

Eficiência Técnica e Retornos de Escala na Produção de Serviços Públicos Municipais: o Caso do Nordeste e do Sudeste Brasileiros*

Maria da Conceição Sampaio de Sousa**

Francisco S. Ramos***

Sumário: 1. Introdução; 2. Aspectos metodológicos; 3. Dados e parâmetros; 4. Mensuração dos níveis de eficiência: discussão dos resultados para os casos do Nordeste e do Sudeste brasileiros; 5. Economias de escala e eficiência: mensurando os custos associados à política atual de descentralização municipal no Brasil; 6. À guisa de conclusão.

Palavras-chave: eficiência técnica; *data envelopment analysis*; gastos municipais.

Códigos JEL: H41 e C61.

Este artigo visa a avaliar a eficiência dos municípios das regiões Nordeste e Sudeste do Brasil no que se refere à utilização dos recursos públicos. A questão é saber se os gestores municipais alocam esses recursos de forma a garantir o máximo de serviços públicos ou, alternativamente, se para uma determinada oferta de bens públicos a despesa exigida é mínima. Por meio de técnicas não-paramétricas de mensuração de eficiência produtiva, busca-se analisar a relação entre o tamanho do município e os níveis de eficiência alcançados, além de investigar o tipo de economia de escala que caracteriza as municipalidades. Os resultados mostram que, nas duas regiões, para a imensa maioria dos municípios com população inferior a 15 mil habitantes, os baixos níveis de eficiência se explicam pela presença de economias crescentes de escala. Os resultados, embora exploratórios, sugerem que a atual política de descentralização no Brasil não promove o uso eficiente dos recursos públicos: a proliferação de pequenas municipalidades conduz ao aumento dos custos médios dos serviços municipais. Portanto, em virtude do seu pequeno tamanho, esses municípios não podem se beneficiar das economias de escala inerentes à produção de determinados serviços públicos. O desperdício de recursos reflete-se nos elevados níveis de excesso de dispêndio que caracterizam as cidades menores.

*Este artigo recebido em jan. 1999 e aceito em out. 1999, foi originariamente apresentado no XX Encontro Brasileiro de Econometria, realizado em Vitória, ES, em 1998. Os autores agradecem os comentários dos participantes do referido encontro, bem como as sugestões de dois pareceristas anônimos. Agradecem, ainda, ao Ipea e ao CNPq pelo apoio durante a elaboração dessa pesquisa.

**UnB (e-mail: mcoss@unb.br).

***UFPE (e-mail: fsr@npd.ufpe.br).

This paper evaluates the efficiency of the municipal districts of the Northeast and Southeast of Brazil in what concerns to the use of public resources. The question is to learn if the municipal administrators allocate those resources to guarantee the maximum of public services or, alternatively, if for a certain offer of public goods the expense is minimum. Using non-parametric techniques to measure productive efficiency, the paper analyzes the relationship between the size of the municipal district and its efficiency levels, besides investigating the existence of scale economies. Results show that, for the immense majority of the municipal districts with population of less than 15 thousand, the low efficiency levels are explained by the presence of increasing scale economies. The results, although exploratory, suggest that the current decentralization policy in Brazil doesn't promote the efficient use of the public resources: the proliferation of small municipalities drives to the increase of the medium costs of the municipal services.

1. Introdução

Existe atualmente um amplo consenso de que a descentralização representa um avanço considerável na produção e na distribuição dos bens e serviços públicos locais. Inúmeras razões explicam por que o poder público local possui uma vantagem evidente na provisão de tais serviços. Em primeiro lugar, a proximidade dos usuários favorece a identificação das prioridades a serem estabelecidas quando da alocação dos recursos públicos. Ademais, torna-se muito mais fácil fiscalizar eficientemente a utilização dos recursos quando eles não necessitam percorrer um longo caminho burocrático antes de chegar ao seu destino; a descentralização reduz, pois, as possibilidades de corrupção e ineficiência comumente encontradas no setor público. Finalmente, as exigências em termos de capacidade de gestão são menores e os problemas existentes mais simples, principalmente nas cidades de pequeno e médio portes.

Reconhece-se, portanto, a necessidade de responsabilizar os governos locais pela provisão de uma parcela significativa dos bens públicos postos à disposição da população. Porém, a mera descentralização, por si só, não garante uma oferta adequada dos bens e serviços mencionados. É preciso também garantir que os governos locais utilizem os recursos disponíveis da melhor maneira possível. No Brasil não existe, porém, qualquer evidência de que os recursos adicionais com os quais as prefeituras foram beneficiadas pela Constituição de 1988 estejam sendo alocados eficientemente, de modo a garantir o máximo de bem-estar à população.

Ademais, a proliferação de pequenos municípios, cuja viabilidade fiscal é discutível e cuja existência provavelmente acarreta custos superiores aos eventuais benefícios advindos de sua criação, tornou-se uma característica do movimento recente de transferência de poder político aos níveis inferiores de governo. Aqui, também, não se pode garantir que esse desmembramento de cidades leve em conta a escala mínima de operações exigida para que essas municipalidades sejam economicamente autônomas e capazes de assegurar aos seus habitantes a provisão eficiente de um determinado nível de serviços públicos sem incorrer em custos desnecessariamente elevados.

Torna-se, pois, necessário definir medidas de eficiência a partir das quais as políticas de gastos públicos possam ser avaliadas. A construção de indicadores para esses níveis de eficiência será, pois, a principal contribuição deste artigo, que visa a avaliar a eficiência dos municípios brasileiros no tocante à utilização dos recursos públicos. A questão é saber se os gestores municipais otimizam esses recursos alocando-os de forma a garantir o máximo de serviços públicos ou, alternativamente, se, para uma determinada oferta de bens públicos, a despesa exigida para o seu custeio é minimizada. Para tal, analisar-se-á a relação entre a despesa agregada e a provisão de bens públicos, em nível municipal. Trata-se, portanto, da determinação de uma fronteira de custo-eficiência. Esta fronteira será aproximada mediante a utilização de técnicas não-paramétricas de mensuração da eficiência produtiva. Atenção especial será dada à análise da relação entre o tamanho do município e os níveis de eficiência alcançados; poder-se-á, então, investigar o tipo de economia de escala que caracteriza as municipalidades analisadas.

