

INEFICIÊNCIA DO GASTO PÚBLICO E PODER DE BARGANHA: UMA AVALIAÇÃO DO *FLYPAPER EFFECT* NA ATENÇÃO BÁSICA À SAÚDE

PATRÍCIA SIQUEIRA VARELA
UNIVERSIDADE REGIONAL DE BLUMENAU

LUIZ PAULO LOPES FÁVERO
UNIVERSIDADE DE SÃO PAULO

RESUMO

O desenho do sistema de transferências intergovernamentais pode provocar comportamentos e resultados indesejáveis ao bem-estar social. A literatura sobre o federalismo prevê que as transferências não-condicionais e sem contrapartida provocam gasto público com desperdício ou ineficientes, ocasionado pelo fenômeno conhecido como *Flypaper Effect*, sendo a redução do poder de barganha dos cidadãos uma das explicações para a sua ocorrência. Assim, este artigo teve por objetivo medir e explicar as variações de desempenho dos Municípios Paulistas quanto à eficiência técnica na aplicação de recursos públicos nas ações de atenção básica à saúde em função do perfil de financiamento dos gastos gerais e específicos de tal área. A eficiência técnica reflete a capacidade de uma entidade obter máximos *outputs* com o menor consumo de *inputs* e foi medida com o uso da técnica *Data Envelopment Analysis (DEA)*: modelo em dois estágios. Para verificar a relação de dependência entre eficiência técnica e perfil de financiamento dos gastos com saúde foi usada a análise de regressão. Os resultados indicaram que seria possível aumentar, consideravelmente, a quantidade de serviços prestados à população sem a necessidade de novas dotações orçamentárias na maioria dos Municípios. E, principalmente, as transferências não-condicionais e sem contrapartida geram um efeito negativo no escore de eficiência, corroborando a hipótese do *Flypaper Effect*. Apesar de parte das transferências sob estudo poderem ser carregadas para outras jurisdições, a redução no poder de barganha ocorreria pelo alto custo de se obter informações, o que poderia ser compensado pelo grau de escolaridade dos municípios que apresentou interdependência positiva com a eficiência.

Palavras-chave: *Flypaper Effect*; Poder de Barganha; Eficiência Técnica; Atenção Básica à Saúde; Municípios Paulistas.

1 INTRODUÇÃO

O Brasil, há mais de um século, adotou a forma federativa de divisão territorial. A Constituição Federal de 1988 reestruturou o federalismo brasileiro, fortalecendo os governos municipais, mormente quanto à repartição das receitas públicas. A descentralização tem sido um processo comum tanto do movimento de redemocratização do país, assim como no contexto de reforma do aparelho do Estado, cujo foco principal é a eficiência dos gastos públicos e a gestão voltada para os cidadãos. Além disso, a Constituição Federal de 1988 definiu competências concorrentes ou partilhadas pelas esferas federal, estadual e municipal para um amplo conjunto de políticas públicas, entre elas a política de saúde que se caracteriza por complexas relações intergovernamentais que têm garantido avanços paulatinos na resolução de problemas quanto à prestação de serviços públicos de saúde à população.

No federalismo fiscal, quando se comparam as prescrições para a alocação eficiente de tributos e de gastos entre níveis de governo, tem-se que diversas ações públicas são passíveis

de descentralização, não ocorrendo o mesmo com o conjunto de tributos que podem ser arrecadados de forma eficiente pelos governos municipais e estaduais. O resultado é um *gap* entre receitas e despesas, denominado *vertical fiscal gap*. Surge, assim, um dos motivos para a existência de transferências intergovernamentais que, de acordo com Shah (2007, p. 2-9), podem ser classificadas como: de propósito geral, por definição não-condicional e sem contrapartida; ou de propósito específico. Esta última pode ser dividida em: sem contrapartida e com contrapartida. A contrapartida, por sua vez, pode ser limitada ou ilimitada.

No Brasil, o federalismo é marcado pela forte dependência dos Municípios em relação aos recursos recebidos por transferência, principalmente do Fundo de Participação dos Municípios (FPM), de origem federal, e a cota-parte do Imposto sobre a Circulação de Mercadorias e Prestação de Serviços (ICMS), de origem estadual. Transferências essas não-condicionais e sem contrapartida.

No caso da função saúde, uma parcela das transferências não-condicionais e sem contrapartida mais parte dos recursos arrecadados pelo próprio Município compõem a base de cálculo para o percentual mínimo de 15% a ser aplicado nas ações de saúde. Os governos locais também recebem transferências do Fundo Nacional de Saúde, sendo uma parte fixa (valor *per capita* vezes a população do Município) e uma parte variável, como forma de incentivo para ações específicas e ampliação de oferta em áreas estratégicas. Além disso, os Municípios recebem transferências fundo a fundo conforme as suas capacidades de ofertar procedimentos de média e alta complexidades.

A análise da literatura sobre federalismo fiscal, mais especificamente sobre o fenômeno conhecido como *Flypaper Effect*, abre uma discussão sobre o comportamento dos burocratas e políticos frente ao orçamento. O *Flypaper Effect* indica que o recebimento de transferências fiscais não-condicionais e sem contrapartida implica em aumento das despesas públicas locais proporcionalmente maior do que um aumento equivalente na renda pessoal. Teoricamente, existe uma equivalência entre transferências para governos e residentes. Dessa forma, o impacto das transferências ou da renda pessoal sobre a despesa pública deveria ser o mesmo. Entretanto, empiricamente, este resultado não tem sido observado.

Wyckoff (1988) usa o modelo burocrático de Niskanen (1971) para explicar o *Flypaper Effect*, em que os cidadãos teriam maior poder de barganha para influenciar as decisões orçamentárias quando a receita municipal fosse oriunda da arrecadação tributária, pois poderiam mudar-se e levar tal receita para outras jurisdições quando sua função utilidade mínima não fosse atendida. No caso das transferências não-condicionais e sem contrapartida, a situação seria mais favorável para os burocratas maximizarem os orçamentos e, conseqüentemente, sua função utilidade que não necessariamente promove o melhor resultado social. Sabe-se que um maior controle sobre as decisões fiscais reduziria o *Flypaper Effect*. Por outro lado, conforme argumentam Bardhan e Mookherjee (2005, p. 683), o grau de informação dos eleitores é uma função direta da sua situação socioeconômica.

Sendo assim, este estudo teve por objetivo medir e explicar as variações de desempenho dos Municípios Paulistas quanto à eficiência técnica na aplicação de recursos públicos nas ações de atenção básica à saúde em função do perfil de financiamento dos gastos gerais e específicos de tal área.

A eficiência técnica reflete a capacidade de uma entidade obter máximos *outputs* com o menor consumo de *inputs* e é influenciada pela forma como os políticos e burocratas lidam com as restrições orçamentárias. O perfil de financiamento dos gastos públicos municipais, aqui especificamente atenção básica à saúde, pode interferir na forma como os políticos e burocratas se posicionam frente ao orçamento, com maior ou menor propensão a elevar os

custos de produção dos serviços em tal área. Nesse sentido, as hipóteses deste estudo podem ser enunciadas, como segue:

- H1: Quanto maior o grau de dependência de recursos transferidos por outras esferas de governo para financiamento das políticas públicas municipais, menor a eficiência técnica na provisão de serviços de atenção básica à saúde;
- H2: Quanto maior o percentual de transferência vinculada do governo federal, geral ou específica da atenção básica, em relação às receitas correntes, maior a eficiência técnica dos Municípios na subfunção atenção básica à saúde.
- H3: Quanto piores as condições socioeconômicas dos municípios, menor a eficiência técnica na subfunção atenção básica à saúde pelos Municípios.

