

## EFICIÊNCIA DO SETOR HOSPITALAR NOS MUNICÍPIOS PAULISTAS

IGOR VIVEIROS SOUZA\*  
MARISLEI NISHIJIMA†  
FABIANA ROCHA‡

### Resumo

O objetivo deste artigo é avaliar o grau de eficiência produtiva do setor público hospitalar em 366 municípios do estado de São Paulo entre os anos de 1998 e 2003. Para tanto é utilizado o método de fronteira estocástica de produção. O modelo estimado com a forma flexível de Fourier usa o complemento da taxa de mortalidade hospitalar como produto e o gasto público com profissionais e o número de leitos por município como insumos. Os resultados sugerem que os municípios mais eficientes são aqueles que contratam mais leitos de hospitais privados, que realizam o maior número de internações (efeito de economia de escala), que possuem menor população (efeito de congestionamento) e que apresentam menor tempo médio de internação.

**Palavras-chave:** hospitais, eficiência, fronteira estocástica, municípios.

### Abstract

The purpose of this article is to assess the degree of productive efficiency of public sector hospitals in 366 municipalities of the state in São Paulo between the years 1998 and 2003. In order to do so we use the stochastic frontier of production approach. The model is estimated using the flexible form of Fourier and uses the complement of the rate of hospital mortality as output and public spending with professionals and the number of beds per municipality as inputs. The results suggest that the most efficient municipalities are those that hire more beds in private hospitals, the ones which perform the highest number of admissions (economies of scale effect), the ones with smaller population (congestion effect), and the ones which show lower average time of hospitalization.

**Keywords:** hospitals, efficiency, stochastic frontier, municipalities.

**JEL classification:** C23, H51, I10

---

\* Fundação João Pinheiro

† EACH/USP

‡ FEA/USP

## 1 Introdução

O setor de saúde brasileiro é um dos grandes demandantes de recursos públicos. Para se ter uma idéia o Governo despendeu uma média anual de 3,23% de seu PIB em saúde entre os anos de 1999 e 2006, conforme estatísticas do Banco Mundial. Em 2003, somente o estado de São Paulo teve um orçamento para o setor de saúde, considerando as três esferas de Governo, de aproximadamente 13,4 bilhões de reais, o que implica um gasto *per capita* em torno de R\$ 346,21<sup>1</sup>.

Apesar da Constituição de 1988 ter estabelecido como modelo a descentralização na provisão e no financiameanto das ações de saúde, por meio do Sistema Único de Saúde (SUS), visando maior eficiência no uso dos recursos, poucos trabalhos empíricos foram realizados para avaliar o desempenho dos municípios na gestão de tais recursos.

Deste modo, este estudo busca contribuir para a redução desta lacuna através da investigação da eficiência produtiva dos gastos públicos no setor hospitalar dos municípios do estado de São Paulo. O período de análise vai de 1998 a 2003 e abrange somente os municípios de São Paulo por questão de disponibilidade e confiabilidade dos dados. São avaliados 366 municípios o que corresponde a 89% do total de municípios que possuem rede hospitalar pública ou privada contratada. Somente para esses tem-se informação disponível para todas as variáveis investidas em 2003. Além disso, em 2003 a população desses 366 municípios representava 94% da população total do Estado.

A literatura sobre fronteira estocástica aplicada especificamente a hospitais em geral trata de estimar fronteiras de eficiência de custos, em que são utilizados grandes volumes de informação sobre preços e quantidades de insumos e produtos, além de informações específicas sobre os pacientes para estimar relações microeconômicas. Bradford & Kleit (2001) utilizam dados de um hospital para comparar a eficiência produtiva de dois tipos de tratamentos alternativos em pacientes cardíacos. Smet (2007) compara a eficiência entre hospitais na Bélgica, usando uma função multi-produto para avaliar seus desempenhos na presença de uma demanda estocástica (dada a sua grande variabilidade). Bernet et al. (2008) verificam como o acesso ao financiamento de investimentos produtivos de hospitais (equipamentos e instalações) altera o grau de eficiência produtiva de uma amostra de hospitais. Finalmente, Brown III & Pagán (2006) avaliam como o sistema de saúde norte americano Managed Care impacta sobre o custo de uma amostra de hospitais em diferentes localidades.

Considerando-se, contudo, a ausência de informações detalhadas sobre insumos utilizados pelos hospitais públicos brasileiros a alternativa adotada foi estimar modelos de fronteira estocástica de produção, seguindo a literatura de eficiência produtiva de gastos públicos baseada em dados agregados, usando informações hospitalares por município.

A literatura sobre eficiência de gastos públicos é farta em estudos que comparam a eficiência relativa de diferentes países na provisão de saúde (Evans et al. (2000), Gupta & Verhoeven (2001), Jayasuriya & Wodon (2002), Greene (2003a), Afonso & St. Aubyn (2004), Herrera & Pang (n.d.), Herrera & Worldwide (2005)). Afonso et al. (2003), por sua vez, constrõem um indi-

---

<sup>1</sup>Valores de 2003, informação retirada do DATASUS.

gador de desempenho para o setor público como um todo para 23 países da OCDE. Esse 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.

Especificamente para o Brasil, [Sousa et al. \(2005\)](#) que buscam calcular *scores* de eficiência sobre os serviços gerais dos municípios brasileiros e de [Marinho \(2001\)](#), que analisa a eficiência da prestação de serviços de saúde em municípios do Rio de Janeiro.

A fim de estimar a eficiência dos hospitais será utilizado o método de Fronteira Estocástica (FE), mais especificamente o modelo implementado por [Battese & Coelli \(1995\)](#). Esse método permite a decomposição das variações do desempenho dos municípios na provisão de serviços hospitalares em relação à fronteira em variações na eficiência técnica e em choques puramente aleatórios. Além disso, permite a investigação empírica objetiva de prováveis variáveis explicativas do termo ineficiência. O uso desse método se constitui num diferencial analítico em relação aos trabalhos anteriormente listados que majoritariamente utilizam metodologias não paramétricas, quais sejam, o *Free Disposable Hull* (FDH) e a *Data Envelopment Analysis* (DEA).<sup>2</sup>

Além desta introdução, o artigo apresenta três seções. Na segunda seção, apresenta-se a base teórica e quantitativa sobre a qual este estudo repousa. Na terceira seção discute-se a base de dados e os principais resultados obtidos das estimativas. Por fim, na quarta seção resumem-se as principais conclusões.

