizados pelo setor público). Em virtude da existência de economias de escala, é possível que a despesa *per capita* se relacione inversamente com a população e com a densidade demográfica. Os resultados encontrados por Mendes e Sampaio de Souza (2006) estão de acordo com essas teorias. Para os autores, em razão da proliferação de pequenas cidades, o tamanho sub-ótimo da maioria das municipalidades brasileiras restringe a exploração de economias de escala que caracterizam a produção de serviços públicos. Os autores também encontram uma relação inversa entre *despesa per capita* e densidade demográfica.

Variáveis de escala também aparecem exercendo efeito negativo sobre os gastos públicos no trabalho de Sampaio de Souza et alii (2005), mas esses autores avaliam os gastos públicos priorizando o lado da oferta, ao destacar os aspectos de eficiência na provisão de bens públicos. Para medir a oferta desses bens, levam em consideração serviços administrativos, educacionais, de saúde e de moradia. Concluem que a eficiência na oferta de serviços públicos municipais relaciona-se positivamente com o tamanho do município e a taxa de urbanização.

O objetivo deste artigo é avaliar a presença de eventuais economias de escala na oferta de serviços públicos de saúde, também a partir de uma análise do

¹ Ver ainda Vickrey (1969), *The American Economic Review*, vol. 59(2), Papers and Proceedings of the Eighty-first, (May, 1969), pp. 251–260, para o efeito das deseconomias de escala (*congestion*) para o setor transportes.

lado da oferta. O foco é o setor de saúde, porque espera-se que o processo de descentralização dos serviços de saúde, com a consolidação do Sistema Único de Saúde (SUS), traga com ele todos os benefícios e dificuldades expostos pelas teorias de federalismo fiscal. Por um lado, descentralização pode implicar aumento da qualidade e do acesso aos serviços de saúde. Por outro lado, pode trazer ineficiências devido à perda de escala.

Para realizar essa análise, os municípios paulistas foram ordenados de acordo com os seus gastos em saúde a partir da derivação de um indicador simples de eficiência relativa, que leva em consideração o desempenho na provisão de certos serviços de saúde e o gasto *per capita* nesse setor. Além de medir a extensão da ineficiência nos gastos com saúde, procura-se estimar o efeito das variáveis de escala sobre a eficiência relativa dos municípios. Também estima-se o efeito direto dessas variáveis sobre os indicadores de desempenho e o gasto, separadamente.

O artigo está organizado da seguinte maneira. A segunda sessão faz uma breve revisão da literatura empírica sobre eficiência dos gastos públicos, procurando ressaltar alguns resultados e identificando as principais variáveis utilizadas como insumo e produto em trabalhos anteriores. A terceira sessão descreve como serão calculados os indicadores de desempenho, fazendo uma breve resenha da técnica que será utilizada para avaliar a eficiência dos municípios, comumente denominada de FDH (*Free Disposable Hull*). A quarta sessão apresenta os dados utilizados, explica o porquê da escolha de determinadas variáveis em detrimento de outras como produto e expõe estatísticas descritivas. A quinta e a sexta seções apresentam, respectivamente, as regressões estimadas e os resultados. Finalmente são colocadas as conclusões, discutindo possíveis prescrições de política e sugestões para futuras investigações.

2. Medindo Eficiência dos Gastos Públicos

A literatura sobre desempenho relativo e absoluto dos governos na provisão de bens públicos está majoritariamente centrada na comparação internacional da eficiência dos gastos públicos a partir de uma estrutura muito agregada.

Evans et alii (2000) adotaram uma abordagem paramétrica para medir a eficiência dos sistemas de saúde de 191 países durante o período 1993-1997. Eles assumem que a função de produção de saúde tem um produto e dois insumos. Como produto foi utilizada a expectativa de vida da população (ajustada) e como insumo foram utilizados os gastos com saúde (públicos e privados) e a média de anos de escolaridade da população adulta. Os resultados indicam que Omã, Chile e Costa Rica apresentam os sistemas de saúde mais eficientes e Zimbábue, Namíbia, Bostwana, Malawi e Lesoto os sistemas de saúde mais ineficientes.

Gupta e Verhoeven (2001) usaram a abordagem de *Free Disposable Hull* (FDH) para avaliar a eficiência dos gastos em educação e saúde em 37 países africanos entre os anos de 1984 e 1995. Os resultados indicam que, comparativamente aos países da Ásia e do hemisfério ocidental, na média, os países da África são os que oferecem

serviços de educação e saúde de forma mais ineficiente. De qualquer maneira, observa-se um aumento na produtividade ao longo do tempo. Adicionalmente, eles encontraram uma relação negativa entre os scores de eficiência de produto e o nível de gasto público, levando à conclusão de que melhoras na educação e saúde exigem mais eficiência e não maiores alocações orçamentárias.

Afonso et alii (2003) construíram um *score* de desempenho do setor público para 23 países da OCDE. Este indicador é composto por 7 sub-indicadores que captam a qualidade das funções administrativas, os resultados em educação, os resultados em saúde, a qualidade da infra-estrutura, o grau de desigualdade, a estabilidade econômica e o desempenho econômico. Construído o indicador, eles consideraram o desempenho do indicador como produto e o gasto público total como insumo de forma a ordenar a eficiência dos países. Para tanto utilizaram como metodologia o FDH. Os *sub-indicadores dos resultados em saúde são compostos pela mortalidade infantil e pela expectativa de vida*.

Outros trabalhos – cuja metodologia foi resumida esquematicamente na Tabela 1 – ainda podem ser destacados como relevantes para o estudo da eficiência na gestão pública dos setores de educação e saúde. Esses estudos variam com relação ao período e região de análise, variáveis utilizadas como insumo e produto nos setores de educação e saúde e técnica adotada para construir *scores* de eficiência.

Para o Brasil, podem ser destacados os trabalhos de Marinho (2003), Sampaio de Souza et alii (2005), Miranda (2006) e Brunet et alii (2006). No trabalho de Sampaio de Souza et alii (2005), citado na introdução deste artigo, foram utilizados dois modelos de DEA para produzir *scores* de eficiência sobre os serviços dos municípios brasileiros. Os autores concluem que a proliferação de municípios como resultado da Constituição determinou uma redução de eficiência nos serviços municipais por questão de escala, captada por duas variáveis distintas, a densidade demográfica e taxa de urbanização. Nas variáveis políticas, o fato da cidade ser governada pelo PMDB e PDT implica uma perda de eficiência. Nas variáveis de gestão, o grau de utilização de computadores e a existência do poder de decisão em conselhos municipais implicam um aumento de eficiência.

3. Metodologia

3.1. Construção dos indicadores de desempenho

A fim de obter uma medida de desempenho do setor público (definido como o resultado as atividades do setor público) e eficiência do setor público (definida como o resultado relativo aos recursos empregados) é necessário agregar um conjunto de indicadores de desempenho das funções do Governo.

Afonso et alii (2003) assumem que o desempenho do setor público (DSP) depende dos valores de certos indicadores econômicos e sociais (I). Se existem n áreas de desempenho do Governo que juntas determinam o desempenho global no país i é possível escrever, então:

Tabela 1

Principais variáveis de insumo e produto e técnicas utilizadas na avaliação da eficiência dos gastos públicos

Autor	Amostra	Insumo	Produto	Técnica
Evans et alii (2000)	191 países	Gastos com saúde (públicos e privados) e média de anos de escolaridade da população adulta	Expectativa de vida da população (ajustada)	Painel
Gupta e Verhoeven (2001)	37 países africanos	Gasto <i>per capita</i> em saúde		FDH
Jarasuriya e Woodon (2002)	76 países em desenvolvimento	Produto <i>per capita</i> Gasto <i>per capita</i> Taxa de alfabetização	Expectativa de vida	Painel
Afonso et alii (2003)	OCDE	Gasto	Mortalidade infantil Expectativa de vida	FDH
Afonso e St. Aubyn (2004)	OCDE			DEA e FDH
Herrera e Pang (2005)	140 Países	Gasto público em saúde <i>per capita</i>	Expectativa de vida no nascimento Imunização tétano, difteria e malária Imunização rubéola Expectativa de vida ajustada	DEA e FDH
Marinho (2003)	74 municípios do Rio de Janeiro	Total de leitos contratados em hospitais <i>per capita</i> Total de hospitais credenciados <i>per capita</i> Total da capacidade ambulatorial instalada <i>per capita</i> Valor médio da internação Valor médio dos procedimento ambulatoriais	Total de internações em hospitais credenciados <i>per capita</i> Total de procedimentos ambulatoriais <i>per capita</i> Taxa de mortalidade	DEA
Sampaio et alii (2005)	Brasil – Municípios	Gasto corrente, número de professores, taxa de mortalidade, serviços hospitalares e de saúde	População total residente população alfabetizada, matrícula por escola, estudantes que frequentam escola, estudantes aprovados por escola, estudantes no ano correto, domicílios com acesso a água potável, domicílios com acesso a esgoto, domicílios com acesso à coleta de lixo.	DEA
Miranda (2006)	Brasil – Municípios	Total da despesa orçamentária municipal	Número de internações na rede Hospitalar municipal	DEA
Brunet et alii (2006)	Brasil – Estados	Despesa dos estados em saúde	Número de hospitais, número de postos de saúde, número de enfermeiros	FDH

Fonte: Elaboração própria.

$$DSP_i = \sum_{j=1}^n DSP_{ij} \quad (1)$$

onde $DSP_{ij} = f(I_1, \dots, I_K)$.