Espera-se que este trabalho contribua para a construção de bases empírico-teóricas para a Controladoria na Gestão Pública, uma vez que busca encontrar alternativas para a avaliação de desempenho no setor público e em uma área tão complexa quanto à saúde.

A avaliação de desempenho envolve dificuldades relacionadas à mensuração dos produtos diretos e, mormente, os finais dos programas governamentais, à avaliação da qualidade e ao relacionamento entre recursos orçamentários e efetividade. No caso específico da saúde, a determinação dos centros de *inputs-outputs* e sobre o que cada gestor público pode ser responsabilizado também se torna um problema, dado que existem interações entre os diversos sistemas de saúde (municipais, estaduais e federal), com a existência de fluxos de pacientes e recursos entre eles, de acordo com a capacidade de cada um de prestar serviços de média ou alta complexidades.

Para contornar esta última situação, optou-se por abordar a subfunção atenção básica à saúde que é exclusiva de cada Município e, assim, os gestores públicos podem ser responsabilizados pelo o que acontece em seu território, desde que seja fruto do seu poder discricionário. Além disso, o trabalho inova ao focar os *outputs* diretos da produção em saúde, simplificando muitas das restrições ao processo de avaliação de desempenho e *accountability*, principalmente quanto à defasagem entre as ações e os seus resultados efetivos.

2 TRANSFERÊNCIAS INTERGOVERNAMENTAIS, *FLYPAPER EFFECT* E PODER DE BARGANHA

Em sistemas federalistas, existem vários motivos para a existência de transferências intergovernamentais: lidar com a falta de emparelhamento entre meios de obter receitas e necessidades de despesa dos governos locais, dar tratamento às iniquidades horizontais, fornecer compensações a governos locais por benefícios gerados para a população fora da sua jurisdição, influenciar as políticas subnacionais de tributação e gasto, entre outros.

A relevância dos governos subnacionais na provisão de serviços públicos e as diferentes formas de transferências realizadas por governos centrais estimulam a necessidade de um melhor entendimento sobre o comportamento orçamentário de governos locais e estaduais, frente ao impacto das transferências intergovernamentais.

Uma abordagem teórica apropriada para analisar este assunto, de acordo com Bradford e Oates (1971, p. 417), deve incorporar, explicitamente, o processo político, isto é, o processo de escolha coletiva através do qual os recebedores de repasses determinam suas decisões quanto ao incremento de recursos. Os autores (1971, p. 422), baseados em um modelo de competição política e informação perfeita semelhante à teoria de equilíbrio competitivo para bens privados, concluíram que os efeitos alocativos e distributivos de uma transferência não-condicional e sem contrapartida para um governo subnacional não seriam diferentes da

distribuição da mesma quantia diretamente para os residentes locais. Em ambos os casos, o processo político conduziria a uma mesma combinação entre bens públicos e bens privados.

Inman (1979, p. 274) trata a função demanda de indivíduos maximizadores de utilidade em um contexto de escolha fiscal. Para isso, as preferências são especificadas completamente sob serviços locais e renda após transferências e tributos, enquanto a restrição orçamentária requer que receitas oriundas de taxas e subvenções sejam iguais às despesas públicas locais sobre o *output* público. De maneira simplificada, as funções de demanda da despesa pública são determinadas para maximizar uma função utilidade $U = U(G, Y)$ sujeita à seguinte restrição orçamentária:

$$I + \tau z = \sum_i \tau c_i (1 - m_i) G_i + Y \quad (2.1)$$

Em que I é a renda individual antes dos tributos, τ a participação dos indivíduos nos tributos *per capita* (varia entre 0 e 1 e mede o efeito de deduções e créditos tributários), z a transferência não-condicional e sem contrapartida *per capita*, c_i o índice de custo para a função despesa i , m_i o percentual de transferência com contrapartida, G_i a quantidade de *output per capita* na função i , e Y a renda privada após tributos. Caso τ fosse igual a 1, a renda do indivíduo antes dos tributos mais as transferências não-condicionais e sem contrapartida deveria ser igual ao custo para prover a parte não subsidiada de uma determinada quantidade de bens públicos mais a renda líquida dos tributos.

A partir deste modelo, Inman (1979, p. 292-293) deriva as condições relacionadas à elasticidade da despesa G em relação a z e a I :

$$E_{G,z} = (\tau/I)E_{G,I} \quad \text{ou} \quad \frac{dG_i}{dz} = \tau \frac{dG_i}{dI} \quad (2.2)$$

A variação da despesa em relação à transferência não-condicional e sem contrapartida é igual à variação da despesa em função da variação da proporção da transferência (ajustada por τ) sob a renda bruta. Isso porque cada dólar de transferência não-condicional e sem contrapartida *per capita* dado ao governo representa o montante que o indivíduo deveria pagar para aumentar a receita ou que ganharia caso os tributos *per capita* fossem reduzidos.

As premissas usadas por Bradford e Oates (1971) e Inman (1979) servem como referencial de análise, mas não são observadas na realidade. Em estudos sobre os efeitos de transferências intergovernamentais, um fato observado é que aquelas do tipo não-condicional e sem contrapartida aumentam as despesas dos governos subnacionais mais do que um aumento equivalente na renda pessoal privada. Este resultado tem sido denominado *Flypaper Effect* para refletir a noção de que “*money sticks where it hits*”¹, isto é, o dinheiro recebido no setor público tende a permanecer no setor público, enquanto o recebido no setor privado tende a permanecer lá (FISHER, 1982, p. 324).

Uma explicação para o *Flypaper Effect* baseia-se nas variáveis determinantes do poder de barganha de grupos ou coalizões de eleitores e o nível de reversão de serviços públicos previsto nas constituições locais ou exigido por eles. Wyckoff (1988, p. 118) busca explicar o *Flypaper Effect* baseado na assimetria de poder de barganha entre burocrata/governante e os eleitores. Nessa situação, os eleitores têm perfeita percepção do total das transferências recebidas, mas estão sujeitos ao comportamento monopolista e de maximização de orçamento dos governantes locais.

O modelo burocrático de escolha pública local de Wickoff (1988) foi desenvolvido com base na *Theory of Supply by Bureaus* de Niskanen (1971), cujos elementos críticos são:

as características que distinguem os *bureaus*, a natureza da relação entre *bureaus* e seu ambiente e a função-objetivo de maximização dos burocratas.

De acordo com Niskanen (1971, p. 15), *bureaus* são organizações que possuem as seguintes características:

- os proprietários ou empregados da organização não se apropriam, como renda pessoal, de qualquer parte da diferença entre receitas e custos.
- uma parte da receita recorrente da organização vem de outras fontes que não a venda de *output* a um determinado preço por unidade.