Este artigo está organizado da seguinte maneira: a seção 2 apresenta e discute a metodologia utilizada para o cálculo dos níveis de eficiência dos municípios brasileiros; a seção 3 apresenta a base de dados do estudo, comenta a escolha dos indicadores utilizados como *proxies* para a oferta de serviços públicos municipais e descreve brevemente os aspectos computacionais; a seção 4 discute os resultados obtidos para as regiões Nordeste e Sudeste; a seção 5 tenta mensurar as economias de escala que prevalecem entre os municípios dessas regiões, e a seção 6 reúne algumas conclusões e sugere extensões.

2. Aspectos Metodológicos

A análise dos gastos públicos pode ser feita através de diferentes perspectivas. Entre elas, destacam-se os estudos que enfatizam o papel desempenhado

pela *demand*a na provisão de bens públicos e aqueles que privilegiam os aspectos ligados à *eficiência técnica* associada à produção destes bens.

A primeira abordagem baseia-se no trabalho pioneiro de Bergstrom e Goodman (1973). Estes autores consideram que os gastos públicos efetivamente realizados representam uma resposta do governo à *demand*a dos cidadãos/eleitores por bens públicos locais. Em consequência, eles relacionam as despesas municipais, desagregadas por tipo de serviço, com variáveis típicas da demanda, como os níveis de renda, os preços relativos e demais variáveis sócio-econômicas que caracterizam a população. Esse tipo de análise é particularmente apropriado para explicar os níveis e a distribuição dos diferentes serviços públicos municipais.

Um outro tipo de enfoque é aquele onde os aspectos de *eficiência* associados ao gasto público primam sobre os demais aspectos analisados. Esta será a abordagem utilizada em nosso estudo. Nessa análise, a questão é saber qual o menor custo ao qual é possível oferecer uma determinada quantidade de bens e serviços. Trata-se, portanto, de um problema clássico de minimização de custos. A ênfase neste tipo de metodologia é posta na identificação de uma fronteira de custo-eficiência referente à produção de bens públicos. A partir dessa fronteira poder-se-ão estimar os níveis de eficiência associados aos diferentes municípios.

2.1 Mensuração dos níveis de eficiência: abordagens paramétricas e não-paramétricas

A mensuração dos níveis de eficiência pode ser feita mediante o uso de diversos métodos. A distinção entre esses métodos encontra-se no tipo de técnica utilizada para descrever a fronteira do conjunto produtivo. Abordagens paramétricas e não-paramétricas podem ser usadas para determinar os limites desse conjunto. No primeiro caso – abordagem paramétrica – postula-se que a fronteira do conjunto produtivo pode ser representada por uma função de produção caracterizada por parâmetros constantes. Este método foi utilizado, pioneiramente, por Aigner e Chu (1968). Nele, especifica-se uma função de produção e, através de métodos estatísticos apropriados, estimam-se

os parâmetros dessa função.¹ Em seguida, para cada observação, a mensuração do nível de eficiência para insumos/produtos é feita em relação a essa função, cujo gráfico representa a fronteira do conjunto de produção. A maior limitação desse tipo de abordagem advém do fato de as medidas de eficiência obtidas através desse método variarem substancialmente de acordo com a forma funcional especificada; a própria definição dos critérios de eficiência está, pois, intrinsecamente ligada à seleção da forma funcional apropriada.

As abordagens não-paramétricas, em vez de se basearem em uma função especificada *a priori*, consideram apenas que o conjunto de produção deve satisfazer determinadas propriedades, tais como livre disponibilidade (*free disposal*), convexidade (implicando rendimentos de escala não-crescentes e rendimentos variáveis) ou proporcionalidade (implicando rendimentos constantes de escala). Nesse enfoque, o ponto central é determinar, de acordo com hipóteses preestabelecidas, que observações podem ser consideradas elementos da fronteira. Essa determinação é feita através da resolução, para cada observação, de um sistema de equações lineares definido para mensurar o nível de eficiência referente a cada observação.

Na extensa literatura existente sobre o assunto, esse método de determinação dos pontos eficientes é conhecido como DEA (*data envelopment analysis*, análise de envoltória de dados).² Inspirados no trabalho de Farrell (1957), precursor no uso deste tipo de abordagem, Charnes, Cooper e Rhodes (1978), em artigo seminal, introduziram o modelo DEA (conhecido como o modelo CCR), a partir do qual essa metodologia iria, subsequente, se desenvolver. A tecnologia é, nesse caso, especificada por meio de uma série de desigualdades que servirão de base para construir a fronteira tecnológica que representa a melhor alternativa de produção (*best practice*). Supondo-se $k = 1, \dots, K$ observações que combinam $n = 1, \dots, N$ insumos para $x^k = (x_{k1}, \dots, x_{kN})$ para produzir $m = 1, \dots, M$ produtos $y^k = (y_{k1}, \dots, y_{kM})$, a tecnologia que satis-

¹ Para uma discussão mais detalhada a respeito de metodologias para estimar fronteiras de produção paramétricas, ver os trabalhos de Førsund, Lovell & Schmidt (1980) e Bauer (1990).