## 2 Metodologia

Para medir eficiência produtiva a teoria utiliza uma medida de distância entre o ponto de operação da unidade tomadora de decisão (neste caso, o município) e a fronteira tecnológica, medida que pode ser dividida em eficiência técnica e alocativa.<sup>3</sup> O conceito de eficiência técnica, que remonta a [Debreu \(1951\)](#), foi consolidado por [Farrel \(1957\)](#), que, adicionalmente desenvolveu o conceito de eficiência alocativa. O conceito de eficiência técnica diz respeito estritamente às relações entre as quantidades produzidas de produto e as quantidades de fatores utilizados na produção. Quando a quantidade produzida por uma firma, dada uma combinação de fatores, fica aquém do máximo possível de ser atingido com aquela combinação tem-se a caracterização da ineficiência técnica.

A eficiência alocativa - que se refere à escolha ótima da proporção de insumos dado o vetor de preços - não será mensurada neste estudo. De acordo com [Farrel \(1957\)](#), a ineficiência alocativa deve ser estudada quando se objetiva verificar se as firmas agem dentro de um arcabouço lucro-maximizador ou custo-minimizador. Como o objetivo aqui é apenas a obtenção de uma relação técnica da função de produção e, além disso, não se dispõe de informações

<sup>2</sup>[Gupta & Verhoeven \(2001\)](#), por exemplo, usam uma abordagem FDH. [Afonso & St. Aubyn \(2004\)](#), [Herrera & Pang \(n.d.\)](#) e [Herrera & Worldwide \(2005\)](#) usam tanto DEA quanto FDH. Uma exceção do ponto de vista metodológico é [Greene \(2003b\)](#) que estima uma fronteira estocástica para avaliar a eficiência dos gastos em saúde usando os dados do World Health Organization (WHO).

<sup>3</sup>A literatura que trata da eficiência do setor público seguiu de perto a literatura que trata da eficiência do setor privado e, assim, países, estados e municípios são considerados analiticamente como iguais às firmas.

sobre preços, pois se trata de serviços prestados pelo setor público, estima-se apenas um modelo de fronteira de produção.<sup>4</sup>

Para mensurar a eficiência técnica, a literatura nos apresenta técnicas paramétricas e não paramétricas. Dentre as técnicas paramétricas, a mais difundida, e com propriedades amplamente pesquisadas, é a fronteira estocástica de produção. Já dentre as técnicas não paramétricas, tem-se o *Free Disposal Hull* (FDH) e *Data Envelopment Analysis* (DEA). Neste estudo opta-se pelo uso de da fronteira estocástica, pois ela apresenta algumas propriedades que não estão disponíveis nos métodos não paramétricos de acordo com Coelli et al (2005). São elas: (1) a possibilidade de se realizar testes de hipóteses sobre os parâmetros das variáveis explicativas, (2) a possibilidade de se incluir variáveis de controle para explicar a ineficiência técnica em apenas um estágio e (3) permite a presença de ruídos aleatórios no ambiente em que a unidade tomadora de decisão opera<sup>5</sup>.

Embora as fronteiras estocásticas apresentem as vantagens descritas anteriormente, exigem a necessidade de imposição de uma forma funcional *a priori* e de hipóteses acerca da distribuição do termo de ineficiência<sup>6</sup>. Estes são custos que a metodologia DEA não impõe uma vez que tal técnica simplesmente mapeia os pontos dados existentes e traça, sobre os mesmos, um envoltório convexo que contém todos os pontos observados. Dessa forma, a única restrição que o DEA impõe é a convexidade dos conjuntos de produção. Já o FDH, assim como o DEA, não apresenta as propriedades descritas para as fronteiras estocásticas. Adicionalmente, esta técnica descarta a hipótese de convexidade dos conjuntos de produção.

As características positivas supracitadas das fronteiras estocásticas de produção fundamentam a escolha de tal metodologia. No entanto, na tentativa de atenuar os efeitos da imposição de uma forma funcional *a priori* que a escolha de tal método traz consigo, optou-se pelo uso da forma funcional *Flexível de Fourier*<sup>7</sup>.

Não existe na literatura um consenso acerca de qual método é superior. Evidências apontam que ambos produzem, em muitos casos, resultados pouco robustos quando comparados entre métodos. Uma dessas evidências pode ser vista em Jacobs (2000). Em tal estudo, examinam-se as propriedades das fronteiras estocásticas e do DEA, em diversas especificações, para o cálculo de eficiência hospitalar no Reino Unido. Seus resultados apontam para uma robustez interna dos métodos, no tocante aos *rankings* de eficiência, mas não entre métodos.

Conforme exposto, as fronteiras estocásticas de produção permitem a decomposição do resíduo em dois componentes: um relativo à ineficiência e outro relativo ao choque puramente aleatório, chamado de erro idiossincrático. Esse método foi introduzido na literatura econômica simultaneamente por Aigner et al. (1977) e Meeusen & Van de Broeck (1977) numa versão *cross-section* e sua especificação para dados em painel feita por Pitt & Lee (1981) é descrita em (1).

---

<sup>4</sup>Apesar disso, a metodologia aqui aplicada permite afirmar que se uma unidade tomadora de decisão é tecnicamente ineficiente, então, ela não maximiza seu retorno (Kumbhakar e Lovell, 2000).

<sup>5</sup>Métodos não paramétricos incorporam esses ruídos no valor da ineficiência técnica.

<sup>6</sup>Maiores detalhes serão dados a seguir.

<sup>7</sup>Propriedades da forma Flexível de Fourier serão apresentadas mais adiante.

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

em que

$\ln y_{it}$  é o logaritmo da quantidade produzida pelo município  $i$  no período  $t$ ;

$\beta_0$  é o intercepto da equação;

$f(\beta_k, x_i^k)$  é a forma funcional adequada;

$\beta_k$  é o vetor de coeficientes tecnológicos;

$x_{it}^k$  é o vetor de insumos utilizado na produção pelo município  $i$  no período  $t$ ;

$v_{it}$  é o choque aleatório não correlacionado com  $x_{it}^k$  e  $u_{it}$  e com distribuição  $N(0, \sigma_v^2)$ ;

$u_{it}$  é o termo de ineficiência não negativo do município  $i$  no período  $t$  também não correlacionado com  $x_{it}^k$ .