Em outras palavras, “*Bureaus* são organizações não-lucrativas que são financiadas, ao menos em parte, por dotação ou transferência periódica.” (NISKANEN, 1971, p. 15). O *bureau* é um especialista na provisão de bens e serviços, cuja preferência por algumas pessoas é que sejam supridos em grandes quantidades e não a um preço de venda unitário, já que não estão preparadas para contratar com organizações lucrativas e, mesmo que assim o fizessem, a quantidade fornecida seria menor. Assim, tais pessoas compõem uma organização coletiva e a ela destinam recursos (NISKANEN, 1971, p. 18). Um *bureau* pode ser uma entidade coletiva privada, como também as várias esferas de governo. Seguindo a acepção de *bureau*, Niskanen (1971, p. 22-23) define burocrata como “[...] o funcionário público *sênior* de qualquer *bureau* com um orçamento separado identificável.” (grifos nossos).

O ambiente do *bureau* é composto pelas relações com três grupos: a organização coletiva que fornece fundos recorrentes para o *bureau*, os fornecedores de trabalho e fatores materiais de produção e, em alguns casos, os consumidores de serviços que são vendidos a um determinado preço. Das três, a primeira é a que mais distingue os *bureaus* de outros tipos de organização e é sobre ela que se pretende discorrer.

A maioria dos *bureaus* é financiada por uma organização coletiva única ou dominante que, por sua vez, é financiada pelas receitas tributárias ou por contribuições mais ou menos compulsórias. Os membros eleitos da organização coletiva (por exemplo, poder legislativo) são responsáveis pelo processo de revisão e aprovação das atividades e orçamentos propostos pelos *bureaus*, controle dos seus métodos e desempenho e, geralmente, aprovam a indicação do chefe do *bureau*. A organização financiadora depende de um *bureau* específico para a provisão de determinado serviço e, por sua vez, o *bureau* não tem uma fonte alternativa comparável de financiamento, o que levaria a pensar numa relação de monopólio bilateral (NISKANEN, 1971, p. 24).

O *bureau* troca, com uma organização financiadora, um conjunto de atividades e de *outputs* esperados dessas atividades por um orçamento, o que é conduzido inteiramente em termos totais e não unitários. Do ponto de vista do *bureau*, as preferências dos financiadores são resumidas em uma função *output*-orçamento. Qualquer ponto nesta função representa o orçamento máximo que o financiador aprovaria para um nível esperado de *output* (NISKANEN, 1971, p. 25).

Outra característica da relação *bureau* e financiador é a grande disparidade na informação disponível para ambos. Um burocrata conhece muito mais sobre custos dos fatores e processos de produção dos serviços do que os membros da organização financiadora. O burocrata necessita de pouca informação, muitas possíveis de serem estimadas pelo comportamento revelado do financiador, para explorar sua posição como um supridor monopolista de um dado serviço. Por outro lado, os membros da organização coletiva precisam de muitas informações e muito pouco pode ser estimado, dificultando a sua posição de comprador monopolista. Além disso, um burocrata tem um incentivo relativo muito forte e pode trabalhar em tempo integral para obter informação relevante para ele (e obscurecer

informações para o financiador). Dessa forma, ao *bureau* é dado o poder de monopólio, predominantemente, dominante (NISKANEN, 1971, p. 29-30).

O último elemento crítico da teoria desenvolvida por Niskanen (1971) é a função-objetivo de maximização do burocrata. O autor (1971, p. 38) argumenta que, entre as diversas variáveis que poderiam entrar na função utilidade, tem-se: salário, privilégios da ocupação, reputação pública, poder, poder de nomear funcionários públicos, *output* do *bureau*, facilidade de fazer mudanças e facilidade de administrar o *bureau*. Com exceção das duas últimas variáveis, todas as outras têm uma relação positiva e contínua com o total do orçamento. Não é necessário que a função utilidade do burocrata seja fortemente dependente de cada uma das variáveis que crescem com o orçamento. Uma *proxy* adequada seria a maximização do orçamento em si, mesmo para burocratas com pouca motivação pecuniária, mas muita vontade de fazer mudanças de acordo com o interesse público (NISKANEN, 1971, p. 38).

A maximização do orçamento por um *bureau* está sujeita à restrição de que o orçamento deve ser igual ou maior aos custos totais mínimos de fornecer o *output* esperado pelo financiador, ou seja, a organização coletiva (NISKANEN, 1971, p. 42).

Wickoff (1988), a partir destes elementos críticos da *Theory of Supply by Bureaus* e da premissa de que as preferências da organização coletiva – conselho da cidade – é a mesma do eleitor mediano, desenvolveu um modelo de caso extremo para a escolha pública local.

Neste modelo, o conselho municipal não tem nenhuma informação sobre o custo de bens públicos e, portanto, do efeito de mudanças marginais nos orçamentos propostos pelo *bureau* e, sendo avesso ao risco, tende a não realizar mudanças no orçamento do *bureau*. À incerteza quanto às mudanças marginais é acrescentado o comportamento estratégico por parte do *bureau* que reage a cortes no orçamento, eliminando os programas mais populares em primeiro lugar. Dessa forma, cabe ao conselho uma escolha binária entre duas alternativas: aceitar o orçamento do *bureau* em sua integridade ou ficar sem o *bureau* inteiro (WICKOFF, 1988, p. 118). Neste caso, o *bureau* deve atender uma função-utilidade mínima do eleitor U_0^V para assegurar a manutenção no cargo.

Wickoff (1988, p. 118) destaca, ainda, dois resultados do modelo teórico de Niskanen (1971). Primeiro, o *output* é sempre elástico com relação a preço porque ao eleitor é colocada uma escolha binária e não marginal; devido à elasticidade maior do que um, o *bureau* maximizador de orçamento reduzirá seu preço até o nível de custos e se esforçará para aumentar *output*. Segundo, um dólar de aumento de transferência não-condicional e sem contrapartida geraria mais do que um dólar de aumento em despesas, uma vez que aumenta a função-utilidade do eleitor, permitindo exploração adicional por parte do *bureau* (mais *outputs* e maiores tributos).

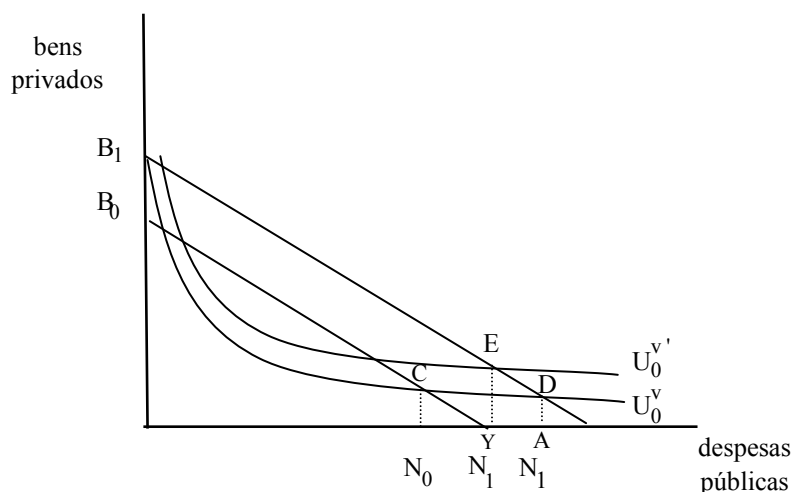


Gráfico 1 - Flypaper Effects no modelo de Niskanen
 FONTE: WICKOFF, 1988, p. 119.