² Ver, a esse respeito, Charnes, Cooper & Rhodes (1978 e 1981), Färe, Grosskopf & Lovell (1994), e Seiford & Thrall (1990), entre outros. A expressão DEA foi introduzida por Charnes, Cooper & Rhodes (1978). Porém, como deverá ficar claro ao longo deste artigo, todos os métodos apresentados envolvem dados. Entretanto, seguindo a tradição estabelecida na literatura, denominaremos DEA apenas aquelas técnicas que supõem alguma forma de convexidade e computam os graus de eficiência através de programas lineares.

faz as propriedades de retornos constantes de escala (C) e forte disponibilidade de insumos (S) pode ser descrita, em termos de insumos, como:³

$$L(y | C, S) = \left\{ x : y \leq zM, zN \leq x, z \in \mathbb{R}_+^K \right\}, \quad y \in \mathbb{R}_+^M \quad (1)$$

O vetor $z = (z_1, \dots, z_K) \in \mathbb{R}_+^K$, que contém as variáveis de intensidade de acordo com as quais as K atividades podem ser constituídas, permite construir variáveis não observadas, porém viáveis a partir de variáveis observadas. Trata-se, portanto, de pesos que possibilitam a construção dos segmentos lineares que definem a fronteira tecnológica.⁴

Podemos agora definir, para cada atividade, uma medida de eficiência técnica na utilização dos insumos, F_i , como:

$$F_i(y^k, x^k | C, S) = \min \left\{ \theta : \theta x^k \in L(y^k | C, S) \right\}, \quad k = 1, \dots, K \quad (2)$$

Essa expressão permite mensurar a eficiência do insumo x^k na produção de y^k , quando a tecnologia utilizada apresenta retornos constantes de escala e a disponibilidade forte dos insumos. F_i é homogênea de grau -1 em x^k , homogênea de grau 1 em y^k e, portanto, homogênea de grau 0 em (y^k, x^k) . Pode-se ainda demonstrar que essa medida é independente das unidades nas quais estão expressos os insumos e os produtos (Deprins, Simar & Tulkens, 1984; Färe, Grosskopf & Lovell, 1985 e 1994).

Utilizando a tecnologia especificada em (1), a medida de eficiência técnica para o município k pode ser computada como a solução do seguinte problema de programação linear:

$$F_i(y^k, x^k | C, S) = \theta^{DEA-C} = \min_{\theta, z} \theta \quad (3)$$

sujeito a:

$$\begin{aligned} \theta x_{kn} - \sum_{k=1}^K z_k x_{km} &\geq 0, & n = 1, \dots, N \\ \sum_{k=1}^K z_k y_{km} &\geq y_{km}, & m = 1, \dots, M \\ \theta, z_k &\geq 0, & k = 1, \dots, K \end{aligned}$$

³ Disponibilidade forte dos insumos define-se da seguinte forma: os insumos são fortemente disponíveis se para $x' > x \in L(y) \Rightarrow x' \in L(y)$, onde $L(y)$ representa o conjunto de todos os vetores de insumos que podem produzir $y \in \mathbb{R}_+^M$.

⁴ Para uma discussão mais detalhada a respeito da construção da tecnologia, ver Färe, Grosskopf & Lovell (1985 e 1994).

A solução desse problema indica a magnitude da redução radial exigida para que a atividade em questão pertença à fronteira tecnológica; portanto, o valor ótimo de θ situa-se entre 0 e 1, com as unidades eficientes apresentando θ igual à unidade. Subtraindo-se da unidade o valor de θ , obtém-se a proporção à qual os insumos poderiam ser reduzidos sem que o nível de produção fosse alterado.

Essa primeira versão do método DEA – doravante mencionado como a técnica DEA-C – implica, como vimos, fortes restrições no que diz respeito à forma do conjunto de produção. Em particular, essa metodologia supõe rendimentos constantes de escala. Essa hipótese pode ser facilmente relaxada modificando-se as restrições que pesam sobre o vetor das variáveis de intensidade, z . Mantendo a hipótese de livre disponibilidade dos insumos, Färe, Grosskopf e Lovell (1985 e 1994) modelaram a existência de rendimentos de escala não-crescentes (N), definindo a tecnologia como:

$$L(y | N, S) = \left\{ x : y \leq zM, zN \leq x, z \in \mathbb{R}_+^K, \sum_{k=1}^K z_k \leq 1 \right\}, \quad y \in \mathbb{R}_+^M \quad (4)$$

Nesse tipo de tecnologia, a soma das variáveis de intensidade não pode exceder a unidade, implicando que as diferentes atividades podem ser contraídas de forma viável, porém não podem ser expandidas da mesma forma. A mensuração da eficiência técnica associada a essa tecnologia pode ser computada como a solução do seguinte problema de programação linear:

$$F_i(y^k, x^k | N, S) = \theta^{DEA-N} = \min_{\theta, z} \theta \quad (5)$$

sujeito a:

$$\begin{aligned} \theta x_{kn} - \sum_{k=1}^K z_k x_{kn} &\geq 0, & n = 1, \dots, N \\ \sum_{j=1}^K z_j y_{jm} &\geq y_{km}, & m = 1, \dots, M \\ \theta, z_k &\geq 0, & k = 1, \dots, K \\ \sum_{k=1}^K z_k &\leq 1 \end{aligned}$$

Os índices de eficiência resultantes dessa mensuração – DEA-N – são também conhecidos como índices FGL. Impondo-se que a soma das variáveis de

intensidade seja exatamente igual à unidade, obtém-se a tecnologia caracterizada por retornos variáveis de escala (V):

$$L(y|V, S) = \left\{ x : y \leq zM, zN \leq x, z \in \mathbb{R}_+^K, \sum_{k=1}^K z_k = 1 \right\}, \quad y \in \mathbb{R}_+^M \quad (6)$$

Aqui as atividades não podem ser expandidas ou contraídas ilimitadamente. O conjunto das atividades viáveis resume-se ao conjunto das combinações convexas das atividades observadas. Exclui-se, ainda, a possibilidade de contração radial para a origem. Tem-se, então, retornos crescentes de escala para níveis mais baixos de produção e retornos decrescentes para níveis mais elevados. Os índices de eficiência associados a essa tecnologia – doravante mencionada como DEA-V – são obtidos através da resolução do problema de programação linear:

$$F_i(y^k, x^k | V, S) = \theta^{DEA-V} = \min_{\theta, z} \theta \quad (7)$$

sujeito a:

$$\begin{aligned} \theta x_{kn} - \sum_{k=1}^K z_k x_{kn} &\geq 0, \quad n = 1, \dots, N \\ \sum_{k=1}^K z_k y_{km} &\geq y_{km}, \quad m = 1, \dots, M \\ \sum_{k=1}^K z_k &= 1 \\ \theta, z_k &\geq 0, \quad k = 1, \dots, K \end{aligned}$$

Na literatura (Färe, Grosskopf & Lovell, 1985 e 1994) este é o modelo BCC – Banker, Charnes e Cooper (1984).

