

FLEXIBILIZAÇÃO DAS RESTRIÇÕES ORÇAMENTÁRIAS E EFICIÊNCIA ECONÔMICA DOS MUNICÍPIOS PAULISTAS: UMA AVALIAÇÃO DA ATENÇÃO BÁSICA À SAÚDE

Patrícia Siqueira Varela
UNIVERSIDADE DE SÃO PAULO

Gilberto de Andrade Martins
UNIVERSIDADE DE SÃO PAULO

Luiz Paulo Lopes Fávero
UNIVERSIDADE DE SÃO PAULO

RESUMO

Este trabalho teve por objetivo levantar, medir e explicar as variações de desempenho dos Municípios Paulistas quanto à eficiência econômica 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 econômica reflete a capacidade de uma entidade obter máximos *outputs* ao menor custo e foi medida com o uso da técnica *Data Envelopment Analysis (DEA)*: modelo em dois estágios. Por sua vez, a eficiência econômica é um dos parâmetros de avaliação de desempenho do setor público e é influenciada pela forma como os políticos e burocratas lidam com as restrições orçamentárias. Tais restrições podem ser mais suaves ou mais fortes em função do perfil de financiamento das políticas públicas, conforme determina a literatura sobre o federalismo fiscal. Para verificar a relação de dependência entre eficiência econômica 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. Além disso, verificou-se que a maior proporção de idosos em uma jurisdição torna a prestação de serviços mais cara, o contrário acontecendo com maiores densidade populacional, grau de urbanização e escala dos estabelecimentos de saúde. E, principalmente, as transferências não-condicionais e sem contrapartida geram um efeito negativo no escore de eficiência, enquanto os repasses de recursos do Sistema Único de Saúde, tanto os não-vinculados quanto os vinculados, efeito positivo.

Agradecimentos: Os autores agradecem à FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) pelo apoio financeiro concedido ao projeto de pesquisa, do qual resultou este artigo.

1 INTRODUÇÃO

O Brasil, há mais de um século, adotou a forma federativa de divisão territorial e, hoje, é formado por 26 Estados, o Distrito Federal e 5563 Municípios divididos em cinco regiões. Um dos principais problemas do federalismo brasileiro é a heterogeneidade econômica entre regiões, também presente entre Estados e Municípios.

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. O Sistema Único de Saúde (SUS) é caracterizado por um sistema

complexo de relações intergovernamentais. O governo federal estabelece normas e age como coordenador das ações dos entes federativos, buscando a concretização do atendimento universal, equitativo e integral ao cidadão. Os Municípios são os principais provedores de serviços na área da saúde, com responsabilidades exclusivas quanto à subfunção atenção básica à saúde.

Enquanto política setorial de governo, o SUS tem sido influenciado pelos esforços recentes de modernização da administração pública, tendo como um dos seus pilares a disseminação dos princípios da administração pública gerencial, cujo foco é a gestão voltada para resultados e o uso eficiente dos recursos. Nesse contexto, o programa aparece como o elemento integrador dos planos e orçamentos e como unidade de gestão das ações governamentais. Assim, a avaliação de desempenho quanto à efetividade, eficácia e eficiência dos programas governamentais ganha destaque.

Todavia, a determinação de parâmetros de avaliação pela Controladoria na Gestão Pública não é uma tarefa simples, apresentando 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 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 foco foi nos *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.

O orçamento é considerado um dos principais veículos do processo de planejamento, execução e controle. Todavia, Merchant e Otley (2007, p. 794), citando Fraser e Hope (2003), argumentam que a determinação de metas depende da conciliação entre os aspectos motivacionais e o planejamento realístico dos números do orçamento. São sugeridos dois princípios para resolver este problema: remover a ênfase em metas fixas e presentes e substituí-las por padrões de desempenho por *benchmark* e ajustar as avaliações de desempenho para refletir a experiência de operação vigente e circunstâncias econômicas enfrentadas durante o período.