O Gráfico 1 mostra as condições de equilíbrio sob o modelo de Niskanen (1971). Inicialmente, o eleitor mediano tem uma restrição orçamentária que intercepta o eixo de bens privados em B_0 . O burocrata propõe o maior orçamento e, portanto, o maior *output* sobre essa restrição orçamentária, o qual é também compatível com a função utilidade $U(X_1, X_2) \geq U_0^v$, representado pelo ponto C com despesas públicas iguais a N_0 .

Se o eleitor recebe um aumento de renda, portanto, a nova restrição orçamentária passa a interceptar o eixo de bens privados em B_1 e a curva de utilidade aumenta para $U_0^{v'}$, uma vez que o eleitor pode levar consigo esta renda extra se decidir mudar para outra jurisdição. O novo ponto de equilíbrio torna-se E com despesa igual a N_1^Y . Por outro lado, se há um aumento em transferências não-condicionais e sem contrapartida, com o mesmo impacto na restrição orçamentária, o eleitor não consegue carregar a sua parcela da transferência para a sua nova localização. Como resultado, U_0^v fica inalterada e o equilíbrio final é o ponto D com despesa igual a N_1^A . O *Flypaper Effect* é observado pelo fato de $N_1^A - N_0$ exceder $N_1^Y - N_0$.

O nível mínimo de utilidade exigido pelo eleitor depende de sua renda, já que um aumento desta permite a ele exigir do governante um aumento na sua função-utilidade, ameaçando-o de levar consigo parte dos recursos orçamentários para outra jurisdição. O mesmo não ocorre com as transferências não-condicionais e sem contrapartida, pois são vistas como um volume adicional de dinheiro pertencente à comunidade e não ao eleitor.

Mesmo se a transferência não-condicional e sem contrapartida pudesse ser carregada pelo eleitor para a nova jurisdição, como é o caso do FPM aqui no Brasil, o *Flypaper Effect*, provavelmente, ocorreria por causa dos custos de se obter informações, já que tais transferências não são uniformes entre as jurisdições (WICKOFF, 1988, p. 120).

3 TRAJETÓRIA METODOLÓGICA

3.1 Censo de Municípios Paulistas

A unidade de análise deste estudo foi o Município, enquanto responsável pela subfunção atenção básica à saúde. A população, portanto, foi composta pelos 645 Municípios do Estado de São Paulo, escolhidos por causa da relevância da sua economia no Brasil e pela diversidade socioeconômica, como apontado pelo Índice Paulista de Responsabilidade Social

(IPRS), divulgado pela Fundação Sistema Estadual de Análise de Dados – SEADE (2008). O estudo foi realizado através de um censo dos Municípios Paulistas.

3.2 Escore de Eficiência Técnica

De acordo com Cohen e Franco (2004, p. 77), os conceitos utilizados em avaliação de projetos sociais, e aí se incluem os serviços de saúde, foram desenvolvidos pela economia e estão associados à função de produção que vincula recursos, insumos, processos e produtos. Os mesmos autores, assim como Shah (2007, p. 11), sinalizam que, no processo de provisão de bens e serviços públicos, a dificuldade em mensurar os *outputs* surge de forma mais aguda do que na produção de bens privados. Bradford, Malt e Oates (1969) já haviam se preocupado com esta questão e buscaram esclarecê-la com o desenvolvimento de um modelo em que os *outputs* são separados entre diretos (*Output-D*) e de interesse final do cidadão (*Output-C*). Por exemplo, o produto final de um programa sentido por uma população não é o número de vacinas aplicadas, mas a redução da mortalidade infantil.

Combinações diferentes de *inputs* (*Input-I*) geram vetores de bens e serviços diretamente produzidos. Os vetores de *outputs* finais, por sua vez, são completamente determinados pelo vetor de *outputs* diretos e por certas variáveis ambientais ou exógenas, por exemplo, as condições de saneamento no caso da saúde. Diante destes pressupostos, as preferências dos indivíduos podem ser expressas como $U = U(C, Z)$. O vetor Z representa o nível de provisão de outros bens públicos e da quantidade de bens privados consumidos pelo indivíduo. Em que $C_k = f_k(D, E)$, com E representando um vetor de variáveis ambientais e f_k uma função indicativa do grau do *output* final sentido pelo indivíduo ligado diretamente aos serviços D produzidos no ambiente E (BRADFORD; MALT; OATES, 1969, p. 187).

Pela análise dos argumentos de Bradford, Malt e Oates (1969) é possível pensar em duas relações: (1) *inputs* e *outputs* diretos e (2) *inputs* e *outputs* finais. Um programa social pode exigir configurações diferentes em termos de ações e, conseqüentemente, produtos diretos, conforme as circunstâncias ambientais. Portanto, haverá gastos diferentes para se alcançar os objetivos relacionados ao bem-estar da população, mesmo que o custo em si dos produtos diretos de cada ação não se altere. É importante observar que a ineficiência do gasto com *outputs* diretos afeta o gasto com os *outputs* de interesse final do cidadão.

As variáveis ambientais ou exógenas poderiam também ser pensadas em termos da transformação de *inputs* em *outputs* diretos. Assim, a função de produção poderia ser representada por $D = f(I, W)$, em que D é a quantidade máxima de *outputs* diretos obtidos, pelo vetor de *inputs* I e influenciada pelo vetor de variáveis ambientais ou exógenas W , com base em uma determinada tecnologia de produção. Uma jurisdição pode fazer mais com menos recursos ou vice-versa, dependendo das suas condições ambientais.

Na avaliação de desempenho, o cálculo das medidas de eficiência pressupõe que a função de produção seja conhecida. Na prática, isto não acontece e ela precisa ser estimada a partir de uma amostra de dados com o uso de métodos paramétricos ou não-paramétricos, sendo este último o caso da *Data Envelopment Analysis* (DEA), termo usado primeiramente por Charnes, Cooper e Rhodes (1978), desenvolvida com base nos trabalhos de Koopmans (1951) e Debreu (1951) sobre produtividade e de Farrell (1957) sobre eficiência.

A DEA é uma técnica de otimização, baseada em programação linear e projetada para estabelecer uma medida de eficiência relativa entre diferentes unidades tomadoras de decisão (*Decision Making Units* - DMUs). Os vários modelos de DEA se baseiam na análise de eficiência das unidades tomadoras de decisão (DMUs) com múltiplos insumos (*inputs*) e produtos (*outputs*) e parte da idéia de construir uma fronteira de eficiência, em que as DMUs mais eficientes se situem sobre a superfície da fronteira enquanto as menos eficientes se

situem internamente abaixo da fronteira. Tais modelos podem assumir retornos constantes de escala, ou seja, eficiência de escala, ou retornos variáveis de escala. Além disso, podem ter orientação para a maximização de *outputs* ou minimização de *inputs*, dependendo de sob quais quantidades o administrador tem mais controle, *input* ou *output*.

Como este trabalho trata do setor público de saúde e sabe-se que os burocratas ou políticos trabalham com um orçamento fixado, a orientação deve ser para os produtos, ou seja, com os insumos disponíveis a entidade pública deve oferecer o máximo de produtos possíveis dentro de certos padrões de qualidade. Quanto ao rendimento de escala, análises preliminares indicaram que seria variável, o que foi seguido no restante dos procedimentos adotados. O escore de eficiência (θ) é dado pela medida de Farrell (1957), $1 \leq \theta < \infty$, e $\theta - 1$ é o aumento proporcional em *outputs* que se poderia alcançar pela DMU, mantendo a quantidade de *inputs* constante. Observa-se que $100/\theta$ - medida de Sheppard (1970) - define o escore de eficiência que varia entre zero e 100%.