Para a estimação dos coeficientes tecnológicos das fronteiras estocásticas de produção são empregados os estimadores de máxima verossimilhança<sup>8</sup>, sendo necessário assumir hipóteses explícitas sobre a distribuição assimétrica do termo de ineficiência,  $u_{it}$ .

Aqui se utiliza o modelo proposto por Battese & Coelli (1995) que desenvolveram uma metodologia de estimação para painéis desbalanceados utilizando uma distribuição normal-truncada. A escolha desta distribuição resulta do fato de ser mais flexível do que outras formas funcionais mais simples como a semi-normal ou a exponencial.<sup>9</sup>

O modelo especificado por Battese & Coelli (1995) também permite a variação da ineficiência técnica ao longo do tempo, que pode ser modelada usando-se características dos municípios que variam ao longo do tempo. A equação (2), em que o termo de ineficiência segue uma distribuição normal truncada  $u_{it} \sim N^+(z_{it}\delta, \sigma_u^2)$ , mostra o termo  $z_{it}\delta$  que representa a média da ineficiência, sendo esta composta pelo vetor de variáveis específicas dos municípios  $z_{it}$  e  $\delta$  o vetor de coeficientes associados a essas variáveis. Assim, o termo ineficiência é modelado como uma média condicional de um conjunto linear de covariadas pré-especificadas. O termo  $w_{it}$  é uma variável correspondente à truncagem de uma normal com média zero e variância  $\sigma_u^2$  no ponto  $(-z_{it}\delta)$ , ou seja,  $w_{it} \geq (-z_{it}\delta)$ .

$$u_{it} = z_{it}\delta + w_{it} \quad (2)$$

Conforme, já explicitado, escolheu-se a forma funcional *Flexível de Fourier*, que consiste numa aproximação global que inclui os termos padrões de uma trans-log mais os termos trigonométricos de Fourier, sua especificação é dada pela equação (3) abaixo.<sup>10</sup>

<sup>8</sup>Sobre metodologia de estimação de fronteiras estocásticas ver Kumbhakar & Lovell (2000).

<sup>9</sup>Para estimações com distribuição semi-normal ou exponencial ver Aigner, Lovell e Schmidt (1977) e para estimações com distribuições gama ver Greene (2000).

<sup>10</sup>Foram feitos testes estatísticos para justificar a escolha da forma flexível de Fourier que serão apresentados e discutidos na seção 3 que trata dos resultados das estimativas.

$$\begin{aligned}
\ln y_{it} = & \beta_0 + \beta_t t + \beta_{tt} t^2 \sum_{k=1}^n \beta_k \ln x_{it}^k + \sum_{j \leq k} \sum_{k=1}^n \beta_{jk} \ln x_{it}^j \ln x_{it}^k \\
& + \sum_{k=1}^n \beta_{kt} \ln x_{it}^k t + \sum_{k=1}^n [\phi_k \cos(h_k) + \omega_k \text{sen}(h_k)] \\
& + \sum_{k=1}^n \sum_{q=k}^n [\phi_{kq} \cos(h_k + h_q) + \omega_{kq} \text{sen}(h_k + h_q)] + v_{it} - u_{it}, \\
& i = 1, \dots, I; k = 1, \dots, n; t = 1, \dots, T.
\end{aligned} \tag{3}$$

Sendo assim é possível, a partir da estimação da equação acima, testar a adequação tanto de uma trans-log, quanto de uma Cobb-Douglas.

De acordo com Berger & Mester (1997), esta forma funcional é uma aproximação global pelo fato dos termos  $\cos(h_k)$ ,  $\text{sen}(h_k)$ ,  $\cos(h_k + h_q)$  e  $\text{sen}(h_k + h_q)$  serem mutuamente ortogonais no intervalo  $[0; 2\pi]$ , aproximando a função a ser estimada de seu verdadeiro caminho.<sup>11</sup> Segue-se a sugestão dos autores para o corte de 10% de cada cauda do intervalo  $[0, 2\pi]$  evitando-se, desta maneira, problemas de aproximação nas fronteiras do intervalo. Como consequência, os termos  $h_k$  e  $h_q$  são calculados por (4).

$$h_k = 0, 2 \times \pi - \mu \times a + \mu \times x_k. \tag{4}$$

Sendo  $[a, b]$  o intervalo<sup>12</sup> transformado em radianos e  $\mu \equiv (0, 9 \times 2\pi - 0, 1 \times 2\pi)/(b - a)$ . Portanto, para obter as estimativas dos parâmetros da equação acima por máxima verossimilhança, basta maximizar a função de verossimilhança de  $\varepsilon_{it}$ . Para tal, é preciso conhecer a distribuição de  $\varepsilon_{it}$ , sendo  $\varepsilon_{it} = v_{it} - u_{it}$ .

Supondo  $v_{it}$  e  $u_{it}$  independentes e com suas respectivas distribuições conhecidas, a distribuição conjunta  $v_{it}$  e  $u_{it}$  é dada por  $f(v_{it}, u_{it}) = f(u_{it}) \times f(v_{it})$  e desde que  $\varepsilon_{it} = v_{it} - u_{it}$  tem-se que  $f(v_{it}, u_{it}) = f(u_{it} + \varepsilon_{it}, u_{it}) = f(\varepsilon_{it}, u_{it})$ . A função de distribuição conjunta assume a expressão dada em (5).

$$f(\varepsilon, u) = \frac{e^{-\frac{1}{2} \left\{ \left[ \frac{(u - \mu_*)^2}{\sigma_u^2} \right] + \left[ \frac{(\varepsilon - z\delta)^2}{(\sigma_v^2 + \sigma_u^2)} \right] \right\}}}{2\pi \sigma_u \sigma_v \Phi(z\delta/\sigma_u)} \tag{5}$$

onde:

$$\begin{aligned}
\mu_* &= \frac{\sigma_v^2 z\delta - \sigma_u^2 \varepsilon}{\sigma_v^2 + \sigma_u^2} \\
\sigma_*^2 &= \frac{\sigma_u^2 \sigma_v^2}{\sigma_v^2 + \sigma_u^2}
\end{aligned}$$

e  $\Phi(z\delta/\sigma_u)$  é a função de distribuição acumulada da normal padrão avaliada no ponto  $(z\delta/\sigma_u)$ .