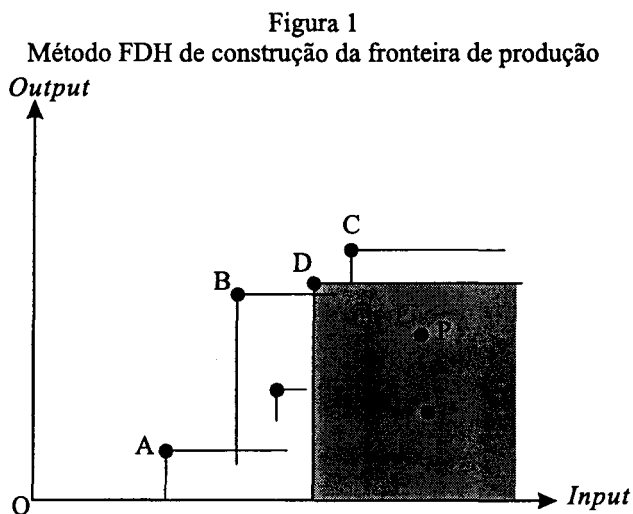
Embora representem um claro avanço em relação às técnicas paramétricas, essas diferentes abordagens baseiam-se em hipóteses demasiado restritivas no que diz respeito à estrutura do conjunto produtivo. Deprins, Simar e Tulkens (1984) propuseram um modelo fundamentado em suposições menos limitantes. Para esses autores, a fronteira do conjunto de produção é simplesmente o limite do cone de livre disponibilidade (*free disposal hull*) do conjunto de dados. A tecnologia associada a esse método (equação 8) supõe forte disponibilidade dos insumos e produtos, além de retornos variáveis de escala; porém, não mais requer a hipótese de convexidade:

$$\begin{aligned} L(y|V-FDH, S) = \left\{ x : y \leq zM, zN \leq x, z \in \mathbb{R}_+^K, \sum_{k=1}^K z_k = 1, \right. \\ \left. z_k \in [0, 1] \right\}, \quad y \in \mathbb{R}_+^M \quad (8) \end{aligned}$$

Nesse método – doravante referido como FDH – a fronteira do conjunto de produção é obtida comparando-se insumos e produtos de forma a estabelecer os pontos dominantes. Uma observação é declarada ineficiente se ela for dominada por pelo menos uma outra observação, dominação significando aqui a capacidade de produzir uma quantidade maior de produto a partir de uma quantidade inferior de insumos. Segue-se, portanto, que uma observação que não é dominada por qualquer outra é declarada FDH-eficiente. Em termos da metodologia DEA, isto equivale a acrescentar ao programa (7) a restrição de integralidade:⁵

$$z_k \in [0, 1], \quad k = 1, \dots, K \quad (9)$$

A figura 1 ilustra essa metodologia para o caso em que se considera um insumo e um produto. Nesse gráfico, os pontos *ABDC* descrevem a fronteira de produção associada ao método FDH. Por construção, toda observação ineficiente é necessariamente dominada por pelo menos uma observação. Então, a observação *D* é eficiente e domina a observação *P*: comparando-se os pontos *P* e *D*, pode-se ver que em *D* é possível produzir mais com menor quantidade de insumos. Por outro lado, a observação *A* não domina nem tampouco é dominada por qualquer outra observação. Diz-se, então que essa observação é eficiente por *default*.

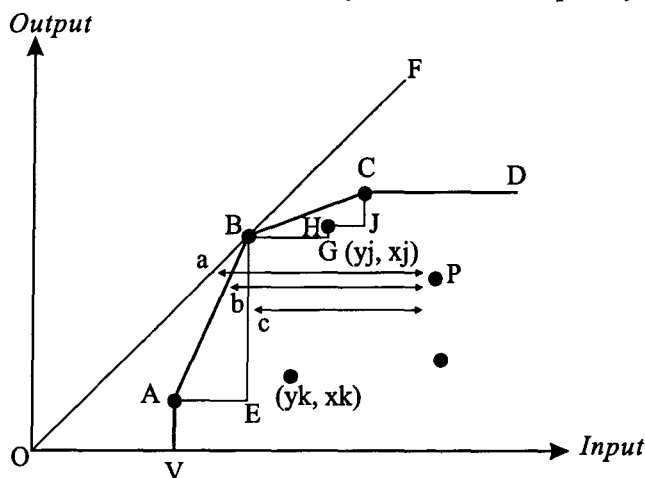


Fonte: Tulkens (1990).

⁵ Mais detalhes podem ser encontrados em Deprins, Simar & Tulkens (1984) e Tulkens (1990).

A figura 2 mostra as fronteiras DEA-C, DEA-N, DEA-V e FDH. As implicações da hipótese de convexidade são claramente percebidas. A fronteira do conjunto de produção associada ao método DEA-C é dada pelo segmento OF . Para o método DEA-N, caracterizado por retornos de escala não-crescentes, a fronteira assume a forma segmentada (*piecewise*) linear, representada pela linha $OBCD$. Excluindo-se a origem, a fronteira do conjunto de produção é dada pelo segmento $VABCD$, correspondente à variante DEA-V que incorpora retornos de escala variáveis. Finalmente, a fronteira FDH é dada pelo segmento $VAEBGHJCD$.

Figura 2
Formas alternativas de construção da fronteira de produção



Fonte: Tulkens (1990).

Pa = DEA-C; Pb = DEA-CD e DEA-V; Pc = FDH.