Desta forma, uma melhora no desempenho do setor público do país i na área j depende da melhora nos valores dos K sub-indicadores sócio-econômicos relevantes:

$$\Delta DSP_{ij} = \sum_{k=1}^K \frac{\partial f}{\partial I_k} \Delta I_k \quad (2)$$

Quanto maior o efeito positivo do gasto público sobre qualquer dos sub-indicadores selecionados, maior será a melhora esperada no indicador de desempenho do setor público.

Em um primeiro passo, os autores definem sete sub-indicadores de desempenho público. Os quatro primeiros olham para resultados administrativos, educação, saúde e infra-estrutura. Estes indicadores tentam refletir a qualidade da interação entre as políticas fiscais e o processo de mercado e entre o processo de mercado e a influência nas oportunidades individuais que isto tem. Estes são chamados de indicadores de “processo” ou “oportunidade”. Os outros três sub-indicadores refletem as tarefas “musgravianas” do Governo, que incluem um indicador de distribuição de renda, um indicador de estabilidade econômica que ilustra o alcance do objetivo de estabilização e outro indicador para avaliar desempenho econômico. A separação conceitual é certamente de alguma forma artificial. Por exemplo, os indicadores de saúde e educação poderiam ser também vistos como indicadores de eficiência alocativa. Finalmente todos os sub-indicadores são reunidos num indicador de desempenho do setor público. Cada indicador é composto por índices sócio-econômicos sobre os quais o Governo tem uma influência significativa, senão exclusiva, e, portanto, refletem o mais próximo quanto possível os resultados das políticas públicas. Por exemplo, com relação à saúde deve-se olhar, por exemplo, para a taxa de mortalidade infantil, imunização de doenças e o acesso a serviços de saúde.

Como o objetivo do trabalho é avaliar o desempenho e a eficiência dos gastos em saúde preocupa-se, na verdade, somente com este sub-indicador do indicador de oportunidade de Afonso et alii (2003). Assim, para construir-se o indicador de desempenho na saúde pondera-se somente os índices que compõem o sub-indicador de saúde. Além disso, como os indicadores são construídos para os municípios, na expressão para o indicador de desempenho (DSP_i) i indica o município, não mais o país. Os valores dos índices são normalizados em relação a média dos municípios de forma que a média dos índices resultantes seja igual a 1, ou seja, cada índice é obtido pela divisão do valor assumido para determinada variável “ x ” de um município i pela média da variável x de todos os municípios. Para calcular o indicador de desempenho da área de saúde a partir dos índices, dá-se inicialmente peso igual a cada um deles. Se, por exemplo, somente dois índices compõem o sub-indicador saúde, cada um deles contribui 50% para o indicador de desempenho saúde. Para

aqueles indicadores, em que números mais altos são mais desfavoráveis usa-se o inverso dos valores originais.

3.2. Método para calcular a fronteira de eficiência

A técnica utilizada para construir a “fronteira” foi a Free Disposable Hull (FDH). A FDH é uma técnica não-paramétrica proposta inicialmente por Deprins et alii (1984), sendo utilizada posteriormente para a avaliação de eficiência de qualquer unidade tomadora de decisão inclusive o setor público. A principal razão para esta escolha é que, com esta metodologia, não é preciso assumir uma fronteira de produção convexa.

Assume-se que um município qualquer pode produzir um bem ou um serviço y com uma quantidade de insumo x :

$$y = F(x) \quad (3)$$

O segredo na montagem do *score* é determinar quem é ou são os governos mais eficientes. Em termos de produto, é considerado eficiente o município que produz mais produtos e serviços com menos insumo. Para um determinado governo de um município A , basta que se selecionem entre todos os governos dos outros municípios quais são mais eficientes do que ele. Se nenhum governo for encontrado, então o governo do município A é considerado como eficiente e estabelece-se para ele um valor igual a 1.² Se um município qualquer não for o mais eficiente (exemplo o município A), seu *score* em termos de produto do município mais eficiente será definido da seguinte forma:

$$MIN_{n \in N} MAX_{p \in P} \frac{y_p(A)}{y_p(n)} \quad (4)$$

onde N é o número de municípios mais eficientes que A , P é o número de produtos, $y_p(n)$ denota a produção do p -ésimo tipo de produto pelo produtor n . Em termos de insumo é considerado eficiente o município que produz o mesmo nível de produto utilizando menos insumos. Neste caso, o *score* para o município A é definido da seguinte forma:

$$MIN_{n \in N} MAX_{m \in M} \frac{x_m(n)}{x_m(A)} \quad (5)$$

onde N é o número de municípios mais eficientes que A , M é o número de insumos, $x_m(n)$ denota o uso do m -ésimo tipo de input pelo produtor n .

Diferentemente do FDH, a técnica de Data Envelopment Analysis (DEA) é uma técnica que constrói as fronteiras de produção usando programação linear. Supondo que existam os mesmos m produtos e k insumos para os n municípios,³ para todo

² Apesar de nosso argumento, muitos governos, não apenas um, podem estar no *score* mais elevado.

³ Veja Coelli et alii (1998) para maiores detalhes.

município existe um y_i (vetor coluna dos produtos ou serviços) e um x_i (vetor coluna de insumos). De maneira complementar, existe uma matriz ($k \times n$) de insumos e uma matriz ($m \times n$) de produto. A partir dessas definições, pode-se apresentar o problema matemático de programação linear da seguinte forma:

$$MIN_{\theta, \lambda} \theta$$

sujeito a $-y_i + Y\lambda \geq 0$

$\theta x_i - X\lambda \geq 0$

$n_1' \lambda = 1$

$\lambda \geq 0$

Nesse problema, θ é um escalar (que satisfaz $\theta \leq 1$) que mede a eficiência técnica do município (x_i, y_i). Na prática é uma medida da distância entre o município e a fronteira de eficiência. Se $\theta < 1$, o município está dentro da fronteira e, portanto, é ineficiente. Se $\theta = 1$, o município está na fronteira e, portanto, é eficiente. O vetor λ é um vetor ($n \times 1$) de constantes com os pesos usados para computar a alocação ineficiente dos municípios se eles forem ineficientes. n_1 é um vetor dimensional de uns. $n_1 \lambda$ impõe a convexidade da fronteira.

Finalmente, uma outra alternativa metodológica disponível é o método de fronteiras estocásticas de produção. Ele tem como principal vantagem sobre o FDH e DEA a possibilidade de fazer a decomposição do resíduo em dois componentes: um relativo à ineficiência e outro relativo ao choque puramente aleatório. Esse método foi introduzido na literatura econômica simultaneamente por Aigner et alii (1977) e Meeusen e Van de Broeck (1977). Sua especificação para dados em *cross-section* é a seguinte:⁴

$$\ln y_i = \beta_0 + f(\beta_k, x_i^k) + v_i - u_i, \quad i = 1, \dots, I; k = 1, \dots, n$$

em que $\ln y_i$ é o logaritmo da quantidade produzida pelo município i ; β_0 é o intercepto da equação; $f(\beta_k, x_i^k)$ é a forma funcional adequada; β_k é o vetor de coeficientes tecnológicos; x_i^k é o vetor de insumos utilizado na produção pelo município i ; v_i é o choque aleatório não correlacionado com x_i^k e u_i , com distribuição $N(0, \sigma_v^2)$ e u_i é o termo de ineficiência não negativo do município i também não correlacionado com x_i^k .

Para a estimação dos coeficientes tecnológicos das fronteiras estocásticas de produção são empregados geralmente os estimadores de máxima verossimilhança.⁵ É necessário assumir hipóteses distributivas sobre o termo de ineficiência u_{it} . Diversas distribuições assimétricas têm sido utilizadas para o termo u_{it} , sendo

⁴ Posteriormente, Pitt e Lee (1981) especificaram o modelo de Aigner et alii (1977) para dados em painel. Assim, a função de produção passa a assumir a seguinte forma:

$$\ln y_{it} = \beta_0 + f(\beta_k, x_{it}^k) + v_{it} - u_{it}, \quad i = 1, \dots, I; k = 1, \dots, n, t = 1, \dots, T$$

em que todas as hipóteses válidas usando em *cross section* valem para o modelo em painel.

⁵ Para outros estimadores que podem utilizados para o cálculo dos coeficientes ver Kumbhakar e Lovell (2000).

as mais populares a distribuição semi-normal, distribuição normal-truncada, distribuição exponencial e distribuição gama.

Quando se escolhe um método econométrico, deve-se avaliar as vantagens existentes entre a escolha feita e as alternativas possíveis. A primeira questão que se coloca é porque não utilizar mínimos quadrados ordinários (MQO). Conforme colocam Kumbhakar e Lovell (2000), o estimador de mínimos quadrados, apesar de não ser viesado para o vetor de coeficientes tecnológicos é viesado para o intercepto da equação a ser estimada.⁶ Além disto, tal metodologia não permite a existência de ineficiência técnica, atribuindo toda variação na produção que não possa ser explicada por variação dos insumos a choques aleatórios. Nesta situação, caso não haja presença de ineficiência, MQO é o estimador mais adequado para a função de produção. Assim, MQO assume municípios plenamente eficientes.

A primeira questão que se coloca é a escolha entre métodos não paramétricos e paramétricos. Um argumento contra os métodos não paramétricos é o de que eles constroem a fronteira a partir dos próprios dados, com o que as medidas de eficiência derivadas são válidas no sentido de que elas refletem somente quão eficiente é um município em relação a outros naquela amostra particular. Dado que a eficiência é estimada relativamente, *outliers* nos dados podem alterar o formato da fronteira e distorcer os *scores* de eficiência.