Para obter a distribuição do erro idiossincrático  $\varepsilon$ , integra-se a função (5) com respeito a  $u$ :  $f(\varepsilon) = \int_0^\infty f(\varepsilon, u) du$ , para obter (6).

<sup>11</sup>O que lhe confere a flexibilidade destacada anteriormente.

<sup>12</sup>São respectivamente os valores mínimo e máximo de  $x_k$ .

$$f(\varepsilon) = \frac{e^{\left\{-\frac{1}{2} \frac{(\varepsilon+z\delta)^2}{(\sigma_v^2+\sigma_u^2)}\right\}}}{\sqrt{2\pi(\sigma_v^2+\sigma_u^2)}[\Phi(z\delta/\sigma_u)/\Phi(\mu_*/\sigma_*)]}, \quad -\infty \leq \varepsilon \leq \infty \quad (6)$$

As estimativas dos parâmetros são obtidas a partir maximização do logaritmo da função de verossimilhança dada por (7).

$$L(\beta, \delta, \sigma_u, \sigma_v) = -\frac{\ln 2\pi + \ln \sigma_s^2}{2} \sum_{i=1}^N T_i - \frac{1}{2} \sum_{i=1}^N \sum_{t=1}^{T_i} \left( \frac{y_{it} - x_{it}\beta + z_{it}\delta}{\sigma_s^2} \right)^2 - \sum_{i=1}^N \sum_{t=1}^{T_i} [\ln \Phi(d_{it}) - \ln \Phi(d_{it}^*)] \quad (7)$$

onde:

$$\begin{aligned} d_{it} &= \frac{z_{it}\delta}{\sigma_s^2 \sqrt{\gamma}} & d_{it}^* &= \frac{\mu_{it}^*}{\sigma_s \sqrt{\gamma(1-\gamma)}} \\ \mu_{it}^* &= (1-\gamma)z_{it}\delta - \gamma(y_{it} - x_{it}\beta) & \sigma_s^2 &\equiv \sigma_u^2 + \sigma_v^2 \\ \gamma &\equiv \frac{\sigma_u^2}{\sigma_s^2}. \end{aligned}$$

A parametrização (7) permite verificar a relevância do termo de ineficiência. Se o termo  $\gamma$  converge para 1 o termo de ineficiência predomina o erro idiossincrático e se  $\gamma$  converge para zero a ineficiência torna-se irrelevante para explicar o termo  $\varepsilon$ . No último caso o emprego da técnica de FE não traz ganhos em relação ao método de mínimos quadrados ordinários<sup>13</sup>. Deste modo, caso não se rejeite a hipótese de ineficiência técnica, obtêm-se as estimativas destas a partir da distribuição condicional de  $u_{it}$  em  $\varepsilon_{it}$  dada pela expressão (8).

$$E(e^{u_{it}}|\varepsilon_{it}) = e^{(-\mu_* + \sigma_*^2/2)} \frac{\Phi\left(\frac{\mu_*}{\sigma_*}\right) - \sigma_*}{\Phi\left(\frac{\mu_*}{\sigma_*}\right)} \quad (8)$$

A partir da estimação da ineficiência é possível calcular os *scores* de eficiência que permitem ordenar os municípios de acordo com seu desempenho relativo. A preocupação, contudo, aqui é com os determinantes de ineficiência e não com a distância dos mesmos em relação à fronteira.

### 3 Dados e resultados econométricos

A amostra é composta por 366 municípios paulistas em painel abrangendo os anos de 1998 a 2003, o que corresponde a 89% do total de municípios com rede hospitalar disponível. Além disso, a população desses 366 municípios corresponde a 94% dos residentes do Estado de São Paulo em 2003.

<sup>13</sup>A aplicação de mínimos quadrados parte do pressuposto de que não existe correlação entre o termo residual e a matriz de variáveis independentes X. Se esta hipótese não puder ser sustentada, então torna-se necessário o uso de estimadores de efeito fixo ou mínimos quadrados de dois estágios.

Tabela 1: Variáveis utilizadas

Variáveis	Obs.	Média	Desvio padrão	Mín.	Máx.
Complemento mortalidade hospitalar	2135	97.2988	1.85975	76.92	100
Pessoal	2129	440191	2886995	23.57	$6,77 \times 10^7$
Leitos	2135	274.87	1345.44	6	28091.6
Tendência (T)	2135	3.46042	1.70676	1	6
População	2135	95696.8	558698	3320	$1,07 \times 10^7$
Internações	2135	6202.09	28076.5	2	561530
Gestão	2135	0.32758	0.44198	0	1
PMDB	2135	0.16815	0.37409	0	1
PSDB	2135	0.27681	0.44753	0	1
PT	2135	0.06464	0.24594	0	1
PFL	2135	0.13068	0.33713	0	1
PTB	2135	0.10258	0.30348	0	1
Demais partidos	2135	0.25714	0.43716	0	1
Internações hospitais públicos	2135	0.21261	0.39081	0	1
Internações hospitais privados	2135	0.76802	0.39903	0	1
Internações hospitais universitários	2135	0.01937	0.10904	0	0.93467
Internações cirurgia clínica	2135	0.15145	0.11261	0	0.61576
Internações obstetrícia	2135	0.19290	0.13439	0	1
Internações clínica médica	2135	0.49633	0.20742	0	1
Internações pediatria	2135	0.12503	0.08572	0	0.75521
Outras internações	2135	0.03429	0.11224	0	1
Gasto público	2134	268.59	510.56	63.36	20722.0
Permanência	2129	6.75514	23.2544	1.3	956.4

Fonte: DATASUS e IBGE

O critério de seleção foi a existência de rede hospitalar no município (SUS e particular contratado<sup>14</sup>) e a disponibilidade de informações. Supõe-se que pacientes de localidades onde não exista rede hospitalar se desloquem para aqueles municípios que possuem hospital, fenômeno este que também pode ocorrer devido às especialidades disponíveis em cada município.