No primeiro estágio do modelo DEA, uma fronteira de produção foi estimada a partir do *input* discricionário e dos *outputs* relativos aos serviços de atenção básica à saúde prestados pelos Municípios Paulistas. A função de produção foi composta de um insumo (*input*) e de seis produtos diretos referentes ao atendimento ambulatorial de atenção básica, ao PSF (Programa de Saúde da Família) e ao PACS (Programa de Agentes Comunitários de Saúde) (*outputs* 1 a 6), conforme pode ser visto no Quadro 1.

Quadro 1 - Variáveis de *Input* discricionário e *Outputs* do modelo de eficiência econômica dos Municípios Paulistas - 2006

Tipo	Variável	Descrição	Fonte
INPUT 1	Despesa com Atenção Básica	Total da despesa liquidada com a subfunção atenção básica à saúde.	SIOPS/DATASUS
OUTPUT 1	Ações de Enfermagem e Outros Procedimentos de Saúde - Nível Médio	Quantidade de procedimentos realizados neste tipo de ação.	SIA/DATASUS
OUTPUT 2	Ações Médicas Básicas	Quantidade de procedimentos realizados neste tipo de ação.	SIA/DATASUS
OUTPUT 3	Ações Básicas em Odontologia	Quantidade de procedimentos realizados neste tipo de ação.	SIA/DATASUS
OUTPUT 4	Ações Executadas por Outros Profissionais – Nível Superior	Quantidade de procedimentos realizados neste tipo de ação.	SIA/DATASUS
OUTPUT 5	Pessoas Cadastradas pelo PSF	Quantidade de pessoas cadastradas pelo Programa Saúde de Família.	SIAB/DATASUS
OUTPUT 6	Pessoas Cadastradas pelo PACS	Quantidade de pessoas cadastradas pelo Programa Agentes Comunitários de Saúde.	SIAB/DATASUS

Nesta fase, 42 observações foram eliminadas por falta ou inconsistência dos dados, o que representa 6,51% da população objeto de estudo.

Os escores de eficiência calculados pelo modelo DEA podem ser, severamente, influenciados pela presença de *outliers* nos dados, pois uma simples observação muito distante da média empurra a fronteira e, conseqüentemente, diminui artificialmente a eficiência para o conjunto de DMUs como um todo e, assim, ocorre um subestimação dos escores de eficiência. A base para a detecção de *outliers*, conforme proposto por Sousa e Stošić (2005), foi a medida *leverage*, termo emprestado da literatura sobre regressão linear e que indica a influência da remoção de uma dada DMU nos escores de eficiência das outras DMUs. Dez municípios foram considerados *outliers*, mas somente quatro foram excluídos da análise por discrepância nos dados. Os outros seis Municípios permaneceram no estudo, uma vez que poderiam servir como referência de boas práticas.

3.3 Variáveis não-Controláveis

O modelo DEA em dois estágios permite a incorporação de variáveis ambientais ou não-controláveis na análise, ou seja, variáveis que não estão sob o controle dos gestores de uma DMU, as quais eles podem, no máximo, administrar os seus efeitos na eficiência. Assim, neste segundo estágio da *Data Envelopment Analysis*, buscou-se identificar variáveis que interferem e em que direção o fazem no grau de reversão da função orçamento-*output*. Existem fatores que, supostamente, influenciam a obtenção do uso eficiente dos recursos públicos que os gestores da área da saúde, a curto prazo, não conseguem alterar, mas precisam estar atentos aos seus efeitos no desempenho dos seus centros de responsabilidade.

Alguns destes fatores relacionam-se com as características populacionais. O maior percentual de população rural e/ou a menor densidade populacional podem tornar os custos da prestação de serviços em saúde mais altos por causa do maior esforço de se chegar até a população, inclusive com perdas de economia de escala. Da mesma forma, um alto percentual da população na faixa-etária acima de 60 anos pode representar altos custos de locomoção ou do paciente até a unidade de saúde ou das equipes de saúde até a sua residência. Igualmente, vale perceber se a população na calda inferior da faixa-etária – 0 a 18 anos – afeta os custos da prestação de serviços em atenção básica. As duas faixas etárias foram escolhidas de acordo com as tabelas utilizadas pelos planos de saúde.

Também se poderia pensar em como a organização dos estabelecimentos de saúde com maior capacidade de atendimento quanto ao volume de pessoas pode favorecer ou não o uso eficiente dos recursos públicos. Outro fator é o percentual de recursos próprios aplicados na função saúde que, por exigência legal, deve ser no mínimo de 15%. Talvez, esta norma possa servir como um incentivo para o gasto com desperdício, já que o governo local precisa atingir este mínimo de qualquer maneira. Todavia, sabe-se que muitos Municípios gastam além do mínimo, indicando a prioridade que dão à função saúde.

Quadro 2 - Variáveis não-controláveis do modelo de eficiência econômica dos Municípios Paulistas - 2006

Tipo	Variável	Descrição	Fonte
IND 1	População Urbana	Percentual da população urbana em relação à população total.	Fundação SEADE
IND 2	Faixa-etária I	Percentual da população com idade entre 0 a 18 anos.	DATASUS
IND 3	Faixa-etária II	Percentual da população com idade de 60 anos ou mais.	DATASUS
IND 4	Densidade Populacional	Relação entre o total da população e a área do Município.	Fundação SEADE
IND 5	Escala dos Estabelecimentos de Saúde	População dividida pela quantidade de estabelecimentos ligados à gestão pública municipal que prestam serviços de atenção básica.	DATASUS
IND 6	Percentual de Recursos Próprios Aplicados na Função Saúde	Percentual de aplicação de recursos próprios aplicados em saúde calculados em conformidade com a Emenda Constitucional 29/2000.	SIOPS/DATASUS

Sendo assim, fez-se a opção pelo uso da análise de regressão, $\theta_i = \beta \cdot z_i + \varepsilon_i \geq 1$, na qual a variável dependente foi o escore de eficiência calculado no primeiro estágio e as variáveis independentes aquelas listadas no Quadro 2. Todas as variáveis são relativas ao ano de 2006, com exceção da IND 4, cujos dados disponíveis mais recentes são de 2005.

O termo de erro da regressão do segundo estágio do modelo DEA reflete os escores de eficiência de cada Município paulista ajustado pelos efeitos dos *inputs* não-controláveis. Para que este termo indique as ineficiências ajustadas, θ_i^A , ele deve variar de $1 \rightarrow \infty$, contemplando, assim, a noção de fronteira. De acordo com Greene (1980, p. 34), com o uso de Mínimos Quadrados Ordinários (MQO), os parâmetros são consistentemente estimados, com exceção do intercepto. Contudo, o maior erro obtido por MQO pode ser usado como

estimador consistente do intercepto. Assim, para cada DMU, somando-se ao erro da regressão o erro máximo, obteve-se o índice de eficiência ajustado; pois com o deslocamento do intercepto, todos os erros passam a ter o sinal correto. A expressão para o escore de eficiência ajustado θ_1^A é: $\theta_1^A = \varepsilon_i + \text{Max}_n \varepsilon_n$.