No caso da técnica de fronteira estocástica, os erros aleatórios podem ser tratados através de inferência estatística dos parâmetros estimados como visto anteriormente. Além do mais, uma vez que caracteriza o comportamento do “município médio” as estimativas são muito menos sensíveis a mudanças na amostra.

A desvantagem é a necessidade de se fazer uma escolha arbitrária da distribuição da eficiência. Muito embora haja distribuições flexíveis que não impõem restrições sobre a média da ineficiência, esta pode sofrer alterações de acordo com distribuição escolhida. Entretanto, há evidências de que a posição relativa das unidades tomadoras de decisão no *ranking* de eficiências não é muito sensível à forma distributiva. assumida para as unidades tomadoras de decisão como mostra Greene (2000).⁷

Neste trabalho, a opção final pelo uso de um método não paramétrico foi feita sobretudo com base no fato de que o método de fronteira estocástica exige apenas uma variável dependente. Assim sendo, seria preciso agregar os outputs selecionados do setor de saúde, ou fazer uma regressão para cada produto. Neste último caso, não conseguiríamos avaliar a eficiência na produção conjunta dos produtos, que é um dos objetivos deste trabalho.

Uma vez feita a opção por um método não paramétrico, resta a escolha entre o FDH e o DEA. O FDH requer apenas uma única hipótese para que exista uma

⁶ Kumbhakar e Lovell (2000) argumentam que com dados em painel e ineficiência variante no tempo, o estimador de MQO será inconsistente para os coeficientes da matriz de regressores. Além disto, o autor argumenta que esta inconsistência pode ser evitada usando-se máxima verossimilhança.

⁷ Neste mesmo estudo, o autor mostra que, para 123 geradoras de eletricidade dos Estados Unidos, as médias das distribuições não variam abruptamente. Seus resultados reportam valores médios de ineficiência iguais a 0,8766 para a distribuição semi-normal, 0,8949 para a distribuição gama, 0,8961 para a distribuição normal truncada e 0,9011 para a exponencial.

fronteira de produção e que seja contínua, qual seja, a de que insumos e produtos possam ser livremente descartados. Não é necessário que o conjunto de possibilidade de produção seja convexo. Um exemplo plausível de não convexidade deste conjunto ocorre quando existe *set up costs* na produção. Como a identificação empírica de que a conjunto de possibilidade de produção no setor público é ou não convexo foge do escopo do trabalho, preferiu-se escolher o método que não depende desta hipótese (FDH).

Assim, o que se percebe é que as vantagens de usar qualquer uma das abordagens tende a salientar as desvantagens de outra abordagem. Assim, optamos por seguir a literatura internacional que usa majoritariamente o FDH.

4. Escolha das Variáveis para Cálculo dos *Scores* e Estatísticas Descritivas

4.1. *Medindo resultados da área de saúde*

A elaboração dos indicadores privilegiou dados que mostrassem **resultados**, ao invés de dados de despesas com produtos. A razão para isso é que produtos da área médica, como número de consultas e distribuição de medicamentos, não necessariamente mostram a qualidade do serviço prestado. Por exemplo, é possível um município apresentar gastos elevados em atendimento a doentes simplesmente porque o serviço de prevenção foi ineficaz.

O trabalho é limitado pelo complicado arcabouço institucional da saúde municipal. A maior complexidade se refere à possibilidade de municípios estabelecerem convênios entre si, principalmente para aproveitar economias de escala em certos tratamentos. Dessa forma, se um município não tem o aparato técnico necessário para tratar certa enfermidade, o paciente pode ser transferido para outra localidade. Com isso, alguns municípios acabam sendo pólos de tratamento, recebendo muitos residentes de outros municípios. Mas esses municípios atratores de pacientes devem receber, a princípio, pagamento por parte dos municípios beneficiados, uma vez que cada município é responsável pela saúde de seus residentes.

Assim sendo, a forma escolhida para a construção das variáveis (de gastos e de resultados) foi o agrupamento pelos **residentes** em certa localidade. Isto é, os gastos de cada município e também os seus resultados são contabilizados tomando como base os seus **residentes**. Mesmo assim, porém, os custos do tratamento de um paciente podem não ser completamente pagos pelo município em que este reside, o que ainda poderia levar a distorções nas despesas municipais.

A escolha de variáveis é direcionada para representar a qualidade do atendimento e minimizar o possível viés ocasionado por características locais. Uma possível variável para os indicadores seria mortalidade infantil, que capta o acesso da população à rede de saúde. Entretanto, sabe-se que a mortalidade infantil pode estar fortemente relacionada a outros fatores não diretamente ligados ao serviço de

saúde. A estatística pode ser bastante influenciada pelo nível de renda do município, pelo nível de escolaridade da população, pelas condições do saneamento básico, entre outros fatores. Para contornar esse problema, uma alternativa seria restringir os dados para que só captassem a mortalidade de crianças com menos de um mês de vida (mortalidade neonatal), ou ainda a quantidade de consultas pré-natais *per capita*. Entende-se que essas duas variáveis dependem fundamentalmente do serviço prestado em hospitais ou postos de atendimento, e menos de fatores regionais ou socioeconômicos. Porém, a seleção esbarra na disponibilidade dos dados. Apesar do sistema de saúde apresentar esse detalhamento, cerca de 50% dos municípios não apresentam dados de mortalidade neonatal, enquanto ao menos 26% dos municípios paulistas não divulgaram a quantidade de consultas pré-natais realizadas.

Levando em consideração as limitações e dificuldades expostas, os indicadores de qualidade do atendimento em saúde no estado de São Paulo selecionados procuram diferenciar os municípios em duas dimensões:

1. **Acesso à rede pública de saúde (acesso ao sistema de saúde):** esse índice verifica o acesso da população aos serviços de saúde. A variável é a razão entre os residentes mortos em hospitais ou outro estabelecimento de saúde, dividido pelos óbitos gerais da população municipal, dentro ou fora do sistema. As duas estatísticas são fornecidas pelo DATASUS. A idéia é que quanto mais fácil for o acesso aos serviços de saúde para a população, maior será a proporção de mortes dentro de hospitais ou outros estabelecimentos de saúde.
2. **Serviços de prevenção (doenças infecciosas):** Permite verificar como os serviços de prevenção municipal atuam. O índice é construído com base no número de internações *per capita* devido a doenças infecciosas, por local de residência. Se doenças infecciosas podem ser evitadas com um serviço de prevenção eficiente, quanto menor o número de internações devido a essas doenças, melhor o serviço de prevenção do município.

Todos os dados se referem ao ano de 2005. Dos 645 municípios do estado de São Paulo, foram utilizados 610 de acordo com a disponibilidade de dados. Isso equivale a 95% dos municípios do estado.

É importante ressaltar que foi feita a tentativa de incluir uma terceira dimensão, que seria a qualidade do atendimento hospitalar. A medida proposta seria a razão entre o número de óbitos nos hospitais e o número de Autorizações de Internação Hospitalar pagas e computadas como internações, pelo local de residência do paciente (mortalidade hospitalar pelo local de residência).

Sabe-se que a mortalidade hospitalar certamente não pode ser usada para comparar a qualidade de diferentes hospitais, pois muitas vezes os hospitais tecnicamente superiores acabam recebendo os casos mais difíceis e graves, o que faz com que a mortalidade nestes seja maior. Ou seja, os melhores hospitais apresentam naturalmente um “viés de gravidade” dos casos, o que leva a uma maior mortalidade de seus pacientes. No entanto, acreditava-se inicialmente não existir um “viés de gravidade” por local de residência. Portanto, foram feitos cálculos de mortalidade hospitalar tomando como base o local de residência dos pacientes.

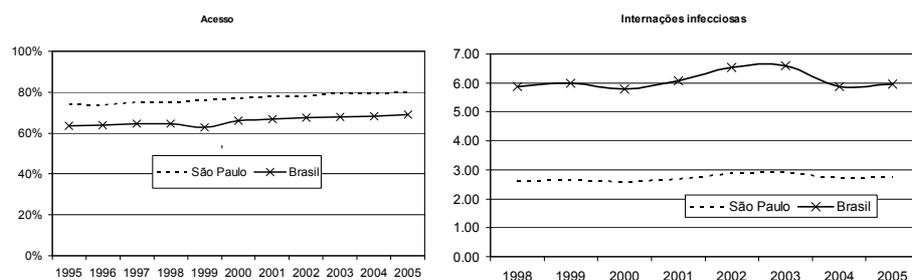
Os dados revelaram, porém, que o Estado de São Paulo tem um índice de

mortalidade hospitalar de seus residentes bastante superior ao resto do país. Em 2007, por exemplo, esse índice foi inferior a 35% para o Brasil e superior a 45% em São Paulo. Esse resultado é contra-intuitivo, pois, em outras estatísticas do sistema de saúde, que incluem o *acesso ao sistema de saúde e as internações por doenças infecciosas*, São Paulo apresenta melhores condições do que a média nacional. Uma possível explicação para esse resultado contra-intuitivo está justamente no fato dos residentes de São Paulo terem maior acesso ao sistema de saúde. Em outros estados, é possível que certos doentes graves não sejam mantidos nos hospitais. Se isso não ocorre em São Paulo, isso geraria um “viés de gravidade” dos internados do Estado de São Paulo. Assim sendo, neste trabalho foi feita a opção de não trabalhar com a variável mortalidade hospitalar e, portanto, não foi possível medir a dimensão *qualidade do atendimento hospitalar*.