Para estimação da função de produção hospitalar dos municípios paulistas considerou-se como produto ( $ly$ ) o logaritmo da variável complemento da taxa de mortalidade hospitalar por município (100 - taxa de mortalidade hospitalar)<sup>15</sup>. Ainda que por questões de indisponibilidade de dados siga-se a literatura de eficiência de gastos públicos em saúde, esta utiliza como medida de produto a mortalidade infantil e a expectativa de vida. Como aqui a preocupação é com a eficiência dos hospitais de procura-se utilizar uma medida que capte a qualidade do atendimento hospitalar do SUS. Quanto menos pessoas morrem nos hospitais teoricamente melhor é a qualidade do atendimento hospitalar.

Para mensurar as variáveis explicativas, trabalho ( $lw$ ) e capital ( $lk$ ), foram consideradas como *proxies* respectivamente, o valor dos serviços de profissionais empregados nas internações hospitalares e o número de leitos contratados pelo SUS em cada município. A variável de natureza monetária ( $lw$ ) foi deflacionada pelo IPCA e está expressa em reais de 1998.

Todas as informações sobre hospitais foram obtidas junto ao Datasus e as informações complementares foram obtidas junto ao IBGE e estão resumidas na Tabela 1.

<sup>14</sup>O SUS aluga a capacidade produtiva do setor privado.

<sup>15</sup>Usou-se o completo para tornar mais fácil a leitura das estimativas da função de produção.



A Tabela 1 apresenta inicialmente as variáveis utilizadas na função de produção: taxa de sobrevivência, que corresponde ao complemento da variável taxa de mortalidade hospitalar; o valor de gastos com profissionais ligados diretamente às internações como *proxy* do insumo trabalho, o número de leitos (do SUS e contratados) como *proxy* do insumo capital e uma variável que descreve uma tendência no tempo (T). As demais variáveis são usadas como “controle”<sup>16</sup> e, assim, aparecem como variáveis explicativas do termo ineficiência estimado. Foram usadas as seguintes variáveis explicativas por município: população; número de internações; gestão, variável com valores compreendidos entre 0 e 1 que descreve o percentual de internações realizadas sob gestão plena do município<sup>17</sup>; *dummies* de partidos que estavam à frente da prefeitura durante os anos de 1998 a 2003; variáveis que descrevem o percentual das internações que ocorreu em leitos públicos (do próprio SUS), privados contratados ou em hospitais universitários, portanto, variando entre 0 e 1; variáveis que descrevem os percentuais de internações realizadas em cirurgia clínica, obstetrícia, clínica médica, pediatria e demais tipos de internação<sup>18</sup>; o valor médio do gasto público por internação; e o tempo médio de permanência na internação.

Considerando que as variáveis usadas como insumos não estão em valores per capita ou por quantidade de internações é preciso controlar pelo tamanho da população e pelo número de internações para tornar compatíveis os resultados.

As *dummies* de partido e a variável de gestão municipal plena buscam captar efeitos da forma de gestão sobre a eficiência produtiva do setor, ou seja, verificar se a gestão descentralizada nas prefeituras é mais eficiente que a gestão mais central por parte dos Estados.

As variáveis que descrevem os percentuais de internações em leitos públicos, privados e universitários buscam inferir se existem efeitos diferenciados entre a provisão de internações públicas diretamente pelo setor público e a produzida pelo setor privado. Já as variáveis de percentuais de tipos de internação buscam reduzir o grau de homogeneidade da informação sobre taxa de mortalidade hospitalar, uma vez que traz informações sobre o tipo de tratamento recebido na internação, controlando assim por diferenças possíveis de tratamentos envolvidos na internação, como por exemplo diferentes graus de complexidade.

A variável valor médio da internação busca controlar gastos diferenciados sobre a taxa de sobrevivência hospitalar e a variável tempo médio de internação os tempos diferenciados dos tratamentos. Além disso, um termo de tendência foi incluído no vetor Z a fim de verificar como a média da eficiência técnica se comporta no tempo, isso é, se os municípios estão, em média, convergindo ou divergindo da fronteira.

A Tabela 2 mostra as estimativas<sup>19</sup> da fronteira estocástica para os hospi-

<sup>16</sup>Colocamos entre aspas para lembrar que estas variáveis não entram diretamente na função de produção que em nosso caso é utilizada a forma flexível de Fourier. A especificação colocando-as como controle, entretanto, não alterou a direção (sinais dos coeficientes) dos valores das estimativas.

<sup>17</sup>A alternativa compreende internações realizadas sob gestão estadual ou composição de gestão em que o município gere apenas a atenção básica em saúde.

<sup>18</sup>Note que tanto estas variáveis como as anteriores equivalem a um conjunto de *dummies* centradas, ou seja, seus valores somam 1, mas não são compostas exclusivamente de zeros e uns.

<sup>19</sup>As estimativas foram obtidas utilizando o pacote Stata.

tais paulistas.

O modelo (1) descreve o modelo inicialmente considerado para a análise. Os modelos de (2) a (5) atestam a robustez dos resultados obtidos, uma vez que consistem nas estimativas do modelo (1) com a retirada consecutiva dos “controles” da ineficiência.

Na primeira parte da tabela estão os resultados das estimativas da função de produção usando a forma flexível de Fourier. Assim, além das variáveis de trabalho (Pessoal) e capital (Leitos) tem-se o produto do logaritmo do capital e do trabalho (Pessoal $\times$ leitos), o tempo ao quadrado (Tendência<sup>2</sup>), o quadrado do logaritmo do trabalho (Pessoal<sup>2</sup>), o quadrado do logaritmo do capital (Leitos<sup>2</sup>), o tempo vezes o logaritmo do trabalho (Tempo $\times$ peessoal), o tempo vezes o logaritmo do capital (Tempo $\times$ leitos) que são os termos tradicionais de uma função trans-log. Os termos  $f\_cos\_w$ ,  $f\_cos\_k$ ,  $f\_sen\_w$ ,  $f\_sen\_k$ ,  $f\_cos\_2w$ ,  $f\_cos\_2k$ ,  $f\_cos\_wk$ ,  $f\_sen\_2w$ ,  $f\_sen\_2k$  e  $f\_sen\_wk$  são os termos de Fourier.