Sendo assim, este estudo teve por objetivo levantar, medir e explicar as variações de desempenho dos Municípios Paulistas quanto à eficiência econômica 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 econômica reflete a capacidade de uma entidade obter máximos *outputs* ao menor custo e é influenciada pela forma como os políticos e burocratas lidam com as restrições orçamentárias. A literatura sobre o federalismo prevê que 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 econômica na provisão de serviços de atenção básica à saúde;

- H2: Quanto maior o percentual de transferências discricionárias em relação à receita corrente, menor a eficiência econômica dos Municípios na subfunção atenção básica à saúde;
- H3: 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 avaliação comparativa de desempenho permite identificar os pontos fracos da gestão, assim como aqueles que servem como referência para outras jurisdições, concretizando uma das vantagens das estruturas federalistas, como é o caso do Brasil, o aprendizado pela diversidade de experiências. Tratar de uma área tão complexa quanto a saúde e buscar avaliar a eficiência dos Municípios em relação ao perfil de financiamento lança luz sobre as possibilidades de transposição das abordagens tradicionais da Controladoria desenvolvidas para empresas para organizações governamentais. Espera-se, construir bases empírico-teóricas para o desempenho das funções de Controladoria na Gestão Pública.

Nos benefícios esperados, incluem-se aumento do controle social sobre os gastos públicos, melhoria dos serviços prestados e redução de custos. Além disso, pode ser uma alternativa para as entidades de controle externo, mormente tribunais de contas, realizarem auditoria operacional nos entes públicos sob sua jurisdição.

2 TRANSFERÊNCIAS INTERGOVERNAMENTAIS E O PROBLEMA DE *SOFT BUDGET CONSTRAINTS*

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.

Todavia, definir um bom sistema de transferências é uma tarefa delicada, pois as características necessárias para alcançar um objetivo não são desejáveis para alcançar outros igualmente importantes. Assim, elas têm se desenvolvido “[...] como produtos de conveniências administrativas ou de pressões políticas e, em muitos casos, podem ser facilmente aperfeiçoadas no tocante a custos técnicos e mesmo custos políticos baixos.” (PRUD’HOMME, 1995, p. 213).

A idéia subjacente a sistemas apropriados de transferências é que tanto o governo central quanto os governos locais têm interesse em prover bens e serviços para as pessoas de uma nação. Entretanto, em certos cenários, as transferências podem servir como um veículo para uma jurisdição explorar outras ao transferir os custos de seus programas para a nação como um todo. Isso vai depender dos incentivos existentes sob certas formas predominantes de instituições fiscais, as quais conduzirão a restrições orçamentárias flexíveis ou rígidas.

De acordo com Kornai (1986, p. 23), em muitos países, os governos locais têm menor ou maior autonomia financeira e espera-se que sejam auto-suficientes, ou seja, capazes de obter tributos e outras receitas para cobrir suas despesas. Quando o governo local obtém fundos adicionais do orçamento governamental de outras esferas, o que envolve poder de barganha, pode ocorrer o fenômeno *Soft Budget Constraint*. Se o governo local incorre em *déficits*, ele pode esperar ser socorrido por autoridades de esferas mais altas de governo. “As chances são muito boas de que mesmo o gasto sem cuidado não leve a uma catástrofe financeira”.

A terminologia *Soft Budget Constraint* tem sua origem no trabalho seminal de Kornai (1980) e se refere ao fenômeno que ocorre quando uma entidade em dificuldades financeiras espera que seus *déficits* orçamentários, parcialmente ou totalmente, sejam cobertos por uma organização de suporte (KORNAI *et al.*, 2003, p. 1097). De acordo com Kornai (1986, p. 5), cada tomador de decisão tem uma percepção subjetiva da distribuição de probabilidade quanto ao recebimento de assistência externa, caso as receitas não sejam suficientes para cobrir as despesas orçamentárias. Quanto maior a probabilidade subjetiva de que o excesso de despesa será coberto por assistência externa, mais flexíveis serão as restrições orçamentárias.