3.4 Teste das hipóteses

O teste de cada uma das hipóteses levantadas foi realizado por meio da análise de regressão clássica - MQO. A variável dependente, *proxy* do gasto público com desperdício, foi o escore de eficiência ajustado θ_1^A .

Já as variáveis independentes foram aquelas representativas do perfil de financiamento do gasto público municipal de saúde que, como indicado pelo referencial teórico, podem influenciar a eficiência econômica dos Municípios. As variáveis independentes são representativas da composição do financiamento das despesas com saúde dos Municípios, geral e restrito à atenção básica, do grau de dependência dos Municípios em relação às transferências não-condicionais e sem contrapartida e sobre o perfil socioeconômico dos municípios. Dados sobre o recebimento de transferências discricionárias (convênios/acordos) dos municípios foram inseridos como variáveis de controle. As especificações de cada uma delas encontram-se no Quadro 3, a seguir.

Quadro 3 - Variáveis independentes do modelo de regressão - 2006

Tipo	Variável	Descrição	Fonte
TG	Grau de dependência dos Municípios de receitas intergovernamentais não-condicionais e sem contrapartida.	Soma das transferências relativas à cota FPM (Fundo de Participação dos Municípios), ICMS (Imposto sobre a Circulação de Mercadorias e Prestação de Serviços), IPVA (Imposto sobre a Propriedade de Veículos Automotores), IPI (Imposto sobre Produtos Industrializados) exportação e compensação financeira pela exploração de recursos naturais dividida pelas receitas correntes (valores de arrecadação)	SIOPS/DATASUS
TD	Percentual de transferências discricionárias em relação à receita corrente.	Percentual de transferências voluntárias recebidas dos governos federais e estaduais em relação à receita corrente do Município (valores de arrecadação).	SIOPS/DATASUS
SUSVHAB	Transferência SUS vinculada a ações e programas específicos.	Total de transferências do Fundo Nacional de Saúde referente a ações estratégicas + Piso da Atenção Básica (PAB)Variável dividido pela receita corrente.	SIOPS/DATASUS
SUSNVHAB	Transferência SUS sem vinculação.	Total de transferências do Fundo Nacional de Saúde, sem vinculação específica, dividido pela receita corrente.	SIOPS/DATASUS
PABV	Transferência SUS vinculada a ações e programas específicos da atenção básica.	Total de transferências do Fundo Nacional de Saúde referente ao PAB-Variável dividido pela receita corrente.	SIOPS/DATASUS
PABNV	Transferência SUS para a atenção básica, sem vinculação.	Total de transferências do Fundo Nacional de Saúde referente ao PAB-Fixo dividido pela receita corrente.	SIOPS/DATASUS
IR	Indicador de Riqueza.	Indicador da dimensão Riqueza do IPRS.	SEADE
EC	Indicador de Escolaridade.	Indicador da dimensão Escolaridade do IPRS.	SEADE

Todas as variáveis são relativas ao ano de 2006, com exceção de IR, cujos dados disponíveis mais recentes são de 2005. Nesta fase, optou-se por não excluir os *outliers*, pois, se assim fosse feito, vários Municípios de maior porte seriam eliminados da análise. Além do que, apesar de alguns Municípios apresentarem valores discrepantes em algumas variáveis, os *outliers* refletem práticas reais.

4 ANÁLISE DOS RESULTADOS

No primeiro estágio, a fronteira de eficiência econômica foi formada por 17 dos 599 Municípios Paulistas sob análise. Os outros 582 Municípios se encontravam abaixo da fronteira, ou seja, deveriam oferecer mais serviços de saúde dada a despesa com a subfunção atenção básica. Os resultados sinalizaram que 2,83% dos municípios foram considerados eficientes e 97,17% ineficientes.

O escore mínimo de eficiência econômica foi de 2,62% e o máximo de 100% - medida de *Shephard*. Dos 599 municípios, 25% apresentaram escores de eficiência menores do que 13,87%, metade apresentou escores menores do que 21,27% e 25% maiores do que 32,93%, entre estes últimos estão aqueles classificados como eficientes.

Quando se analisou somente os Municípios ineficientes, a média dos escores de eficiência foi de 25,8%, indicando que se poderia aumentar a oferta de produtos diretos da atenção básica em 74,2% sem necessidade de direcionar mais recursos do orçamento público para a área da saúde. Pela avaliação geral das metas de melhorias dos *outputs*, a ação dos Municípios deveria estar voltada para aumentar a cobertura do PACS e do PSF e a quantidade de procedimentos ofertados por outros profissionais de nível superior e de odontologia.

No segundo estágio, os *inputs* não-controláveis foram as variáveis independentes e os logaritmos naturais dos escores de *Farrell*-eficiência (o inverso da medida *Shephard*-eficiência usada no primeiro estágio) a variável dependente da análise de regressão múltipla.

A Tabela 1 demonstra o resultado da análise de regressão clássica realizada com o método *stepwise*. Os pressupostos do modelo de regressão – normalidade dos resíduos e homocedasticidade – foram testados e os resultados apontaram para a aceitação do modelo. Todos os coeficientes podem ser considerados diferentes de zero a um nível de significância de 0,05. As variáveis faixa-etária I (população entre 0 e 18 anos) e percentual de recursos próprios não foram adicionadas ao modelo, indicando que, na presença das outras variáveis, tais fatores não ajudaram a explicar a eficiência da obtenção de *outputs* diretos.

Tabela 1 – Resultados da regressão do modelo DEA em dois estágios

$R^2 = 0,1817$ e R^2 Ajustado = $0,1762$

	Coefficientes não Padronizados	Coefficientes Padronizados	Sig.
Constante	1,47382		0,00
Faixa-etária II	0,06528	0,2391	0,00
Densidade Populacional	-0,00010	-0,1627	0,00
População Urbana	-0,00671	-0,1381	0,00
Escala dos Estabelecimentos	-0,00001	-0,0921	0,02

Lembrando que a medida *Farrell*-eficiência varia de 1 a ∞ , quanto maior o escore mais ineficiente é o Município. Pelos coeficientes padronizados da regressão, é possível perceber que a maior proporção de idosos na população aumenta a ineficiência de uma jurisdição e esta foi a variável que mais impactou nos escores de eficiência. Por outro lado, algumas variáveis de estado ou *inputs* não-controláveis favoreceram a obtenção de melhores resultados pelos Municípios, como aconteceu com a densidade populacional, percentual de população urbana e a escala de estabelecimentos. De fato, ter uma população menos dispersa e com maior grau de urbanização reduz o custo da prestação de serviços em atenção básica. Além disso, apesar de menos relevante, a escala com que se trabalha impacta no desempenho quanto à eficiência. Portanto, esses aspectos deveriam ser considerados nas decisões sobre a descentralização das ações de saúde.

Com base nos resultados da regressão, os escores de *Farrell*-eficiência foram ajustados somando-se aos termos de erro de regressão o erro máximo obtido. Para que os escores de eficiência do primeiro e do segundo estágio pudessem ser comparados, calculou-se o inverso do exponencial da medida de eficiência ajustada gerada neste segundo estágio. O escore ajustado mínimo de eficiência econômica foi de 1,89% e o máximo de 96,45%. Dos 599 municípios, 25% apresentaram escores de eficiência ajustados menores do que 9,20%, metade apresentou escores menores do que 13,58% e 25% maiores do que 20,05%. Os Municípios mais eficientes foram: Tuiuti, Nova Guataporanga, Sabino, Lins e Santos.