A partir da relação entre essas fronteiras torna-se, então, evidente que, por construção, os índices de eficiência associados a essas metodologias podem ser ordenados. Esse ordenamento se expressa na desigualdade FGL (Färe, Grosskopf & Lovell, 1994), ampliando-a para incluir o índice FDH:

$$0 < F_i(y^k, x^k | C, S) \leq F_i(y^k, x^k | N, S) \leq F_i(y^k, x^k | V, S) \\ \leq F_i(y^k, x^k | V - FDH, S) \leq 1 \quad (10)$$

Claramente, as variantes que apresentam retornos variáveis de escala envolvem melhor os dados. Assim, os índices de eficiência associados ao método FDH são mais elevados porque esse método utiliza hipóteses menos restritivas;

no outro extremo, o método DEA-C produz índices menores em virtude de sua hipótese de proporcionalidade (rendimentos constantes).

A ordenação dessas medidas de eficiência provê também informações sobre as economias de escala locais que caracterizam as diferentes atividades.⁶ De fato, comparando-se os valores dos índices de eficiência para os diferentes modelos DEA (programas 3, 5 e 7), pode-se investigar a existência de economias de escala. Definindo-se a seguinte relação:

$$S_i(y^k, x^k) = F_i(y^k, x^k | C, S) / F_i(y^k, x^k | V, S), \quad k = 1, \dots, K \quad (11)$$

a atividade (y^k, x^k) é eficiente de escala nos insumos se ela for eficiente tanto na tecnologia DEA-C quanto na tecnologia DEA-V, o que implica $S(y^k, x^k)$ igual à unidade. Em virtude de (10) e (11), tem-se que:

$$0 < S(y^k, x^k) \leq 1 \quad (12)$$

Note-se que $S(y^k, x^k) = 1$ se, e somente se, $F_i(y^k, x^k | V, S) \cdot x^k$ apresentar rendimentos constantes de escala (Färe, Grosskopf & Lovell, 1994:74). No caso em que $S(y^k, x^k) < 1$, a questão é saber se a ineficiência deve-se à existência de um nível de produção excessivamente baixo, em uma região onde imperam as economias crescentes de escala, ou se essa ineficiência resulta de um nível de produção demasiado elevado em uma região caracterizada por retornos decrescentes. O gráfico 2 ilustra bem esse ponto: a ineficiência da atividade (y^k, x^k) resulta da presença de retornos crescentes de escala, enquanto a ineficiência da atividade (y^j, x^j) advém da existência de retornos decrescentes de escala. Pode-se, portanto, inferir as economias de escala locais que caracterizam as atividades comparando-se os índices de eficiência apropriados. Portanto, se $S(y^k, x^k) < 1$ e $F_i(y^k, x^k | C, S) = F_i(y^k, x^k | N, S)$, as ineficiências resultam da existência de economias crescentes de escala; se $S(y^k, x^k) < 1$ e $F_i(y^k, x^k | C, S) < F_i(y^k, x^k | N, S)$, então a ineficiência de escala nos insumos deve-se à presença de retornos decrescentes de escala.⁷

⁶ Para mais detalhes sobre esse ponto, ver Färe, Grosskopf & Lovell (1994).

⁷ Alternativamente, podem-se avaliar as economias de escala mantendo-se $S(y^k, x^k) < 1$ e comparando-se $F_i(y^k, x^k | N, S)$ com $F_i(y^k, x^k | V, S)$. Nesse caso, se $S(y^k, x^k) < 1$ e $F_i(y^k, x^k | N, S) = F_i(y^k, x^k | V, S)$, as ineficiências resultam da existência de economias decrescentes de escala; se $S(y^k, x^k) < 1$ e $F_i(y^k, x^k | N, S) < F_i(y^k, x^k | V, S)$, então a ineficiência de escala nos insumos deve-se à presença de retornos crescentes de escala.

Note-se que a mensuração desse fenômeno, dentro da metodologia DEA, foi iniciada por Banker (1984). Partindo do modelo DEA-C (programa 3), esse autor, utilizando as restrições que pesam sobre as variáveis de intensidade, estabeleceu uma relação entre os retornos de escala e a soma dessas variáveis, $u_k = \sum_{k=1}^K z_k$. Nesse enfoque, o valor ótimo de u_k é interpretado como um indicador dos retornos de escala locais. Assim, $u_k < 1$ implica que os retornos de escala são crescentes; $u_k > 1$ implica retornos de escala decrescentes; e quando $u_k = 1$ os retornos de escala são constantes (Banker, 1984). Essa interpretação baseia-se na suposição da existência de uma solução única para u_k . Porém, como demonstrado por Chang e Guh (1988), os valores de u_k podem, em alguns casos, ser superiores e inferiores à unidade dentro de um mesmo programa. Aparentemente, esse problema não ocorre com a avaliação FGL anteriormente descrita.

3. Dados e Parâmetros

A implementação das metodologias descritas na seção 2 requer, para o cálculo do custo-eficiência, informações sobre o custo total de cada município e a quantidade de serviços públicos postos à disposição da população. Dispúnhamos, inicialmente, de dados para 4.984 municipalidades brasileiras. Após uma análise cuidadosa dessas informações, foram excluídos os municípios para os quais não foi possível coletar as informações básicas; o mesmo ocorreu com os municípios cujas informações eram claramente insatisfatórias.⁸ Além disso, para aumentar o grau de homogeneidade e facilitar as comparações, consideramos unicamente as cidades com população inferior a 100 mil habitantes. A base de dados efetivamente utilizada compõe-se, pois, de 4.157 municipalidades agregadas em cinco regiões. A escolha de 1991 como ano-base deve-se ao fato de ser esse o ano mais recente para o qual dispomos de *informações censitárias*.

⁸ A razão pela qual as informações são inexistentes ou precárias é a intensa criação de municípios no Brasil. De fato, algumas municipalidades, embora legalmente criadas, não tinham sido ainda desmembradas do município-mãe. Conseqüentemente, não declararam valores para os indicadores de produto. Ademais, para algumas cidades os dados para a despesa corrente não foram reportados por não terem sido divulgados em tempo hábil pela Secretaria do Tesouro Nacional.