Um dos impactos da flexibilização das restrições orçamentárias é sobre a eficiência alocativa que não pode ser alcançada quando as combinações de *input-output* não se ajustam aos sinais de preço. Este fato, por sua vez, acontece porque, diferentemente da situação de *Hard Budget Constraints*, as entidades têm uma fraca compulsão para ajustar a demanda aos preços relativos (KORNAI, 1986, p. 10).

Conforme Kornai *et al.* (2003, p. 1097), *Soft Budget Constraints* não se restringem às relações intergovernamentais, podendo ser também encontradas em empresas estatais e privadas, bancos, organizações não-lucrativas e até economias nacionais. Neste trabalho, especificamente, o termo se refere a governos locais que buscam socorro financeiro nos governos de esferas mais altas quando estão em situação fiscal difícil. O socorro financeiro (*bailout*) pode vir na forma de transferências intergovernamentais ou, alternativamente, o governo central pode assumir dívidas ou realizar empréstimos com o intuito de evitar crises fiscais de governos locais.

Oates (2005, p. 361), com base nos estudos de Goodspeed (2002) e Wildasin (1997), descreveu os incentivos que o governo central teria para prestar socorro financeiro, quais sejam: (i) nível de *output* local abaixo do desejado pode ter conseqüências eleitorais para as autoridades centrais; (ii) falhas de governos locais podem repercutir em outras jurisdições, tendo maior conseqüência quanto maior for a jurisdição provocadora do problema “*Too big to fail*”; (iii) colapso financeiro muito grande por um governo local pode sugerir que as autoridades centrais, de alguma maneira, falharam em estabelecer responsabilidades apropriadas; e (iv) elementos de corrupção presentes em formas padrão de suborno na fonte do problema.

Por outro lado, os principais argumentos para *Soft Budget Constraints* nas entidades que esperam receber assistência de outras entidades são: (i) proteção de empregos; (ii) proteção da produção doméstica contra competição estrangeira; (iii) objetivos redistributivos em nome da equidade, justiça social e solidariedade; (iv) salvaguarda do indivíduo e, enfim, da sociedade como um todo contra flutuações e incertezas, o que se relaciona com os argumentos (i) e (iii); (v) importância social do propósito particular da instituição (KORNAI, 1986, p. 25). Apesar de o autor falar das organizações de maneira geral, com exceção do argumento (ii), todos os outros são perfeitamente aplicáveis às relações intergovernamentais.

De acordo com Oates (2005, p. 362), para evitar os efeitos desestabilizadores e distorcivos das *Soft Budget Constraints*, é relevante determinar os tipos de instituições fiscais que podem acomodar a descentralização fiscal a fim de concretizar as vantagens políticas e os ganhos econômicos pelo controle local. As condições essenciais para tornar as restrições orçamentárias mais fortes seriam:

- A existência de um sistema de tributação local confiável e efetivo para fornecer as receitas necessárias ao financiamento dos programas locais. As jurisdições locais precisam determinar o peso dos benefícios dos programas propostos contra seus custos. Um bom sistema tributário torna os custos aparentes para o eleitorado local e constitui a fonte de recursos financeiros.

- O sistema de transferências intergovernamentais deve atender suas funções alocativas e distributivas básicas sem estar sujeito a manipulações para prestar socorro financeiro, ou seja, não pode ser expansivo a ponto de permitir transferência adicional para aliviar pressões fiscais locais.

Rodden *et al.* (2003, p. 13) afirmam que diversos aspectos das relações fiscais intergovernamentais podem desempenhar um papel crítico para tornar as restrições orçamentárias mais fortes: (i) transparência e capacidade de prever as transferências intergovernamentais; (ii) habilidade de responder às necessidades locais pelo levantamento de receitas ao invés de contar, pesadamente, com transferências locais (dependência de transferências); e (iii) designação de despesas para os níveis apropriados de governo e flexibilidade adequada na determinação das despesas locais.