Os resultados da Tabela 2 sugerem, como esperado, que capital e trabalho afetam positivamente a produção de sobrevivência hospitalar. Atestam também que o capital ao quadrado, que mede uma grande capacidade produtiva instalada, proporciona menor taxa de mortalidade hospitalar. As significâncias estatísticas dos vários termos de Fourier atestam a escolha adequada desta forma funcional.

Na Tabela 3 são apresentados os coeficientes estimados das variáveis explicativas da ineficiência. Embora as estimativas da função de produção e da ineficiência tenham sido feitas conjuntamente, optou-se por apresentar os resultados separadamente para facilitar a análise.

Inicialmente é importante observar que a hipótese de inexistência da ineficiência pode ser rejeitada. A estimativa gama é muito próxima da unidade, indicando a predominância do termo ineficiência<sup>20</sup> sobre a variância do erro idiossincrático.

Testes de adequação das formas funcionais também foram realizados, indicando que a escolha da forma funcional Flexível de Fourier é adequada (Tabela 4). Outro teste reportado nessa mesma tabela é o teste conjunto nos coeficientes que explicam a ineficiência. O objetivo de tal teste é verificar se a ineficiência é melhor descrita por uma constante do que condicionada a características individuais que poderiam explicá-las. Assim, as diferenças observadas entre firmas na ineficiência, seriam aleatórias. Esse teste também apresenta o resultado favorável do teste à modelagem da ineficiência. Assim sendo o conjunto de características descrito anteriormente parece ser eficaz na determinação do valor da ineficiência técnica.

Os resultados das estimativas do termo ineficiência mostram que quanto maior a população do município maior é a ineficiência do setor hospitalar. Esse resultado pode refletir um efeito congestionamento. As indivisibilidades características da produção de certos serviços de saúde (por exemplo, serviços de alta complexidade) restringiriam a provisão desses serviços aos grandes centros urbanos, implicando um custo de congestionamento. Já com relação ao número de internações realizadas por um município quanto maior é este

---

<sup>20</sup>A estatística LR não pode ser calculada para o teste conjunto de todos os coeficientes que compõem o termo de ineficiência porque não houve convergência dados os valores iniciais. Seus valores críticos estão em Koode e Palm (1986). Entretanto, estimando por MQO, o teste de Coelli comprova a ineficiência dos resíduos.

Tabela 2: Estimativas da função de produção

	(1)	(2)	(3)	(4)	(5)
Pessoal	0.0135*** (0.00453)	0.0147*** (0.00452)	0.0156*** (0.00453)	0.0170*** (0.00448)	0.0196*** (0.00437)
Leitos	0.0189*** (0.00453)	0.0187*** (0.00452)	0.0179*** (0.00453)	0.0178*** (0.00448)	0.0188*** (0.00437)
Tendência	0.000234 (0.00157)	0.000409 (0.00157)	0.00058 (0.00157)	0.000758 (0.00156)	0.000884 (0.00155)
Tendência <sup>2</sup>	-0.00015 (0.00009)	-0.000144 (0.00009)	-0.000154* (0.00009)	-0.000158* (0.00009)	-0.000148 (0.00009)
Pessoal×leitos	-0.00578*** (0.00093)	-0.00572*** (0.00093)	-0.00566*** (0.00093)	-0.00576*** (0.00092)	-0.00536*** (0.00090)
Pessoal <sup>2</sup>	0.000274 (0.00033)	0.000198 (0.00033)	0.000149 (0.00033)	9.83 × 10 <sup>-5</sup> (0.00033)	-0.000106 (0.00032)
Leitos <sup>2</sup>	0.00506*** (0.00075)	0.00504*** (0.00075)	0.00503*** (0.00074)	0.00517*** (0.00074)	0.00459*** (0.00073)
Tempo×pessoal	0.000139 (0.00021)	0.000179 (0.00021)	0.000158 (0.00021)	0.000134 (0.00020)	0.000117 (0.00020)
Tempo×Leitos	-0.000391 (0.00027)	-0.000513* (0.00027)	-0.000487* (0.00027)	-0.000480* (0.00026)	-0.000479* (0.00026)
F_cos_w	-1.017*** (0.29700)	-1.063*** (0.29500)	-1.082*** (0.29300)	-1.167*** (0.29300)	-1.278*** (0.29300)
f_sen_w	1.088*** (0.31000)	1.136*** (0.30800)	1.156*** (0.30600)	1.249*** (0.30600)	1.368*** (0.30600)
f_cos_k	0.0273** (0.01300)	0.0297** (0.01280)	0.0308** (0.01270)	0.0341*** (0.01260)	0.0311** (0.01260)
f_sen_k	-0.0322 (0.02130)	-0.0335 (0.02110)	-0.0342 (0.02100)	-0.0379* (0.02100)	-0.0386* (0.02190)
f_cos_wk	-0.0522* (0.03020)	-0.0533* (0.03000)	-0.0543* (0.02980)	-0.0594** (0.02960)	-0.0615** (0.03010)
f_sen_wk	-0.0784*** (0.02580)	-0.0816*** (0.02550)	-0.0833*** (0.02530)	-0.0879*** (0.02520)	-0.0929*** (0.02580)
f_cos_2w	-0.0699*** (0.02460)	-0.0727*** (0.02440)	-0.0739*** (0.02410)	-0.0796*** (0.02410)	-0.0849*** (0.02460)
f_sen_2k	0.0668*** (0.02480)	0.0702*** (0.02460)	0.0719*** (0.02430)	0.0778*** (0.02410)	0.0775*** (0.02440)
Constante	4.564*** (0.02530)	4.559*** (0.02510)	4.557*** (0.02500)	4.552*** (0.02500)	4.547*** (0.02520)

Obs: \*\*\*, \*\* e \* implicam significância estatística aos níveis de 1%, 5% e 10%, respectivamente.

número menor é a ineficiência hospitalar, sugerindo que municípios que possuem maior escala de produção são mais eficientes.<sup>21</sup>

Os coeficientes negativos das variáveis interações em leitos privados e internações em leitos universitários significam que municípios onde predominam hospitais com estas formas de gestão são mais eficientes (o sinal negativo implica que a variável tem um impacto de reduzir a ineficiência). Em relação à significância dos leitos em hospitais privados contratados pelo SUS, pode-se pensar que a provisão dos bens de saúde pelo setor privado tende a ser mais eficiente do que a oferta direta pelo setor público.