4.2. Evolução estadual dos dois Indicadores de Saúde elaborado para os Municípios Paulistas

O acesso ao sistema de saúde em São Paulo é superior ao resto do Brasil. Em média, 80% dos óbitos no estado ocorrem em hospitais ou outras unidades de saúde. Esse mesmo índice no Brasil é de 69%. Tal resultado já era esperado e, de acordo com as hipóteses do trabalho, revela que o alcance da rede pública de saúde em São Paulo é maior do que no resto do país.

A série histórica da variável acesso ao sistema de saúde mostra que a diferença entre as médias do estado e do país se mantém entre os anos de 1995 e 2005, ambos com tendência ascendente. Isso mostra que o atendimento hospitalar é cada vez mais abrangente.



*Razão entre a quantidade de óbitos em unidades de saúde e a quantidade de óbitos total da região (por local de residência).

**Razão entre a quantidade de internações por doenças infecciosas e a população, no período, multiplicada por 1000 (por local de residência).

Fonte: Datasus – Elaboração própria.

Fig. 1. Evolução do acesso ao sistema de saúde* e evolução de internações por doenças infecciosas** em São Paulo e no Brasil

São Paulo mantém taxas de internação por doenças infecciosas inferiores ao restante do país, embora a diferença venha diminuindo marginalmente nos últimos

anos.

4.3. *Análise descritiva dos scores de eficiência*

Como vimos anteriormente, as metodologias de mensuração de eficiência permitem o cálculo da “fronteira de possibilidade de produção” e de *scores* de eficiência de insumo e de eficiência de produto, a fim de ordenar os municípios que fazem parte da amostra em termos de eficiência de gastos em saúde. Esses *scores* de eficiência se situam entre 0 e 1 e todos os municípios colocados na fronteira de possibilidade de produção receberão o *score* máximo de 1. O *score* de eficiência de um dado município indica quanto insumo a menos este município poderia usar para alcançar o mesmo nível de produto. O *score* de eficiência de produto de um dado município diz quanto produto a mais o município poderia produzir com a mesma quantidade de recursos que ele está correntemente usando.

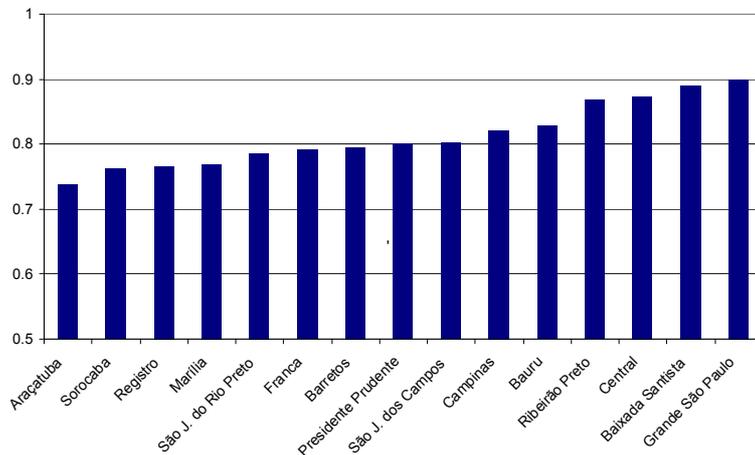
As estatísticas descritivas serão apresentadas organizando os municípios de acordo com três diferentes critérios: região administrativa, nível de renda dos municípios e tamanho da população. Como a variável mortalidade hospitalar foi considerada inadequada para captar a qualidade do atendimento hospitalar, o índice médio é calculado somente a partir do acesso ao sistema de saúde e internações por doenças infecciosas.

Análise dos scores de eficiência segundo Regiões Administrativas

O primeiro *score* examinado, na Figura 2, é baseado no acesso ao sistema de saúde. A ineficiência média das regiões administrativas é menor do que em internação por infecções (ver Figura 3). Além disso, o *score* do acesso ao sistema de saúde apresenta uma variação menor entre as diversas regiões. Araçatuba, com o menor índice, chega a uma ineficiência de aproximadamente 30%, enquanto a Grande São Paulo, a mais eficiente, tem uma ineficiência de aproximadamente 10%.

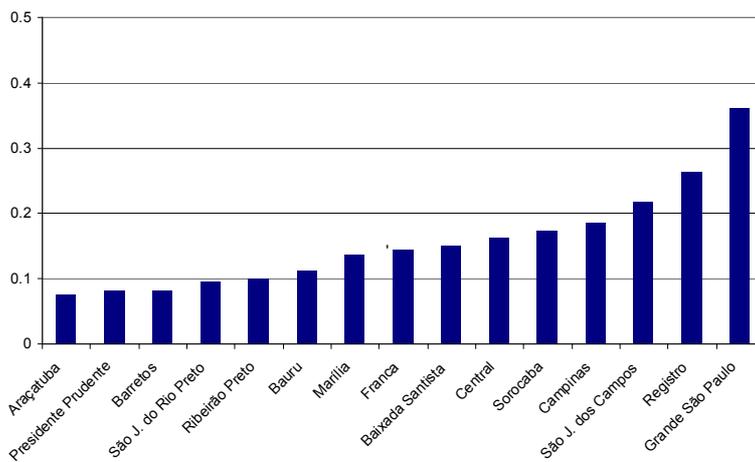
Se o acesso ao sistema de saúde se destaca por relativamente baixa ineficiência média, no que diz respeito a internações por doenças infecciosas, o que chama atenção é o desempenho médio ruim em todas as regiões. Isso revela que, com relação à prevenção de doenças infecciosas, há poucos municípios na fronteira ou próximos à fronteira de possibilidades de produção. Ou seja, uma grande proporção de municípios está bem distante da fronteira nesse critério. A Região Administrativa da Grande São Paulo, com o *score* médio de 36%, é a menos ineficiente, seguida por Registro e São José dos Campos. Araçatuba, Presidente Prudente, Barretos e São José do Rio Preto obtiveram *scores* médios com grau de ineficiência superior a 90%.

O *score* do indicador médio, construído com base nas duas dimensões selecionadas (acesso ao sistema e doenças infecciosas), mostra que as regiões da Grande São Paulo, Registro e São José dos Campos são as mais eficientes na área



Fonte: Elaboração própria.

Fig. 2. Scores do índice de acesso ao sistema de saúde por Região Administrativa

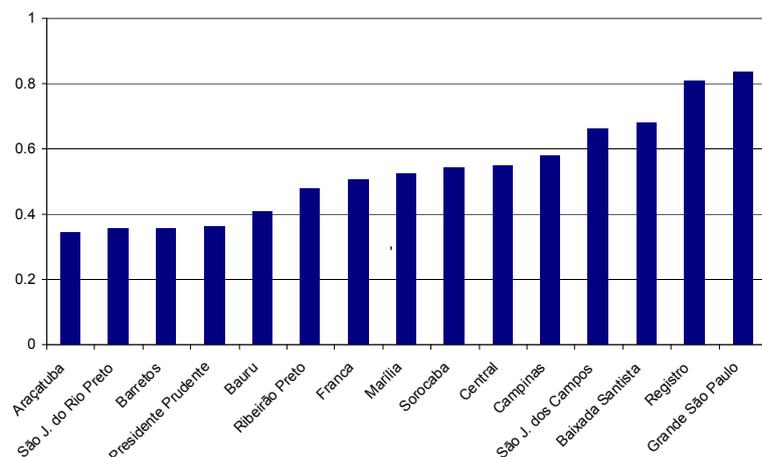


Fonte: Elaboração própria.

Fig. 3. Scores do índice de internações por doenças infecciosas por Região Administrativa

da saúde. As regiões do interior do estado, como Araçatuba, Barretos e Presidente Prudente, por outro lado, detêm os piores indicadores médios.

Por último, ordenamos os municípios de acordo com cada um dos três *scores* e separamos conforme o quartil de desempenho em cada índice. Deste modo, podemos observar, para cada região administrativa, qual é a proporção de municípios alocados no quartil superior, isto é, aqueles entre os 25% mais eficientes para cada indicador examinado. Para ter uma melhor avaliação de cada Região Administrativa, podemos dividir o número de cidades localizadas no quartil superior pelo total de municípios na região administrativa. Chamamos esse



Fonte: Elaboração própria.

Fig. 4. Scores do indicador médio por Região Administrativa

indicador de indicador de Eficiência Intra-Região (EIR).

Os resultados dos EIR's estão alinhados com os números apresentados anteriormente. Os EIR's das regiões administrativas mostram que a Grande São Paulo é a região com maior número de cidades, proporcionalmente, no quartil superior em todas as áreas analisadas. São José dos Campos é outra região que obteve uma grande proporção de municípios no quartil superior. Outras regiões apresentam bons resultados em apenas uma dimensão, não sendo consistente com a outra, como Baixada Santista (eficiente em acesso) e Registro (eficiente em prevenção de doenças infecciosas). Araçatuba possui uma proporção pequena de municípios eficientes de um modo geral, assim como outras regiões interioranas.

Análise dos scores de eficiência segundo PIB per capita municipal

Esta seção apresenta o desempenho médio dos municípios de acordo com o quartil de renda de sua população. Os municípios foram agrupados em quatro diferentes quartis de renda, onde o primeiro representa os municípios com menor renda *per capita* e o quarto os de maior renda.