Em relação ao primeiro aspecto, a não ser que o financiamento entre governos seja completamente independente das escolhas das jurisdições receptoras, provavelmente, existirá o problema de risco moral por parte de cada jurisdição por ter o incentivo inadequado de que os cidadãos de outras jurisdições pagarão por parte de seus programas. Assim, a tendência é de expansão dos programas e aumento ou exagero de seus custos. Jurisdições nacionais tentarão explorar o centro (e cada uma das outras jurisdições) e o desafio do desenho institucional é conter os custos deste jogo (RODDEN *et al.*, 2003, p. 14).

Quanto ao segundo aspecto, um governo local altamente dependente de transferências e empréstimos do governo central, quando enfrenta choques adversos, tem pouca flexibilidade para obter receita adicional. Assim, as pressões dos eleitores, empregados e credores, possivelmente, serão sobre o governo central que pode resolver a crise corrente e os políticos locais não serão considerados responsáveis pelos *outcomes* fiscais. Entretanto, além de verificar o percentual de transferências sobre a receita total (grau de dependência), é preciso apurar se as regras não são facilmente manipuláveis (RODDEN *et al.*, 2003, p. 14-15).

Já no último aspecto, o governo central poderia ser incapaz de opor-se a demandas de *bailout*, se governos locais falhassem em fornecer certos serviços-chave para o eleitorado nacional, o que é agravado se existirem políticas nacionais, tais como de padrão mínimo de serviço. Justamente por isso, transferência de custos ou débitos seria mais difícil se o governo central não se envolvesse na provisão de bens públicos puramente locais. Assim como a dependência por transferências, a regulação central de responsabilidades fiscais locais pode emitir um forte sinal para políticos, eleitores e credores de que o governo central, em última instância, é responsável pelos resultados fiscais locais (RODDEN *et al.*, 2003, p. 15).

Diante do exposto, é possível dizer que o sistema de transferências pode favorecer ou não o fortalecimento das restrições orçamentárias e, conseqüentemente, a eficiência na provisão de serviços públicos. Isso vai depender de como o sistema for desenhado e de outras variáveis, tais como os limites legais impostos ao desempenho fiscal dos governos locais.

3 PRODUÇÃO DE BENS PÚBLICOS E DATA ENVELOPMENT ANALYSIS

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 *et al.* (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 *et al.*, 1969, p. 187).

Pela análise dos argumentos de Bradford *et al.* (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. 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%.

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. No primeiro estágio, uma fronteira de produção é estimada a partir dos *inputs* discricionários e *outputs* de uma amostra de DMUs, sem considerar as variáveis não-controláveis. No segundo estágio, os escores de eficiência são tomados como variável dependente em um modelo de regressão e as variáveis ambientais como variáveis independentes. Os sinais dos coeficientes dos termos da regressão indicam a direção da influência das variáveis não-controláveis nos escores de eficiência e os testes de hipóteses padrões são usados para avaliar a força da relação. A regressão do segundo estágio também permite ajustar todos os escores de eficiência pelo uso dos coeficientes estimados, em um ambiente comum.

4 TRAJETÓRIA METODOLÓGICA

4.1 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.

4.2 Escore de Eficiência Econômica

Neste primeiro estágio, 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.

Para levantar e medir a eficiência relativa da aplicação de recursos na subfunção atenção básica à saúde pelos Municípios Paulistas, foi usada a *Data Envelopment Analysis* com orientação para a maximização de *outputs*, com o intuito de verificar o grau de reversão da função orçamento-*output*. Quanto ao rendimento de escala, análises preliminares indicaram que seria variável, o que foi seguido no restante dos procedimentos adotados.