3.1 Indicadores de insumo e produto para as municipalidades brasileiras

Os custos totais de cada município, computados como o valor das despesas correntes,⁹ serviram como indicador para o insumo agregado. No tocante à mensuração dos produtos, devido à impossibilidade de se quantificar diretamente a oferta de serviços públicos, um conjunto de indicadores observáveis foi utilizado como *proxy* para esses serviços. Entre os indicadores coletados, selecionamos cinco, com base em testes estatísticos.¹⁰ A descrição dos insumos e produtos utilizados encontra-se no quadro da p. 446.

As informações sobre a despesa corrente foram obtidas da Secretaria do Tesouro Nacional (STN). Os indicadores de produto 1 a 4 do quadro foram obtidos no censo das populações de 1991. Dados sobre matrículas nas escolas municipais, no primário e no secundário, referem-se ao ano de 1991 e foram obtidos junto à Coordenadoria Geral do Seec/Inep, do Ministério da Educação e dos Desportos (MEC).

3.2 Aspectos computacionais

Os índices de eficiência para as municipalidades brasileiras foram calculados através da solução dos programas lineares definidos na seção 2. Estes programas são formulados e resolvidos para cada município. A computação das diferentes variantes do programa linear – modelos DEA, apresentadas na seção 2, requer a utilização de algoritmos apropriados. Para tal foi utilizado o algoritmo CPLEX incorporado na linguagem Gams (Broke, Kendrick & Meerhaus, 1988). Este algoritmo é particularmente apropriado para tratar grandes massas de dados e resolver modelos de grande porte. Em relação ao método FDH, apesar da restrição de integralidade que caracteriza essa técnica, sua resolução não requer a utilização de algoritmos desenvolvidos para resolver problemas de programação inteira. Na verdade, sequer é necessário o uso de técnicas de programação linear. Um sistema de comparações exaustivas, entre as observações, permite calcular os níveis de eficiência e as outras informações relevantes.

⁹ Eliminamos as despesas com investimento, pelo fato de serem menos confiáveis.

¹⁰ Ver Sousa, Ramos & Silva (1997) para mais detalhes.

Descrição dos insumos e produtos

Insumo	1. Despesa corrente	DSP92_C	Gastos correntes dos municípios: despesa corrente agregada, correspondente ao montante de gastos correntes, conforme definido pela Lei nº 4.320/64, efetuados por cada município brasileiro no exercício financeiro de 1992 e expressos em cruzeiros de 1992.
Indicadores de serviços públicos (produtos)	1. População total	POP91_T	População residente entre 31-8 e 1-9-1991.
	2. Domicílios com água	DOM91_A	Número de domicílios servidos por rede geral, com canalização interna ou não.
	3. Domicílios com coleta de lixo	DOM91_L	Número de domicílios com coleta de lixo: refere-se ao número de domicílios com coleta de lixo feita direta ou indiretamente por serviço, empresa pública ou particular. No primeiro caso, o lixo é coletado no domicílio, e no segundo o lixo é depositado em caçambas, tanques ou depósitos e depois coletado.
	4. (1/Número de analfabetos)	(1/ANALF91)	Inverso do número de analfabetos: refere-se às pessoas com cinco anos ou mais, incapazes de ler e escrever um bilhete simples no idioma que conhecessem.
	5. Alunos matriculados	EDUC	Número de alunos matriculados no pré-primário, primeiro e segundo graus.

Essa variável representou, em 1992, 75% das despesas totais do município.

4. Mensuração dos Níveis de Eficiência: Discussão dos Resultados para os Casos do Nordeste e do Sudeste Brasileiros

Com o intuito de identificar o grau de eficiência das municipalidades brasileiras, as metodologias discutidas na seção 2 foram aplicadas aos municípios das regiões Nordeste e Sudeste. As tabelas 1 a 4 sintetizam os resultados obtidos.

As tabelas 1 e 2 categorizam 1.429 municipalidades da região Nordeste e 1.334 municípios da região Sudeste, agregados em classes segundo a população, em eficientes e ineficientes, utilizando-se três metodologias distintas: DEA-C, DEA-V e FDH.¹¹

Tabela 1
Caracterização das municipalidades segundo as diferentes metodologias:
região Nordeste

Classes de população	Total de municípios	Municípios eficientes						Municípios ineficientes					
		DEA-C		DEA-V		FDH		DEA-C		DEA-V		FDH	
		Nº	%	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
0-4.999	125	1	0,8	1	0,8	63	50,4	124	99,2	124	99,2	62	49,6
5.000-9.999	307	0	0,0	0	0,0	158	51,5	307	100,0	307	100,0	149	48,5
10.000-14.999	319	2	0,6	3	0,9	146	45,8	317	99,4	316	99,1	173	54,2
15.000-19.999	202	0	0,0	0	0,0	94	46,5	202	100,0	202	100,0	108	53,5
20.000-24.999	147	0	0,0	0	0,0	68	46,3	147	100,0	147	100,0	79	53,7
25.000-29.999	94	4	4,3	4	4,3	43	45,7	90	95,7	90	95,6	51	54,3
30.000-34.999	58	1	1,7	1	1,7	29	50,0	57	98,3	57	98,3	29	50,0
>35.000	177	7	4,0	15	8,5	125	70,6	170	96,0	162	91,5	52	29,4
Total	1.429	15	1,0	24	1,7	726	50,8	1.414	99,0	1.405	98,3	703	49,2

Tabela 2
Caracterização das municipalidades segundo as diferentes metodologias:
região Sudeste