Renda *per capita* não parece configurar importante controle para eficiência nos gastos em saúde. A Tabela 3 mostra que não há uma sequência do mais pobre para o mais rico, ou vice-versa, em nenhum índice. A maior eficiência em acesso está no terceiro quartil de renda, seguido do quarto e segundo quartis. A maior eficiência em prevenção de doenças infecciosas está no primeiro quartil (mais pobre), porém seguido do último quartil de renda (mais rico).

Tabela 2

Indicador de eficiência intra-região (proporção de municípios no quartil mais eficiente)

Região administrativa	Acesso ao sistema de saúde	Infecciosas	Indicador médio
Araçatuba	7%	5%	5%
Baixada Santista	50%	25%	13%
Barretos	22%	0%	6%
Bauru	32%	16%	16%
Campinas	24%	29%	27%
Central	44%	16%	24%
Franca	15%	30%	30%
Grande São Paulo	62%	81%	84%
Marília	3%	20%	28%
Presidente Prudente	28%	8%	12%
Ribeirão Preto	41%	14%	9%
São J. do Rio Preto	21%	12%	12%
São J. dos Campos	37%	40%	37%
Sorocaba	16%	30%	27%
Registro	8%	75%	67%

Fonte: Elaboração própria.

Tabela 3

Scores dos índices por quartil do PIB *per capita*

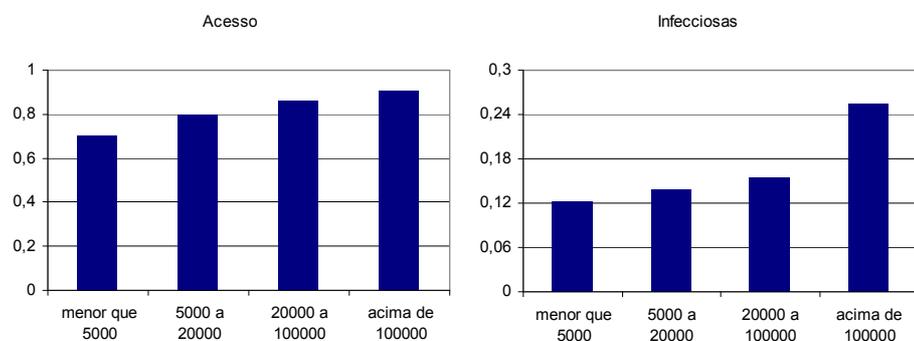
Quartis	Acesso ao sistema	Infecciosas	Média
1o. quartil	0,776	0,208	0,575
2o. quartil	0,800	0,117	0,419
3o. quartil	0,823	0,137	0,471
4o. quartil	0,815	0,143	0,554

Fonte: Elaboração própria.

Análise dos scores por Classes de Tamanho da População dos Municípios

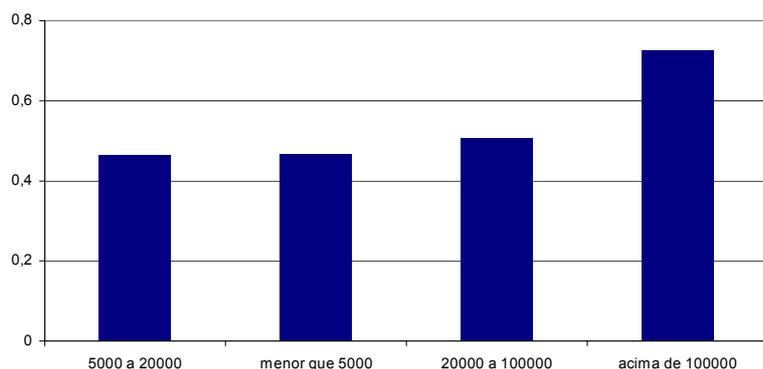
Na análise por tamanho da população, os municípios foram agrupados de acordo com o tamanho municipal pelo critério do IBGE, somando-se o grupo de 500.000 ou mais habitantes com o grupo de 100.001 a 500.000 habitantes.

Em geral municípios mais populosos apresentam *scores* mais elevados. Municípios pequenos, principalmente aqueles com menos de 5.000 habitantes, apresentaram baixo grau de eficiência nos gastos em saúde, o que reforça a idéia de economias de escala no setor.



Fonte: Elaboração própria.

Fig. 5. *Scores* do índice de acesso ao sistema de saúde e internações por doenças infecciosas, por tamanho da população do município



Fonte: Elaboração própria.

Fig. 6. *Scores* do índice de médio por tamanho da população do município

5. As Regressões

5.1. Análise econométrica dos scores de eficiência

O objetivo da análise econométrica é verificar como características municipais afetam eficiência na saúde, principalmente aspectos de escala gerados por um maior número de habitantes. Para tanto foi utilizado o seguinte modelo econométrico:⁸

$$\begin{aligned} score_i = & \beta \log(populacao)_i + \theta concentracao_pop_i + \lambda pop_urbana_i \\ & + \Theta CONTROLES_i + \mu_i \end{aligned}$$

A variável dependente, *score*, são os *scores* de eficiência. Serão feitas regressões para as duas diferentes dimensões analisadas e para a média de ambas. O primeiro *score* reflete a qualidade no acesso à saúde no município *i*. O segundo mede a eficiência nos serviços de prevenção e o terceiro é o *score* médio calculado a partir dos dois anteriores.

O logaritmo da população é a principal variável de interesse. Queremos testar se um município, por ter um maior número de habitantes e, assim, uma maior escala nos serviços de saúde, apresenta uma maior eficiência. Para melhor controlar o efeito de escala populacional, também adicionamos ao modelo variáveis que mostram a concentração populacional e o grau de urbanização do município representado pela razão de pessoas que moram na região urbana. Para estimar o efeito *ceteris paribus* do tamanho da população, inserimos no modelo sete diferentes grupos de controles que podem estar correlacionados com eficiência na saúde.⁹

A utilização de um *score* de eficiência como variável dependente numa regressão por mínimos quadrados ordinários não necessariamente gera estimativas consistentes dos parâmetros, uma vez que esses *scores* são tipicamente correlacionados. Simar e Wilson (2003) apontam que enquanto o processo gerador dos dados não for descrito apropriadamente, fica difícil qualquer tipo de inferência. Neste sentido, é preciso assumir a condição de separabilidade entre o espaço gerado pelas variáveis insumo e produto e as variáveis de controle usadas na regressão. Os controles agem somente no sentido de deslocar *para fora da fronteira de eficiência* os governos locais, não influenciando a posição da própria fronteira.

Simar e Wilson descrevem o que é chamado de *sensível* processo gerador de dados, consistente com a análise realizada neste trabalho. Ou seja, sob o conjunto de hipóteses A1-A8 (ver páginas 34-37), naquele artigo pode-se inferir que o *score* de eficiência não paramétrico utilizado neste artigo é consistente apesar da (relativa) baixa taxa de convergência. Para estimar a regressão, em princípio bastaria assumir

⁸ O uso de regressões em dois estágios em que se utiliza a medida de eficiência (primeiro estágio) como resposta em modelos de regressão (segundo estágio) é questionável. Ver, por exemplo, Simar e Wilson (2003). Uma tentativa de contornar o problema, no entanto, é apresentada na subseção 5.2.

⁹ Os conjuntos de controles são: demografia, renda, saneamento e vacinação, educação, violência, clima e ideologia do partido no poder municipal. Para uma apresentação mais detalhada das variáveis de controle, ver Apêndice.

que o vetor de parâmetros (β) é finito, especificar a distribuição dos erros e estimá-los por máxima verossimilhança, uma vez que a variável dependente está truncada.¹⁰ Os autores propõem o uso do método regressão truncada e com isso os parâmetros da regressão seriam consistentemente estimados. Para fazer inferência, dado que a variável dependente possui correlação serial, a proposta é a realização de um *bootstrap* para o cálculo do intervalo de confiança.

Assim sendo, neste estudo, além da análise de regressão clássica por mínimos quadrados, também foram realizadas estimativas por máxima verossimilhança (truncadas) e foi feito o cálculo do intervalo de confiança a partir de *bootstrap*, seguindo as sugestões de Simar e Wilson (2003). Foi efetuado o algoritmo#1 proposto pelos autores com 400 repetições.¹¹ Os resultados apresentados na seção 6 sugerem um intervalo de confiança semelhante ao obtido quando assumimos a distribuição normal truncada dos erros para seu cálculo.

Por fim, buscou-se identificar o efeito das variáveis explicativas da regressão sobre o insumo e os produtos separadamente, levando em conta a correlação entre os regressores, procedimento detalhado na próxima subseção.