Para os Municípios Paulistas, foi calculada uma função de custo 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	SIA/DATASUS

		neste tipo de ação.	
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.

4.3 Variáveis não-Controláveis

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	DATASUS

		mais.	
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_i^A é: $\theta_i^A = \varepsilon_i + \text{Max}_n \varepsilon_n$.

4.4 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_i^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, do recebimento de transferências discricionárias. Dados sobre o perfil socioeconômico 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.

5 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*. 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 em atenção básica.

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

R² = 0,1817 e R² Ajustado = 0,1762

	Coeficientes		Sig.
	não Padronizados	Padronizados	
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

Os pressupostos do modelo de regressão – normalidade dos resíduos e homocedasticidade – também foram testados. Os resultados apontaram para a aceitação do modelo.

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: Tututi, 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

R² = 0,3441 e R² 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.

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.
- Não se pode aceitar H2: Quanto maior o percentual de transferências discricionárias em relação à receita corrente, menor a eficiência econômica dos Municípios na subfunção atenção básica à saúde.
- Parcialmente, não se pode rejeitar H3: 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.

6 CONSIDERAÇÕES FINAIS

O presente artigo teve como objetivo levantar, medir e explicar as variações de desempenho dos Municípios Paulistas quanto à eficiência econômica 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 como um todo.

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.

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*.

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, apresentando uma metodologia que auxilie os avanços nos estudos sobre Controladoria na Gestão Pública.

REFERÊNCIAS

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.

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.

FRASER, J.; HOPE, R. *Beyond budgeting: how managers can break free of the annual performance trap*. Boston: Harvard Business School Publishing Corporation, 2003 *apud* MERCHANT, Kenneth A.; OTLEY, David T. *A review of the literature on control and accountability*. In: CHAPMAN, Christopher S. *et al. Handbook of Management Accounting Research*. Oxford: Elsevier, v. 2, 2007.

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

GOODSPEED, T. J. *Bailouts in a federation*. **International Tax and Public Finance**. Netherlands, v. 9, p. 409-421, 2002 *apud* OATES, Wallace E. *Toward a second-generation theory of fiscal federalism*. **International Tax and Public Finance**. Netherlands, v. 12, n. 4, p. 349-373, Aug. 2005.

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

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 Wiley and Sons, Inc. 1951.

KORNAI, J. **Economics of shortage**. Amsterdam: Elsevier, 1980.

_____. *The soft budget constraint*. **Kyklos**. Bern, v. 39, n. 1, p. 3-30, Feb. 1986.

KORNAI, J.; MASKIN, E.; ROLAND, G. *Understanding the soft budget constraint*. **Journal of Economic Literature**. Pittsburgh, v. 41, n. 4, p. 1095-1136, 2003.

MERCHANT, Kenneth A.; OTLEY, David T. *A review of the literature on control and accountability*. In: CHAPMAN, Christopher S. *et al. Handbook of Management Accounting Research*. Oxford: Elsevier, v. 2, 2007.

OATES, W. E. *Toward a second-generation theory of fiscal federalism*. **International Tax and Public Finance**. Netherlands, v. 12, n. 4, p. 349-373, Aug. 2005.

PRUD'HOMME, R. *The dangers of decentralization*. **The World Bank Research Observer**. Cary, v. 10, n. 2, p. 201-220, Aug. 1995.

RODDEN, J.; ESKELAND, G. S.; LITVACK, J. *Introduction and overview*. In: _____. (Ed.). **Fiscal decentralization and the challenge of hard budget constraints**. Cambridge: MIT PRESS, 2003.

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.

WILDASIN, D. E. *Externalities and bailouts: hard and soft budget constraints in intergovernmental fiscal relations*. *Unpublished Paper*, 1997 apud OATES, Wallace E. *Toward a second-generation theory of fiscal federalism*. ***International Tax and Public Finance***. Netherlands, v. 12, n. 4, p. 349-373, Aug. 2005.