Como discutido anteriormente, a justificativa para a inclusão de *dummies* de partido está no fato da saúde, atualmente, ser um serviço público descentralizado. Dessa forma, as prefeituras administram boa parte dos recursos repassados pelo SUS. Acredita-se que prefeituras administradas por um mesmo partido partilhem de práticas comuns de gestão resultantes, entre outras coisas, de programas partidários de Governo. Adicionalmente, para uma melhor inferência sobre gestão é necessário considerar se a gestão é municipal plena ou estadual. A insignificância estatística de todas estas variáveis para expli-

<sup>21</sup>Para uma avaliação do papel do congestionamento e da presença de economias de escala na provisão de serviços públicos municipais de saúde e educação ver Mendes e Sousa (2006). O trabalho de Mendes e Sousa avalia essas questões olhando a função de demanda por serviços públicos e neste sentido é complementar ao presente estudo que enfatiza o lado da oferta.

Tabela 3: Determinantes da ineficiência

	(1)	(2)	(3)	(4)	(5)
População	0.0960*** (0.02090)	0.0922*** (0.02540)	0.103*** (0.02550)	0.0840*** (0.02060)	0.101*** (0.02200)
Internações	-0.113*** (0.02140)	-0.103*** (0.02500)	-0.123*** (0.02530)	-0.111*** (0.02110)	-0.126*** (0.02230)
Tendência	-0.0151** (0.00767)	-0.0032 (0.00692)	-0.00371 (0.00845)	-0.00717 (0.00767)	-0.00718 (0.00711)
Internações hospitalares públicas	-0.181*** (0.05230)	-0.199*** (0.07070)	-0.242*** (0.07280)	-0.223*** (0.06710)	
Internações hospitalares universitários	-0.561* (0.32600)	- 608 (0.38100)	- 686 (0.46400)	- 678 (0.45700)	
PMDB	0.00389 (0.03270)	0.0131 (0.03600)	0.0187 (0.04340)		
PT	- 66 (0.04620)	-0.0829* (0.04830)	-0.0879 (0.05940)		
PFL	0.0186 (0.03690)	24 (0.03990)	0.0128 (0.04750)		
PTB	0.000294 (0.03770)	0.0148 (0.03980)	0.0306 (0.04780)		
Demais	0.00543 (0.02870)	23 (0.03250)	0.0128 (0.03640)		
Gestão	0.0374 (0.02560)	0.0278 (0.02660)	0.0253 (0.03190)		
Internações obstetrícia	- 124 (0.13200)	- 177 (0.13700)			
Internações clínica médica	- 164 (0.12700)	- 134 (0.12300)			
Internações pediatria	-0.471*** (0.16000)	-0.469*** (0.17500)			
Outras internações	-0.670*** (0.21700)	-0.0724 (0.14000)			
Gastos públicos	-0.00468 (0.04170)				
Permanência	0.120*** (0.03630)				
Constante	- 384 (0.28100)	- 324 (0.20700)	-0.547*** (0.20100)	-0.385** (0.15800)	-0.516*** (0.17100)
ilgtgamma Cons	4.036*** (0.25200)	4.075*** (0.33600)	4.332*** (0.26700)	4.214*** (0.27000)	4.174*** (0.24400)
lnsigma2 Cons	-5.556*** (0.23200)	-5.522*** (0.31600)	-5.267*** (0.24800)	-5.368*** (0.24900)	-5.464*** (0.22100)
$\sigma^2$	0.0038625	0.003996	0.0051577	0.0046625	0.0042371
$\gamma$	0.9826323	0.9832849	0.9870321	0.9854348	0.9848355
$\sigma_u^2$	0.0037954	0.003929	0.0050908	0.0045946	0.0041729
$\sigma_v^2$	0.0000671	0.0000668	0.0000669	0.0000679	0.0000643
Observações	2129	2129	2129	2190	2190

Obs: \*\*\*, \*\* e \* implicam significância estatística aos níveis de 1%, 5% e 10%, respectivamente.

car o termo de ineficiência sugere que os resultados eleitorais não alteram o padrão de eficiência produtiva do setor hospitalar para o período estudado. Além disso, para o caso particular da variável de gestão municipal plena, parece não existirem ganhos de eficiência decorrente da descentralização da provisão de bens de saúde.

O termo de tendência nos coeficientes de ineficiência indica que os municípios ao longo do tempo vão melhorando sua eficiência, isto é, vão convergindo para a fronteira.

As variáveis que descrevem os percentuais por tipos de internação - clínica cirúrgica, que é a categoria base, obstetrícia, clínica médica, pediatria e outros (cuidados crônicos prolongados, psiquiatria, fisiologia e reabilitação) indicam internações de pediatria e outros motivos de internação são estatisticamente significantes para redução da ineficiência em relação à clínica cirúrgica. Este resultado era esperado uma vez que a pediatria atende somente crianças en-

**Tabela 4:** Testes para a forma funcional da função de produção e modelagem da ineficiência

Testes para a Forma Funcional					
	Estimativas por				Conclusão:
	MQO		MV		
	LR	p-valor	LR	p-valor	
Cobb-Douglas × Translog	502,35	0,00	126,77	0,00	Rejeita Cobb-Douglas em favor da Translog
Translog × Fourier	188,17	0,00	51,44	0,00	Rejeita Translog em favor de Fourier
Teste para a modelagem de eficiência*			teste est.	p-valor	Rejeita não modelagem de eficiência
			820.5996	0,000	

\* Distribuição Qui-quadrado mista. Valores críticos em Koode e Palm (1986)

quanto as demais modalidades atende a todas as idades, o que aumenta a probabilidade de morte durante o tratamento, pois esta aumenta com o avanço da idade. Pela composição da categoria outras internações também se explica o fato de estas terem menos impacto sobre a mortalidade hospitalar, resultando em maior eficiência para a produção.

Por fim, a variável gasto médio por internação não se mostra estatisticamente significativa para explicar o termo de ineficiência e a variável tempo médio de internação sugere que quanto maior o tempo de internação mais ineficiente é o município. Este último resultado, num primeiro momento poderia sugerir que uma redução no tempo de internação poderia ser ideal, entretanto, deve ser tomado com cautela, pois se deve ter em conta que um tempo maior de internação pode indicar um estoque de saúde muito baixo, de modo que o risco de morte durante a internação é alto.