5.2. Medindo os efeitos de escala sobre desempenho e gastos separadamente

Para determinar o canal pelo qual o ganho de escala afeta o *score* de eficiência, estimamos também uma equação para cada produto (acesso a saúde e serviços de prevenção) e para o insumo (gastos com saúde *per capita*). De acordo com nossa estrutura adotada para calcular os indicadores de eficiência, nota-se que estas três variáveis estão conectadas pela função de produção. O *score* de eficiência capturaria como estas variáveis (gastos com saúde *per capita*, acesso a saúde e serviços de prevenção) estão inter-relacionadas. Desta forma, parece razoável assumir que a forma mais apropriada para estimar os efeitos seja o modelo de regressões aparentemente não relacionadas (SUR, em inglês), uma vez que os erros destas diferentes equações podem estar relacionados. Usamos as mesmas variáveis explicativas em todas as regressões estimando o seguinte conjunto de equações simultâneas.

$$\begin{aligned} \text{Gastos_saude}_i &= \beta_1 \log(\text{populacao})_i + \theta_1 \text{concentracao_pop}_i + \lambda_1 \text{pop_urbana}_i \\ &\quad + \Theta_1 \text{CONTROLES}_i + \eta_i \\ \text{Acesso}_i &= \beta_2 \log(\text{populacao})_i + \theta_2 \text{concentracao_pop}_i + \lambda_2 \text{pop_urbana}_i \\ &\quad + \Theta_2 \text{CONTROLES}_i + \varepsilon_i \\ \text{Prevencao}_i &= \beta_3 \log(\text{populacao})_i + \theta_3 \text{concentracao_pop}_i + \lambda_3 \text{pop_urbana}_i \\ &\quad + \Theta_3 \text{CONTROLES}_i + v_i \end{aligned}$$

¹⁰ Esta variável pode ser truncada à direita ou à esquerda dependendo da medida de eficiência usada: *input* ou *output*.

¹¹ Foi o número máximo suportado pelo Stata e utilizado por Simar e Wilson (2003). Detalhes estão a disposição mediante requisição aos autores.

onde as variáveis dependentes são gastos com saúde *per capita*, acesso a saúde e serviços de prevenção para o município i e as independentes foram descritas acima. A principal vantagem desta estrutura é permitir que os resíduos η_i, ε_i e v_i sejam correlacionados, algo sugerido pelo método usado para calcular o *score* de eficiência.

Estas estimativas buscam identificar como informações adicionais relacionadas ao processo de produção afetam o desempenho do setor público mas não podem ser considerados nem insumos nem produtos. Estamos assumindo, assim, que o conjunto de variáveis explicativas do SUR é exógeno ao processo de produção embora possa explicar parte dele. São variáveis que fogem do controle do poder público local quando no momento da provisão do bem (ver Bartelsman e Doms (2000) e Morroni (2006) para mais detalhes).

6. Resultados

6.1. Resultados das regressões por OLS para os scores de eficiência

Conforme podemos observar na tabela a seguir, as regressões por mínimos quadrados ordinários e as regressões truncadas apresentam coeficientes com mesmo sinal e magnitude muito semelhante. As regressões por mínimos quadrados sugerem que um maior número de pessoas no município acarreta um aumento na eficiência em saúde em todas as dimensões analisadas. Contudo, para as regressões truncadas, a variável *logpop* deixa de ser significativa para a eficiência nos serviços de prevenção e o *score* médio. Por outro lado, no que se refere à eficiência no acesso aos serviços de saúde, o coeficiente é sempre significativo ao nível de 1%.

Concentração urbana não parece ser relevante, porque, mesmo significante, a amplitude de seu coeficiente é desprezível. Já população urbana influencia os *scores* de acesso e de prevenção, porém a variável *age* de maneira contrária para as duas variáveis dependentes. Quanto maior a população urbana municipal, maior a qualidade do acesso aos serviços de saúde, porém menor eficiência em prevenção. Possivelmente, dada a externalidade negativa de doenças infecciosas, há uma maior dificuldade de combater enfermidades onde há uma maior razão de pessoas vivendo em áreas urbanas.

A maioria das variáveis de controle que apresentou efeito significativo segue o sinal esperado.¹² Exceções são os sinais negativos para renda *per capita* e cobertura de planos de saúde privados sobre a eficiência na prevenção de doenças infecciosas. Por outro lado, melhores instalações sanitárias aumentam a eficiência em prevenção e maior proporção de idosos e clima mais quente reduzem essa eficiência, conforme esperado.

¹² No Apêndice foi colocada uma tabela com os sinais esperados das variáveis de controle. Espera-se que as variáveis que medem renda e educação da população tenham efeito positivo sobre a eficiência e também sobre o desempenho na área de saúde. Para alguns controles, como ideologia política do prefeito, não há nada que se possa dizer *a priori*.

Prefeituras de partidos classificados como de direita e esquerda tendem a ser mais eficientes que as de centro – ideologia em que estão classificados os partidos PMDB, PSC, PSL e PRB, mas apenas o primeiro detinha o poder de mais de cinco prefeituras no período analisado. A classificação utilizada para a ideologia foi a de Coppedge (1997).

Para compreender melhor o efeito das características municipais sobre a eficiência na gestão da saúde pública, a seção a seguir analisa separadamente os efeitos sobre os indicadores de desempenho e sobre o gasto *per capita* em saúde.

6.2. Resultados das regressões por SUR para os indicadores de desempenho e gastos per capita

Os resultados das regressões por SUR revelam que municípios maiores têm menor despesa *per capita* com saúde e, ao mesmo tempo, oferecem melhores condições de acesso ao sistema de saúde para seus habitantes, mesmo controlando pela concentração populacional e pela taxa de urbanização, que também tem efeito positivo no desempenho em termos de acesso.

A taxa de urbanização da população tem efeito negativo no desempenho de prevenção de doenças infecciosas, reforçando a idéia de que é mais difícil combater essas doenças em regiões onde há maior proporção de pessoas vivendo em áreas urbanas.

Em consonância com os resultados de Sampaio de Souza et alii (2005), os resultados deste estudo sugerem fortemente que a excessiva descentralização na gestão pública da saúde pode levar à perda de eficiência na oferta de serviços de saúde. Fica claro que municípios menores, além de terem maior gasto *per capita* com saúde, não conseguem oferecer aos seus habitantes a mesma facilidade de acesso ao sistema de saúde que os grandes municípios. Levando em consideração que os municípios maiores tendem a concentrar maior variedade e sofisticação de serviços, os resultados deste estudo sugerem que os acordos feitos entre pequenos e grandes municípios para que os primeiros possam compartilhar tratamentos alocados nos outros não garantem igual acesso a todos os cidadãos. Fica a dúvida se a centralização da gestão melhoraria essa situação.

O tamanho da população não tem efeito significativo sobre incidência de doenças infecciosas, mas o coeficiente tem sinal negativo, associando positivamente, portanto, escala e bom desempenho em prevenção de doenças infecciosas. É certo, portanto, que pequenos municípios não têm melhor desempenho que os grandes na prevenção de doenças infecciosas, sendo que as políticas de prevenção fazem parte das políticas de responsabilidade primordialmente municipal e as prefeituras de municípios menores têm, a princípio, melhores condições de conhecer as necessidades específicas de seus cidadãos. Assim sendo, segundo os resultados deste trabalho, conclui-se que a descentralização na gestão da prevenção de doenças infecciosas não melhora o desempenho nessa área de atuação.

Quanto às variáveis de controle, pode-se dizer que aquelas que apresentaram

Tabela 4
Resultados das estimações para os *scores* de eficiência

	Acesso ols coef/se	Acesso trunc coef/se	Acesso trunc bootstrap coef/se	Infec ols coef/se	Infec trunc coef/se	Infec trunc bootstrap coef/se	Médio ols coef/se	Médio trunc coef/se	Médio trunc bootstrap coef/se
logpop	0,009*** (0,003)	0,013*** (0,004)	0,013*** -	0,011** (0,005)	0,005 (0,004)	0,005 -	0,009* (0,006)	0,004 (0,004)	0,004 -
conpop	0,000 (0,000)	0,000 (0,000)	0,000 -	0,000 (0,000)	0,000 (0,000)	0,000 -	0,000 (0,000)	0,000* (0,000)	0,000 -
popurb	0,246*** (0,047)	0,272*** (0,054)	0,272*** -	-0,242*** (0,093)	-0,155*** (0,054)	-0,155*** -	-0,103 (0,087)	-0,026 (0,053)	-0,026 -
idosos	-0,002 (0,004)	-0,002 (0,004)	-0,002 -	-0,015*** (0,005)	-0,014*** (0,004)	-0,014*** -	-0,015*** (0,005)	-0,014*** (0,004)	-0,014*** -
crianças	-0,001 (0,007)	-0,001 (0,008)	-0,001 -	0,002 (0,011)	0,001 (0,008)	0,001 -	0,005 (0,010)	0,004 (0,008)	0,004 -
rendapc	-0,000 (0,000)	-0,000 (0,001)	0,000 -	-0,001* (0,001)	-0,001 (0,001)	-0,001 -	-0,001* (0,001)	-0,001* (0,001)	-0,001* -
gini	-0,048 (0,236)	-0,034 (0,341)	-0,034 -	-1,480*** (0,401)	-0,986*** (0,319)	-0,986*** -	-1,120*** (0,374)	-0,723** (0,316)	-0,723*** -
tv	0,004* (0,002)	0,004 (0,003)	0,004 -	0,003 (0,003)	-0,000 (0,003)	0,000 -	0,003 (0,003)	0,000 (0,003)	0,000 -
carro	0,001 (0,001)	0,001 (0,001)	0,001 -	-0,000 (0,001)	-0,000 (0,001)	0,000 -	0,002 (0,001)	0,002 (0,001)	0,002 -
indigent	-0,000 (0,002)	0,001 (0,002)	0,001 -	0,006* (0,003)	0,003 (0,002)	0,003 -	0,005* (0,003)	0,003 (0,002)	0,003 -
renda	0,000 (0,000)	0,000 (0,000)	0,000 -	0,001*** (0,000)	0,000** (0,000)	0,0003** -	0,000** (0,000)	0,000** (0,000)	0,000 -
5quartil eletrica	-0,007* (0,004)	-0,007* (0,004)	-0,007 -	-0,002 (0,006)	-0,003 (0,004)	-0,003 -	-0,002 (0,005)	-0,003 (0,004)	-0,003 -
privada	0,057 (0,040)	0,091* (0,054)	0,091* -	-0,158*** (0,058)	-0,098** (0,047)	-0,098** -	-0,177*** (0,054)	-0,129*** (0,046)	-0,129*** -
inst_san	0,248 (0,406)	0,283 (0,260)	0,283 -	0,770*** (0,244)	0,455* (0,263)	0,455* -	0,708*** (0,253)	0,444* (0,260)	0,444* -
cobert	-0,002** (0,001)	-0,002*** (0,001)	-0,002*** -	0,001 (0,001)	0,001* (0,001)	0,001* -	0,001 (0,001)	0,000 (0,001)	0,000 -
vacin	-0,004* (0,002)	-0,004 (0,003)	-0,004 -	-0,007 (0,004)	-0,007** (0,003)	-0,007** -	-0,006 (0,004)	-0,006** (0,003)	-0,006** -
mortal	0,000 (0,000)	0,001 (0,001)	0,001 -	-0,000 (0,000)	0,000 (0,000)	0,000 -	0,000 (0,000)	0,000 (0,000)	0,000 -
cau_ext	0,002 (0,003)	0,003 (0,004)	0,003 -	-0,013*** (0,005)	-0,012*** (0,004)	-0,012*** -	-0,014*** (0,005)	-0,013*** (0,003)	-0,013*** -
temp	0,000 (0,000)	0,000 (0,000)	0,000 -	0,001 (0,000)	0,000 (0,000)	0,000 -	0,001 (0,000)	0,000 (0,000)	0,000 -
plum	0,000 (0,000)	0,000 (0,000)	0,000 -	0,001 (0,000)	0,000 (0,000)	0,000 -	0,001 (0,000)	0,000 (0,000)	0,000 -
ano	0,036*** (0,011)	0,041*** (0,013)	0,041*** -	0,005 (0,014)	-0,005 (0,013)	-0,005 -	0,005 (0,013)	-0,004 (0,013)	-0,004 -
direita	0,036*** (0,010)	0,044*** (0,012)	0,044*** -	0,026** (0,013)	0,013 (0,011)	0,013 -	0,022* (0,012)	0,011 (0,011)	0,011 -
esquerda	0,691 (0,493)	0,539 (0,461)	0,539 -	0,339 (0,592)	0,846* (0,464)	0,846* -	0,352 (0,572)	0,778* (0,459)	0,778* -
_cons	0,337			0,217			0,21		
Adj R ²	561	553	553	561	554	554	561	554	554
Num obs									