Os Municípios de Piracicaba, Rio Grande da Serra, Santo André, São Bernardo do Campo, São Caetano do Sul e São Paulo foram classificados como eficientes no primeiro estágio, mas não permaneceram no grupo dos 17 mais eficientes quando os *inputs* não-controláveis foram inseridos na análise. Observou-se que, no primeiro estágio, a população dos Municípios era maior do que no segundo estágio. Dos grandes municípios permaneceram Santos e São Carlos.

Borá apareceu como o Município mais ineficiente nos dois estágios, vale ressaltar que este é o menor município do Estado de São Paulo. Com exceção de Louveira, todos os 10 Municípios mais ineficientes possuíam população abaixo de 10.000 habitantes. Mesmo com a exclusão do efeito de variáveis que poderiam favorecer os grandes Municípios na avaliação de desempenho comparativa, somente duas jurisdições saíram do grupo dos 10 mais ineficientes: Santana da Ponte Pensa e Aspásia. Aparentemente, pelas suas características, tais municípios são ineficientes por natureza.

Com base nos escores de eficiência ajustados (medida de Farrell) foi realizado o teste de hipóteses. A Tabela 2 demonstra o resultado da análise de regressão clássica realizada com o método *stepwise*. Todos os coeficientes podem ser considerados diferentes de zero a um nível de significância de 0,05. Os pressupostos do modelo de regressão – normalidade dos resíduos e homocedasticidade – também foram testados. Os resultados apontaram para a aceitação do modelo.

Tabela 2 – Resultados da regressão do teste de hipóteses

$R^2=0,3441$ e R^2 Ajustado = 0,3386

	Coeficientes não Padronizados	Coeficientes Padronizados	Sig.
Constante	2,41743		0,00
TG	0,01285	0,2816	0,00
PABF	-0,34289	-0,2085	0,00
PABV	-0,25456	-0,4326	0,00
IE	-0,00806	-0,1035	0,00
SUSNV	-0,03015	-0,1111	0,01

Pelos coeficientes padronizados, percebe-se que a variável que mais influenciou a eficiência ajustada foram as transferências do Piso da Atenção Básica Variável (PABV). Quanto maior a representatividade de tais transferências nas receitas correntes menor a ineficiência do Município. Aparentemente os mecanismos de incentivos do governo federal têm contribuído para a otimização da utilização dos recursos orçamentários. Contudo, não se pode perder de vista que a provisão de serviços públicos pelos governos locais se justifica pela maior possibilidade de ajuste às preferências dos cidadãos, o que seria prejudicado pela maior vinculação de recursos pelo governo federal.

A segunda variável mais significativa foram as transferências gerais (TG) compostas pelos principais repasses de recursos não-condicionais e sem contrapartida para os Municípios. Neste caso, como previsto pela literatura, quanto maior a dependência por este

tipo de transferência maior a chance de ocorrer o gasto público com desperdício, haja vista os incentivos para a flexibilização das restrições orçamentárias.

Pela ordem de relevância das variáveis, em seguida, têm-se as transferências do Piso da Atenção Básica Fixo (PABF) e do SUS não-vinculadas (SUSNV), esta última inclui a primeira mais as transferências para o atendimento de média e alta complexidades. O PABF caracteriza-se por uma transferência em bloco para a subfunção atenção básica, apesar de ser não-condicional e sem contrapartida influenciou positivamente o desempenho dos Municípios. Uma das possíveis explicações para este comportamento seria o fato de os governos locais terem que estabelecer metas e compromissos nas comissões inter-gestores e essa dinâmica do SUS favorecer a otimização do gasto. Em relação ao SUSNV, além do mencionado para o PABF, podem ter ocorrido interações positivas entre o atendimento nos diversos níveis de complexidade.

Por fim, o indicador de escolaridade (IE) do IPRS apresentou interdependência negativa com o grau de ineficiência dos Municípios. Cidadãos mais informados influenciam positivamente o desempenho do setor público ligado à atenção básica, talvez por favorecer a transparência e *accountability* da gestão pública municipal, facilitando o acesso à informação e, assim, aumentando o poder de barganha por parte dos cidadãos.

Três variáveis não foram incluídas no modelo pelo método *stepwise*: Transferências discricionárias (TD), transferências dos SUS vinculadas (SUSV) e Indicador de Renda (IR). Apesar de a variável TD não apresentar correlações com as outras variáveis acima de 0,3, ela não foi inserida na análise, indicando que de fato não influencia os escores de eficiência, aparentemente, não sendo fonte de flexibilização das restrições orçamentárias. No caso do SUSV, a variável possuía correlação de 0,9 com a variável PABV, haja vista ser essa a principal fonte de recursos vinculados do SUS. Já o Indicador de Renda apresentou correlações negativas de 0,55 com TG, 0,31 com SUSV e 0,42 com PABV, ou seja, quanto mais rico o Município menores valores destes três tipos de transferência ele recebeu. IR também apresentou correlação positiva de 0,31 com o SUSNV. Diante deste quadro, a variável IR foi excluída do modelo.

Com base nos resultados da análise de regressão, as conclusões sobre as hipóteses do estudo foram as seguintes:

- Não se pode rejeitar H1: Quanto maior o grau de dependência de recursos transferidos por outras esferas de governo para financiamento das políticas públicas municipais, menor a eficiência econômica na provisão de serviços de atenção básica à saúde.
- Parcialmente, não se pode rejeitar H2: Quanto maior o percentual de transferência vinculada do governo federal, geral ou específica da atenção básica, em relação às receitas correntes, maior a eficiência econômica dos Municípios na subfunção atenção básica à saúde. A influência é somente das transferências específicas da atenção básica.
- Parcialmente, não se pode rejeitar H3: Quanto piores as condições socioeconômicas dos municípios, menor a eficiência econômica na subfunção atenção básica à saúde pelos Municípios. Quando se analisa o indicador de escolaridade, não é possível rejeitar a hipótese, o mesmo não ocorrendo com o indicador renda. Todavia, há que se pensar que este último está sendo retratado pelas outras variáveis incluídas no modelo com as quais apresentava altas correlações.

5 CONSIDERAÇÕES FINAIS

O presente artigo teve como objetivo medir e explicar as variações de desempenho dos Municípios Paulistas quanto à eficiência técnica na aplicação de recursos públicos nas ações

de atenção básica à saúde em função do perfil de financiamento dos gastos gerais e específicos de tal área.

Para isso, foi usado o modelo DEA em dois estágios. No primeiro estágio, os escores de eficiência dos Municípios Paulistas foram calculados com base em um *input* discricionário – despesa liquidada com atenção básica – e seis *outputs* diretos representativos da atenção básica à saúde – número de procedimentos neste nível de atenção e cobertura dos programas PSF e PACS. Os resultados indicaram que somente 17 dos 599 Municípios sob análise foram considerados eficientes e que era possível aumentar, consideravelmente, a quantidade de serviços prestados à população sem a necessidade de aumentar as dotações orçamentárias para a área de saúde.