#### 4 Conclusões e considerações

O objetivo deste trabalho é avaliar a eficiência técnica produtiva do setor hospitalar nos municípios do Estado de São Paulo utilizando como variável dependente a taxa de sobrevivência hospitalar (o complemento da taxa de mortalidade hospitalar) e os gastos com profissionais de saúde empregados nestas internações e o número de leitos disponíveis em cada município como variáveis explicativas/insumos.

Os resultados obtidos, entretanto, devem ser analisados considerando que as informações hospitalares utilizadas são definidas por local de internação e não por moradia. Além disso, deve-se ter em conta o alto grau de agregação das informações e o fato de os modelos estimados não estabelecerem uma relação de causalidade entre as variáveis e a ineficiência.

As estimativas indicam que os municípios com maior população são mais ineficientes, neste caso a população é a residente e as demais informações são de local de internação. Também indicam que são mais eficientes os municípios que: contratam um maior percentual de leitos de hospitais privados e de leitos universitários; possuem maior escala de atendimentos, ou seja, com

maior número de internações por ano; atendem internação por pediatria e por uma categoria denominada outras internações (cuidados crônicos prolongados, psiquiatria, fisiologia e reabilitação); e que apresentam menor tempo médio de internação.

Como o produto considerado é a taxa de sobrevivência hospitalar era esperado que internações por pediatria e outras internações se mostrassem mais eficientes, pois têm impactos diretos sobre a taxa de mortalidade. Sendo assim, estas variáveis podem ser vistas como controles para o grau de complexidade do tratamento oferecido na internação. Com relação à significância dos leitos privados e universitários pode-se pensar que a provisão pelo setor privado parece mais eficiente do que a oferta direta pelo setor público. A redução do tempo médio de internação, por sua vez, não deve ser vista como promotora da eficiência, pois pacientes que ficam internados por muito tempo devem na média apresentar capital-saúde baixo e alto risco de morte.

O fato de o número de internações estar correlacionado com o termo ineficiência com sinal negativo sugere existirem ganhos de escala na produção de qualidade de atendimento hospitalar medida pela taxa de sobrevivência hospitalar. De fato, os municípios que se mostraram menos ineficientes possuem número de internações anuais muito baixos comparados com os demais.

O fato de a população estar correlacionado com o termo de ineficiência com sinal positivo sugere existirem custos de congestionamento.

Por último, variáveis de gestão (partidos políticos e percentual de gestão municipal plena) foram testadas e se mostraram não significantes. Essa ausência de correlação sugere que a qualidade dos produtos de internação hospitalar independe do ciclo político e que a gestão municipal não apresenta ganhos quando comparada com a estadual.

## 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, Economics Working Paper 1/2004/DE/CISEP/UECE, ISEG-UTL.

Aigner, D., Lovell, C. A. K. & Schmidt, P. (1977), 'Formulation and estimation of stochastic frontier production function models', *Journal of Econometrics* 6, 21–37.

Battese, G. & Coelli, T. (1995), 'A model for technical inefficiency effects in a stochastic frontier production function for panel data', *Empirical Economics* 20(2), 325–332.

Berger, A. & Mester, L. (1997), 'Inside the black box: What explains differences in the efficiencies of financial institutions', *Journal of Banking & Finance* 21, 895–947.

Bernet, P. M., Rosko, M. D. & Valdmanis, V. G. (2008), 'Hospital efficiency and debt', *Journal of Health Care Finance* 34(4), 66–88.

- Bradford, W. D. & Kleit, A. N. (2001), 'Stochastic frontier estimation of cost models within the hospital', *The Review of Economics and Statistics* **83**(2), 302–309.
- Brown III, H. S. & Pagán, J. A. (2006), 'Managed care and the scale efficiency of us hospitals', *International Journal of Health Care Finance and Economics* **6**, 278–289.
- Debreu, G. (1951), 'The coefficient of resource utilization', *Econometrica* **19**(3), 273–292.
- Evans, D., Tandon, A. & e J.A. Lauer, C. J. M. (2000), The comparative efficiency of national health systems in producing health : an analysis of 191 countries, GPE Discussion Paper Series 29, World Health Organization.
- Farrel, M. J. (1957), 'The measurement of productive efficiency', *Journal of the Royal Statistical Society* **120**, 253–281. Series A, General.
- Greene, W. (2003a), Distinguishing between heterogeneity and inefficiency : stochastic frontier analysis of the world health organization 's panel data on national health care systems, NYU Working Paper EC-03-10, NYU.
- Greene, W. H. (2003b), *Econometric Analysis*, Prentice-hall.
- Gupta, S. & Verhoeven, M. (2001), 'The efficiency of government expenditures experiences from África', *Journal of Policy Modeling* **23**, 433–467.
- Herrera, S. & Pang, G. (n.d.), Efficiency of public spending in developing countries : an efficiency frontier approach. Mimeo.
- Herrera, S. & Worldwide, W. W. (2005), Efficiency of public spending in developing countries : an efficiency frontier approach, Policy Research Working Paper 3645, The World Bank.
- Jacobs, R. (2000), Alternative methods to examine hospital efficiency: Data envelopment analysis and stochastic frontier analysis, Working papers 177chedp, Centre for Health Economics, University of York.
- Jayasuriya, R. & Wodon, Q. (2002), Measuring and explaining country efficiency in improving health and education indicators, Technical report, The World Bank.
- Kumbhakar, S. & Lovell, K. (2000), *Stochastic Frontier Analysis*, Cambridge University Press.
- Marinho, A. (2001), Avaliação da eficiência técnica nos serviços de saúde dos municípios do estado do rio de janeiro, Texto para discussão 842, IPEA.
- Meeusen, W. & Van de Broeck, J. (1977), 'Efficiency estimation from cobb-douglas production functions with composed error', *International Economic Review* **18**, 435–444.
- Pitt, M. & Lee, L. (1981), 'The measurement and sources of technical inefficiency in indonesian weaving industry', *Journal of Development Economics* **9**, 43–64.

Smet, M. (2007), 'Measuring performance in the presence of stochastic demand for hospital services: an analysis of belgian general care hospitals', *Journal of Productivity Analysis* 27, 13–29.

Sousa, M. C. S., 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(2), 287–313.