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

efeito significativo apresentaram, quase todas, o efeito com sinal esperado.¹³ A maior porcentagem de indigentes e de analfabetos está associada a maior gasto *per capita* e a maior renda no quintil mais rico da população está associada a menor gasto *per capita*. Contudo, maior renda *per capita* está associada a maior gasto *per capita*.

Além disso, os dois únicos resultados inesperados são:

- (i) menor taxa de incidência de doenças infecciosas em municípios com mais crianças e
- (ii) menor gasto *per capita* em municípios com mais idosos.

Tais resultados devem estar relacionados a algum viés de omissão. Contudo, sem dados em painel, que permitam controlar o efeito fixo municipal, não é possível dar uma resposta definitiva para essa questão.

Por fim, observa-se que tanto as prefeituras de direita quanto de esquerda apresentam desempenho significativamente melhor que as de centro, tanto para acesso quanto para prevenção de doenças infecciosas (lembrando que maior incidência de doenças infecciosas indica pior desempenho). Conforme dito anteriormente, na ideologia de centro foram classificados os partidos PMDB, PSC, PSL e PRB, mas apenas o primeiro detinha mais de cinco prefeituras no período analisado. Esses resultados estão mais uma vez em acordo com o trabalho de Sampaio de Souza et alii (2005), no qual encontrou-se que o PMDB e o PDT estão associados a menor eficiência na provisão de bens públicos.

Resumindo, observamos que população afeta negativamente (e estatisticamente) o insumo (gasto com saúde) e positivamente e estatisticamente somente um produto (taxa de acesso ao sistema). Estes efeitos combinados sobre insumo e produto são consistentes com o sinal estimado da variável população quando as variáveis dependentes são os *scores* de eficiência.

7. Conclusões

Para atribuir a competência de gasto entre as diferentes esferas de Governo, a literatura de federalismo fiscal parte do princípio de que bens e serviços públicos devem ser ofertados pelo nível de Governo mais próximo da população. Portanto, a provisão descentralizada poderia aumentar o bem estar social, aumentando a variedade e qualidade dos bens e serviços oferecidos a partir do entendimento das diferentes preferências e custos locais.

Por outro lado, a oferta pulverizada de alguns tipos de bens e serviços pode levar a significativas perdas de escala. Um argumento a favor da centralização seria, então, o de que a produção de alguns serviços exige escala de produção incompatível com a demanda do município. Por exemplo, a existência de muitos municípios com população pequena faz com que não se justifique a oferta por parte destes de todos os níveis de complexidade dos serviços de saúde.

¹³ Ver tabela do Apêndice, conforme já foi destacado na subseção anterior.

Tabela 5

Resultados do modelo SUR para taxa de acesso ao sistema, taxa de internações por doenças infecciosas e gasto *per capita* em saúde

Variáveis	Taxa de acesso	Taxa de internações	Gasto <i>per capita</i>
	ao sistema	por doenças infecciosas	em saúde
População (log)	0.005*	-0.096	-9.014**
	(0.003)	(0.138)	(3.677)
Controles demográficos			
Concentração populacional	0.000	0.000**	0.000***
	(0.000)	(0.000)	(0.000)
População urbana	0.131***	7.418***	-222.343***
	(0.037)	(1.826)	(48.555)
Idosos	-0.004	-0.070	-13.687***
	(0.003)	(0.146)	(3.885)
Crianças	-0.002	-1.094***	-5.378
	(0.005)	(0.268)	(7.118)
Controles de renda			
Renda p.c.	0.000	-0.036*	1.575***
	(0.000)	(0.018)	(0.477)
Gini	-0.039	8.075	-129.328
	(0.219)	(10.910)	(290.071)
TV	0.004**	-0.019	0.589
	(0.002)	(0.090)	(2.396)
Carro	0.000	0.044	-1.480
	(0.001)	(0.039)	(1.025)
Indigentes	0.001	-0.011	3.279
	(0.002)	(0.078)	(2.081)
Renda 5o. quintil	-0.000	0.007	-0.321**
	(0.000)	(0.006)	(0.150)
Eletricidade	-0.004	0.071	4.681
	(0.003)	(0.143)	(3.806)
Saúde privada	0.089***	-0.616	75.632*
	(0.032)	(1.617)	(42.985)
Controles para saneamento e vacinação			
Instalação sanitária	0.336*	-9.538	223.062
	(0.182)	(9.057)	(240.806)
Cobertura vacinação	-0.001	-0.011	2.783***
	(0.001)	(0.026)	(0.689)
Controle educacional			
Analfabetismo	-0.001	-0.266***	6.911***
	(0.002)	(0.094)	(2.512)
Controle violência			
Mortes por causas externas	0.000	-0.001	-0.795*
	(0.000)	(0.017)	(0.456)
Controles climáticos			
Temperatura	0.006**	0.168	13.049***
	(0.002)	(0.121)	(3.215)
Chuvas	0.000	-0.008	0.045
	(0.000)	(0.009)	(0.240)
Controles ideológicos			
Direita	0.034***	-0.880**	-3.182
	(0.009)	(0.439)	(11.675)
Esquerda	0.027***	-0.785**	-14.332
	(0.008)	(0.389)	(10.352)

Número de observações: 561.

*, **, ***: Variável significativa a 10%, 5% e 1% respectivamente.

Este artigo procurou avaliar a presença de economias de escala na oferta de serviços públicos de saúde dos municípios paulistas, com o objetivo de avaliar os possíveis efeitos do processo de descentralização dos serviços de saúde a partir da consolidação do Sistema Único de Saúde (SUS). Por um lado, a descentralização pode implicar aumento da qualidade e do acesso aos serviços de saúde. Por outro lado, pode trazer ineficiências devido à perda de escala.

Para realizar essa análise, os municípios paulistas foram ordenados de acordo com os seus gastos em saúde a partir da construção de um indicador simples de eficiência relativa, que leva em consideração o desempenho na provisão de serviços de saúde e o gasto *per capita* nesse setor. O desempenho na provisão de serviços de saúde foi medido a partir de duas dimensões: *acesso da população ao sistema e prevenção de doenças infecciosas*. O indicador de eficiência de gastos foi então construído pela técnica do FDH (Free Disposable Hull). Além de medir a extensão da ineficiência nos gastos com saúde, procurou-se estimar o efeito das variáveis de escala não só sobre a eficiência relativa dos municípios, mas também diretamente sobre os indicadores de desempenho e o gasto *per capita*, separadamente.

Municípios maiores são mais eficientes na provisão de bens públicos de saúde, ao menos no que se refere às dimensões analisadas neste estudo. Há um claro efeito negativo da escala municipal sobre o gasto *per capita* em saúde, mesmo controlando a taxa de urbanização da população, que também exerce efeito negativo sobre esses gastos. Além disso, observa-se que municípios maiores oferecem melhores condições de acesso ao sistema de saúde para seus habitantes, novamente controlando pela taxa de urbanização, que também exerce efeito positivo no desempenho em termos de acesso ao sistema.