No segundo estágio do modelo DEA, buscou-se identificar as variáveis que afetam o desempenho do gestor público e a direção em que o fazem. Verificou-se que a maior proporção de idosos em uma jurisdição tornou a prestação de serviços mais cara. Por outro lado, maiores densidade populacional, grau de urbanização e escala dos estabelecimentos de saúde favoreceram o gasto público com eficiência. Em média, os escores de eficiência ajustados foram menores do que os originalmente calculados, indicando que as variáveis não-controláveis tiveram um impacto líquido positivo.

Apesar de as variáveis não estarem sob o controle do gestor público municipal, ele não pode ignorá-las, pois afetam, diretamente, o desempenho do sistema de saúde municipal. Além do que, se for pensado que o SUS é uma política compartilhada entre os entes federativos, estes aspectos devem ser considerados nas decisões conjuntas, principalmente, no processo de regionalização da atenção à saúde e na pactuação de metas. A atenção básica quando bem realizada pode evitar diversos custos adicionais ao sistema de saúde.

Com o ajuste dos escores de eficiência, Municípios considerados eficientes na primeira análise deixaram de sê-lo no segundo estágio, sinalizando que o desempenho do gestor municipal tinha sido favorecido pelas variáveis ambientais. Analisando o grupo dos mais ineficientes, percebeu-se que a composição quase não se alterou e que, predominantemente, era formado por municípios de pequeno porte. Os resultados são indicativos de que a configuração destes municípios, por natureza, leva-os à ineficiência.

Sabe-se que Municípios pequenos são fortemente dependentes de transferências intergovernamentais e, a despeito de se constituírem em um dos principais instrumentos fiscais, o desenho dos sistemas de transferências pode provocar comportamentos e resultados indesejáveis ao bem-estar social. A literatura sobre o federalismo prevê que as transferências não-condicionais e sem contrapartida provocam gasto público com desperdício, ocasionado pelo fenômeno conhecido como *Flypaper Effect*, sendo a redução do poder de barganha dos cidadãos uma das explicações para tal fenômeno. Mesmo no caso do FPM, principal transferência, que poderia ser carregado para outra jurisdição, o poder de barganha seria reduzido pelo alto custo de se obter informações.

A partir dos escores de eficiência ajustados e por meio da análise de regressão, constatou-se que as transferências não-condicionais e sem contrapartida aumentaram a ineficiência do gasto público em atenção básica, conforme previsto pela literatura. Por outro lado, os repasses de recursos do SUS, tanto os não-vinculados quanto os vinculados, reduziram a ineficiência, indicativo dos avanços alcançados pela gestão do SUS. Outro fator que apresentou interdependência positiva com a eficiência foi o indicador de escolaridade, sinalizando que uma população mais bem educada pode favorecer a avaliação de desempenho e *accountability*, ou seja, o controle social, propiciando, destarte, maior poder de barganha dos cidadãos.

Este trabalho focou a atenção básica para permitir uma avaliação comparativa de desempenho entre Municípios. Entretanto, também é importante avaliar as outras subfunções de saúde, como o atendimento ambulatorial de média e alta complexidades e o hospitalar. Para isso, poder-se-iam segmentar as atividades com as mesmas características de produção entre os Municípios ou, segmentar os Municípios conforme sua capacidade de produção. Outra forma seria avaliar as regiões ou macrorregiões de saúde como centro de responsabilidades e não o Município.

Enfim, para além dos resultados empíricos apresentados, acredita-se que este trabalho tenha contribuído para indicar possibilidades e restrições quanto à realização de avaliações de desempenho comparativas em sistemas federalistas, apresentando uma metodologia que auxilie os avanços nos estudos sobre Controladoria na Gestão Pública.

REFERÊNCIAS

BARDHAN, P.; MOOKHERJEE, D. *Decentralizing antipoverty program delivery in developing countries*. **Journal of Public Economics**. Amsterdam, v. 89, n. 4. p. 675-704, Apr. 2005.

BRADFORD, D. F.; MALT, R. A.; OATES, W. E.. The rising cost of local public services: some evidence and reflections. **National Tax Journal**. Washington, v. 22, n. 2, p. 185-202, Jun. 1969.

BRADFORD D. F.; OATES, W. E. *The analysis of revenue sharing in a new approach to collective fiscal decisions*. **Quarterly Journal of Economics**. Cambridge, v. 85, p. 416-439, 1971.

BRASIL. Constituição (1988). Constituição da República Federativa do Brasil. **Coleção Saraiva de Legislação**. 21. ed. São Paulo: Saraiva, 2001.

_____. Constituição (1988). Emenda Constitucional nº 29, de 14/02/2000. **Legislação do SUS**. Brasília: CONASS, 2003.

CHARNES, A.; COOPER, W.W.; RHODES, E. Measuring the efficiency of decision making units. **European Journal of Operational Research**. Washington, v.2, p.429-444, 1978.

COHEN, E.; FRANCO, R. **Avaliação de projetos sociais**. 6. ed. Petrópolis, RJ: Vozes, 2004.

DEBREU, G. *The coefficient of resource utilisation*. **Econometrica**. Evanston, v. 19, n. 3, p. 273-292, 1951.

FARRELL, M. J. *The measurement of productive efficiency*. **Journal of the Royal Statistical Society**. London, Series A., v. 120, n. 3, p. 253-290, 1957.

FISHER, R. C. *Income and grant effects on local expenditure: the flypaper effect and other difficulties*. **Journal of Urban Economics**. New York, v. 12, n. 3, p. 324-345, 1982.

FUNDAÇÃO SEADE. Disponível em: <<http://www.seade.sp.gov.br>>. Acesso em: 05/03/2008.

GREENE, W. H. *Maximum likelihood estimation of econometric frontier functions*. **Journal of Econometrics**. North Holland, v. 13, n. 1, p. 27-56, May. 1980.

INMAN, R. P. *The fiscal performance of local governments: an interpretative review*. In: MIESZKOWSKI, P.; STRASZHEIM, M. (Eds.). **Current issues in urban economics**. London: The Johns Hopkins Press Ltd., 1979.

KOOPMANS, T. C. *An analysis of production as an efficient combination of activities*. In: _____. *Activity analysis of production and allocation*. Cowles Commission for Research in Economics, Monograph n. 13. New York: John Weley and Sons, Inc. 1951.

NISKANEN, W. A. *Bureaucracy and representative government*. Chicago: Aldine-Atherton, 1971.

SHAH, A. *A practitioner's guide to intergovernmental fiscal transfers*. In: BOADWAY, R.; SHAH, A. *Intergovernmental fiscal transfers: principles and practice*. Washington, D.C.: World Bank, 2007.

SHEPHARD, R. W. *Theory of Cost and Production Function*. Princeton: Princeton University Press, 1970.

SOUSA, M. C. S. de; STOŠIĆ, B. *Technical efficiency of the Brazilian municipalities: correcting nonparametric frontier measurements for outliers*. *Journal of Productivity Analysis*. Norwell, n. 24, p. 157-181, 2005.

WYCKOFF, P. G. *A bureaucratic theory of flypaper effects*. *Journal of Urban Economics*. Nova York, v. 23, n. 1, p. 115-129, Jan. 1988.

ⁱ “dinheiro cola onde ele bate”.