Classes de população	Total de municípios	Municípios eficientes						Municípios ineficientes					
		DEA-C		DEA-V		FDH		DEA-C		DEA-V		FDH	
		Nº	%	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
0-4.999	331	1	0,3	2	0,6	130	39,2	330	99,7	329	99,4	201	60,7
5.000-9.999	319	0	0,0	3	0,9	161	50,5	319	100,0	316	99,1	158	49,5
10.000-14.999	208	1	0,5	2	1,0	113	54,3	207	99,5	206	99,0	95	45,7
15.000-19.999	132	1	0,8	1	0,8	84	63,6	131	99,2	131	99,2	48	36,4
20.000-24.999	76	1	1,3	2	2,6	51	67,1	75	98,7	74	97,4	25	32,9
25.000-29.999	48	1	2,1	2	4,2	24	50,0	47	97,9	46	95,8	24	50,0
30.000-34.999	45	1	2,2	2	4,4	24	53,3	44	97,7	43	95,6	21	46,7
>35.000	175	5	2,9	19	10,9	112	64,0	170	97,1	156	89,1	63	36,0
Total	1.334	11	0,8	33	2,5	699	52,4	1.323	99,2	1.307	97,5	635	47,6

Com base nas informações contidas nessas tabelas, nota-se que os resultados associados à metodologia FDH divergem sensivelmente daqueles encontrados através das duas formas utilizadas do método DEA (DEA-C e DEA-V). Enquanto a metodologia FDH atesta a existência de um grande número de municipalidades eficientes, a variante DEA-C mostra que somente 0,8 e 0,5% das municipalidades são eficientes, respectivamente, nas regiões Nordeste e

¹¹ Os resultados para a variante DEA-N encontram-se em Sousa, Ramos & Silva (1997).

Sudeste; essa proporção aumenta muito pouco quando a variante DEA-V é utilizada. Várias razões podem ser invocadas para explicar tamanha divergência.

Em primeiro lugar, a fronteira de custo-eficiência associada à metodologia FDH baseia-se em uma noção mais restritiva de *dominância* quando comparada com as fronteiras calculadas pelo método DEA nas duas versões apresentadas. Uma municipalidade é *FDH-dominada* somente quando *todos* os seus indicadores são inferiores àqueles de uma municipalidade eficiente, efetivamente observada, com a qual ela se compara, e sua despesa corrente é igual ou superior às despesas da municipalidade dominante. Por outro lado, uma municipalidade é *DEA-dominada* por uma observação *fictícia*, definida como sendo uma combinação linear (convexa para o caso do DEA-V) de um conjunto de municipalidades eficientes. Portanto, fronteiras do tipo FDH dependem de comparações entre municípios existentes, enquanto as fronteiras do tipo DEA “fabricam” suas próprias possibilidades de comparação. Quando, por falta de informações, é impossível estabelecer comparações, o método FDH declara que a municipalidade em análise é *eficiente por default*. Este procedimento conduz a um aumento do número de municípios ditos eficientes. Essa questão será, posteriormente, discutida com mais detalhes.

Além disso, a hipótese de convexidade incorporada nas fronteiras do tipo DEA restringe, desnecessariamente, o tipo de tecnologia através da qual os serviços públicos podem ser produzidos. Em particular, a variante DEA-C, que ajusta uma fronteira de custo-eficiência caracterizada por retornos constantes de escala, ao ignorar a presença de não-convexidades locais, sistematicamente subestima o grau de eficiência das municipalidades e, portanto, pode não representar um ajustamento apropriado da fronteira produtiva. Considerando que os dados referentes aos municípios das regiões analisadas sugerem a presença de economias de escala para a maioria dos pequenos municípios, essa limitação pode reduzir significativamente a credibilidade da fronteira calculada através desse tipo de metodologia. Este ponto será discutido com mais detalhes na seção 5.

Finalmente, como vimos anteriormente, as fronteiras engendradas pelos métodos FDH, DEA-V e DEA-C são imbricadas umas nas outras, com a fronteira FDH situando-se mais perto das observações e a fronteira DEA-C

encontrando-se a uma maior distância dos dados (ver equação 10). Portanto, a metodologia FDH envolve melhor os dados e esse melhor envolvimento se traduz em um maior número de municipalidades eficientes.

4.1 Eficiência por *default* e *outliers*

No enfoque que ora discutimos, dois pontos merecem destaque: eficiência por *default* e *outliers*.

Eficiência por *default*

Este problema, embora também presente na metodologia DEA, assume particular importância no método FDH, podendo, inclusive, comprometer o poder discriminatório dessa técnica. Isto porque, como dito anteriormente, na ausência de um número suficiente de municípios similares com os quais uma determinada cidade possa ser cotejada, esse município, ao invés de estabelecer uma relação dominante-dominada, é declarado *eficiente por default*. Esse tipo de eficiência não resulta necessariamente de superioridade atestada em relação a outras cidades, mas decorre da insuficiência de informações que permitam estabelecer comparações relevantes. Além disso, pela forma como a técnica FDH é construída, o conceito de *eficiência por default* aplica-se, ainda, à municipalidade que apresenta o menor nível de despesa, assim como àquelas que apresentam o valor mais elevado para pelo menos um dos indicadores de serviços públicos. Essa forma extrema do *sparsity bias* certamente contribui para que o método FDH apresente uma superestimação significativa do número de municipalidades eficientes e, portanto, constitui uma séria limitação dessa abordagem.

As tabelas 3 e 4 apresentam as informações referentes a esse problema. Ressalte-se que o fenômeno da eficiência por *default* relaciona-se com o grau de homogeneidade das informações. Ela depende não somente do número de comparações potenciais, mas também da dispersão dessas mesmas informações. Assim, por exemplo, entre as municipalidades maiores da região Sudeste, onde os municípios são mais homogêneos, a eficiência por impossibilidade de comparação é significativamente inferior àquela apresentada pelas cidades de mesmo porte da região Nordeste. Efetivamente, entre as 175 cidades do Sudeste eficientes, com mais de 35 mil habitantes, apenas 40% foram declaradas eficientes por *default*, contra 71% das 177 cidades nordestinas do mesmo porte e também eficientes.