Municípios maiores são mais eficientes na provisão de acesso aos serviços de saúde. Há um claro efeito negativo da escala municipal sobre o gasto *per capita* em saúde, mesmo controlando a taxa de urbanização da população, que também exerce efeito negativo sobre esses gastos. Além disso, observa-se que municípios maiores oferecem melhores condições de acesso ao sistema de saúde para seus habitantes, novamente controlando pela taxa de urbanização, que também exerce efeito positivo no desempenho em termos de acesso ao sistema.

Levando em consideração que os municípios maiores tendem a concentrar maior variedade e sofisticação de serviços de saúde, os resultados deste estudo sugerem que os acordos feitos entre pequenos e grandes municípios para que os primeiros possam compartilhar tratamentos oferecidos nos outros não garantem igual acesso aos cidadãos de ambos. Fica a dúvida se uma maior centralização da gestão poderia melhorar essa situação.

Observou-se também que o tamanho da população não tem efeito significativo direto sobre incidência de doenças infecciosas, mas o coeficiente tem sinal negativo, associando positivamente, portanto, escala e bom desempenho em prevenção de doenças. É certo, portanto, que pequenos municípios não têm melhor desempenho que os grandes na prevenção de doenças infecciosas, sendo que as políticas de prevenção fazem parte das políticas de responsabilidade primordialmente municipal e que as prefeituras de municípios menores têm, a princípio, condições de conhecer

melhor as necessidades de seus cidadãos. Assim sendo, pode-se dizer que a descentralização excessiva na gestão da prevenção de doenças infecciosas não melhora o desempenho nessa área de atuação.

Em consonância com os resultados de Sampaio de Souza et alii (2005), os resultados deste estudo sugerem fortemente que a excessiva descentralização na gestão pública da saúde pode levar à perda de eficiência na oferta de serviços de saúde, aumentando os gastos sem melhorar, aparentemente, a qualidade. Uma forma de consolidar e aprofundar esses resultados seria construir e trabalhar com novas medidas de desempenho na área de serviços de saúde.

Por fim, foi observado que tanto as prefeituras de direita quanto de esquerda apresentam desempenho significativamente melhor que as de centro – ideologia em que foram classificados os partidos PMDB, PSC, PSL e PTB, mas apenas o primeiro e o último detinham mais de cinco prefeituras no período analisado. Esses resultados estão mais uma vez em acordo com o trabalho de Sampaio de Souza et alii (2005), no qual encontrou-se que o PMDB e o PDT estão associados a menor eficiência na provisão de bens públicos.

Referências bibliográficas

- Afonso, A., Schuknecht, L., & Tanzi, V. (2003). Public sector efficiency: An international comparison. Working Paper 242, European Central Bank.
- Afonso, A. & St. Aubyn, M. (2004). Non-parametric approaches to education and health: Expenditure efficiency in OECD countries. Mimeo. Technical University of Lisbon.
- Aigner, D., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6:21–37.
- Alesina, A. & Spolaore, E. (1997). On the number and size of nations. *The Quarterly Journal of Economics*, 112:1027–56.
- Bartelsman, E. & Doms, M. (2000). Understanding productivity: Lessons from longitudinal microdata. *Journal of Economic Literature*, 38:569–594.
- Bolton, P. & Roland, G. (1997). The breakup of nations: A political economy analysis. *The Quarterly Journal of Economics*, 112:1057–90.
- Brunet, J. F. G., Borges, C., Bertê, A., Busatto, L., & Brunet, L. (2006). Estados comparados por funções do orçamento. Uma avaliação da eficiência e efetividade dos gastos públicos estaduais. Menção Honrosa I, Prêmio IPEA-Caixa.
- Coelli, T. J., Rao, D. S. P., & Battese, G. (1998). *An Introduction to Efficiency and Productivity Analysis*. Kluwer Academic Publishers, Boston.
- Coppedge, M. (1997). A classification of Latin America political parties. The Hellen Kellog Institute. Working Paper 244.
- Deprins, D., Simar, L., & Tulkens, H. (1984). Measuring labor inefficiency in post offices. In Marchand, M., Pestieau, P., & Tulkens, H., editors, *The Performance of Public Enterprises: Concepts and Measurements*, pages 243–267. North-Holland, Amsterdam.
- Evans, D., Tandon, A., Murria, C. J. L., & Lauer, J. A. (2000). The comparative efficiency of national health systems in producing health: An analysis of 191 countries. World Health Organization. GPE Discussion Paper Series 29.
- Galasso, E. & Ravallion, M. (2001). Decentralized targeting of an anti-poverty program.

- Mimeo, Development Research Group, World Bank.
- Greene, W. (2000). A gamma distributed stochastic frontier model. *Journal of Econometrics*, 46:141–164.
- Gupta, S. & Verhoeven, M. (2001). The efficiency of government expenditures experiences from Africa. *Journal of Policy Modeling*, 23:433–467.
- Herrera, S. & Pang, G. (2005). Efficiency of public spending in developing countries: An efficiency frontier approach. Mimeo.
- Kumbhakar, S. & Lovell, K. (2000). *Stochastic Frontier Analysis*. Cambridge University Press, Cambridge.
- Marinho, A. (2003). Avaliação da eficiência técnica nos serviços de saúde dos municípios do Estado do Rio de Janeiro. *Revista Brasileira de Economia*, 57:515–534.
- Meeusen, W. & Van de Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18:435–444.
- Mendes, C. C. & Sampaio de Souza, M. C. (2006). Estimando a demanda por serviços públicos nos municípios brasileiros. *Revista Brasileira de Economia*, 60:281–296.
- Miranda, R. B. (2006). Uma avaliação da eficiência dos municípios brasileiros na provisão de serviços públicos usando ‘data envelopment analysis’. Boletim de Desenvolvimento Fiscal, IPEA.
- Morroni, M. (2006). *Knowledge, Scale and Transactions in the Theory of the Firm*. Cambridge University Press.
- Oates, W. (1972). *Fiscal Federalism*. Hartcourt, New York.
- Person, T. & Tabellini, G. (2000). *Political Economics*. The MIT Press, Cambridge, Massachusetts.
- Pitt, M. & Lee, L. (1981). The measurement and sources of technical inefficiency in Indonesian weaving industry. *Journal of Development Economics*, 9:43–64.
- Sampaio de Souza, M. C., Cribari-Neto, F., & Stosic, B. D. (2005). Explaining DEA technical efficiency scores in an outlier corrected environment: The case of public services in Brazilian municipalities. *Brazilian Review of Econometrics*, 25:289–315.
- Simar, L. & Wilson, P. W. (2003). Estimation and inference in two-stage, semiparametric models of production processes. Discussion Paper 0307, Institut de Statistique, UCL, Belgium.
- Vickrey, W. S. (1969). Papers and proceedings of the eighty-first. *The American Economic Review*, 59:251–260.

Apêndice

Tabela A.1 – Variável de interesse e controles

Variável	Descrição	Ano	Fonte
log pop	log da população	2005	IBGE
Controles demográficos			
conpop	Concentração populacional	2005	IBGE
popurb	População urbana (%)	2000	IBGE
idosos	Pessoas com 65+ anos (%)	2000	IBGE
criança	Pessoas com 10- anos (%)	2000	IBGE
Controles de renda			
rendapc	Renda <i>per capita</i>	2000	IBGE
gini	Índice de gini	2000	IBGE
tv	Domicílios com TV (%)	2000	IBGE
carro	Pessoas com carro (%)	2000	IBGE
indigentes	Porcentagem indigentes	2000	IBGE
renda5quartil	Renda média domiciliar <i>per capita</i> do quintil mais rico da distribuição	2000	IBGE
eletrica	Domicílios com eletricidade (%)	2000	IBGE
saude privada	Total de beneficiários de planos de saúde dividido pela população total	2005	IBGE
Controles sanitários e vacinação			
cobert_vac	Cobertura da imunização, considerando a população alvo, para os imunobiológicos e doses definidos pelo ministério da saúde	2005	Datasus
inst_sanit	Domicílios com saneamento (%)	2000	IBGE
Controles educacionais			
analfa	Pessoas analfabetas (%)	2000	IBGE
Controles violência			
mort_causa_ext	Mortalidade devido a causas externas, definida pelo CID-10, dividido pela população ($\times 1000$)	2005	Datasus
Controles climáticos			
temp_ano	Temperatura média anual	2000	IBGE
plum_ano	Índice pluviométrico médio anual	2000	IBGE
Controles ideologia			
direita	Partido do prefeito é de direita	2004	Coppedge (1997)
esquerda	Partido do prefeito é de esquerda	2004	Coppedge (1997)

Tabela A.2 – Sinal esperado

Variável	<i>Scores</i>	Acesso	Internações	Gasto
logpop	?	?	?	-
conpop*	+	+	?	-
popurb*	+	+	?	?
idosos	-	?	+	+
criança	?	?	+	?
rendapc	+	+	-	+
gini	-	-	+	?
tv	+	+	-	+
carro	+	+	-	+
indigentes	-	-	+	+
renda5quartil	+	+	-	-
eletrica	+	+	-	-
saude privada	+	+	-	-
cobert_vac	+	+	-	?
Inst_sanit	+	+	-	-
analfa	-	-	+	+
mort_causa_ext	-	?	?	+
temp_ano	-	?	+	+
plum_ano	-	?	+	+
direita	?	?	?	?
esquerda	?	?	?	?

* para *score* de prevenção, negativo.