Tabela 3
Caracterização das municipalidades eficientes (método FDH):
região Nordeste

Classes de população	Total de municípios	Municípios eficientes	Municípios eficientes e dominantes		Municípios eficientes por <i>default</i>	
			Nº	%	Nº	%
0-4.999	125	63	23	36,5	40	63,5
5.000-9.999	307	158	104	65,8	54	34,2
10.000-14.999	319	146	119	81,5	27	18,5
15.000-19.999	202	94	69	73,4	25	26,6
20.000-24.999	147	68	51	75,0	17	25,0
25.000-29.999	94	43	28	65,1	15	34,9
30.000-34.999	58	29	12	41,4	17	58,6
>35.000	177	125	36	28,8	89	71,2
Total	1.429	726	442	60,9	284	39,1

Tabela 4
Caracterização das municipalidades eficientes (método FDH):
região Sudeste

Classes de população	Total de municípios	Municípios eficientes	Municípios eficientes e dominantes		Municípios eficientes por <i>default</i>	
			Nº	%	Nº	%
0-4.999	331	130	87	66,9	43	33,1
5.000-9.999	319	161	126	78,3	35	21,7
10.000-14.999	208	113	74	65,5	39	34,5
15.000-19.999	132	84	51	60,7	33	39,3
20.000-24.999	76	51	38	74,5	13	25,5
25.000-29.999	48	24	19	79,2	5	20,8
30.000-34.999	45	24	14	58,3	10	41,7
>35.000	175	112	67	59,8	45	40,2
Total	1.334	699	476	68,1	223	31,9

Por fim, a metodologia FDH é particularmente apropriada para identificar os casos mais óbvios de ineficiências, tendo em vista que essa técnica é bastante assertiva no que diz respeito à detecção de ineficiências. De fato, para cada municipalidade declarada *FDH-ineficiente*, existe *pelo menos uma* municipalidade na amostra que apresenta *performance* claramente superior à municipalidade dominada. Do ponto de vista gerencial, essa informação é extremamente relevante, pois constitui evidência factual de que é possível, com os mesmos recursos, obter melhores resultados. Note-se que, como se trata de uma noção relativa de eficiência, pode ocorrer que as municipalidades domi-

nantes possuam características outras que sejam substancialmente diferentes daquelas que são dominadas. Nesse caso, aquilo que é mensurado como ineficiência pode simplesmente refletir a ação de tais características. Porém, na ausência de tais fatores, o método FDH revela, de forma convincente e simples, as ineficiências técnicas que caracterizam o processo produtivo.

Outliers

Por definição, as fronteiras de eficiência não-paramétricas são definidas pelos valores extremos do espaço dimensional dos insumos e produtos. Conseqüentemente, a presença de *outliers*, observações atípicas que divergem significativamente do restante dos dados, pode influenciar consideravelmente a mensuração dos índices de eficiência. Assim, no método FDH, quando se consideram os *municípios eficientes*, pode-se, por exemplo, detectar os municípios que, além de dominarem vários outros, aparecem sistematicamente como sendo a municipalidade *mais dominante* (aquela que possui a menor despesa) para outras cidades e, portanto, influenciam significativamente a mensuração dos níveis de eficiência. Essas cidades, de alguma forma, detêm a “melhor tecnologia” para a produção dos serviços públicos, em relação aos seus “pares”.¹² Como elas definem a fronteira tecnológica, a remoção de uma dessas municipalidades implica redução das exigências para pertencer à fronteira e, assim, poder fazer com que municípios antes declarados ineficientes passem a ser eficientes.

Note-se que esse problema é ainda mais acentuado no que diz respeito às metodologias DEA e suas variantes DEA-C e DEA-V. Nesses modelos, um número surpreendentemente pequeno de municipalidades contribui para determinar a fronteira de produção. Essas observações anômalas claramente deslocam os limites do conjunto produtivo e se refletem nos índices de eficiência, podendo, inclusive, contribuir para torná-los artificialmente baixos. Deve-se, portanto, verificar se tais desvios não resultam de erros de avaliação. Porém, uma vez que se garanta que os dados estão corretamente mensurados, esse tipo de observação pode revelar informações extremamente úteis que poderiam, de outra forma, ser ignoradas. Tais informações podem, portanto, servir para aumentar a eficácia das políticas públicas mediante a disseminação de práticas de gestão de recursos mais eficientes.

¹² Em metodologias do tipo DEA essa “tecnologia” é dada pelos multiplicadores z . Ver, a esse respeito, a seção 2.

5. Economias de Escala e Eficiência: Mensurando os Custos Associados à Política Atual de Descentralização Municipal no Brasil

A análise dos índices de eficiência para os diferentes municípios levanta uma questão importante que merece ser discutida: a relação entre o tamanho do município e o nível de eficiência. As diferentes metodologias parecem mostrar que municípios menores tendem a ser mais *ineficientes*. De fato, tanto nos métodos do tipo DEA quanto na metodologia FDH, o ajustamento da fronteira melhora sensivelmente com o tamanho das municipalidades (tabelas 1 e 2). Isto resulta da presença de economias crescentes de escala, em nível local. As tabelas 5 e 6, que apresentam os testes de Färe, Grosskopf e Lovell (FGL) e de Banker para avaliar as economias de escala locais (ver item 2.1), ilustram bem esse ponto. Observa-se aí que, tanto na região Nordeste quanto na Sudeste, para a imensa maioria dos municípios com população inferior a 15 mil habitantes, os baixos níveis de eficiência se explicam pela presença de economias crescentes de escala. Este fenômeno é particularmente importante para a região Sudeste, onde 64,3% das cidades encontram-se nessa faixa populacional. Nessa região, os rendimentos decrescentes somente começam a ser relevantes para cidades com população igual ou superior a 25 mil habitantes. Esse resultado é confirmado tanto pelo teste FGL quanto pelo teste de Banker que, nesse estudo, apresentam resultados praticamente idênticos.¹³

A existência de retornos crescentes de escala para municípios menores explica também por que o envolvimento dos dados pelo método DEA-C é precário. De fato, apenas sete municipalidades na região Nordeste e seis na região Sudeste, entre aquelas com até 30 mil habitantes, atingem a fronteira de custo-eficiência, caracterizada por retornos constantes de escala, postulada por esse método (tabelas 5 e 6).

¹³ No que se refere à mensuração das economias de escala nos modelos DEA, embora Banker, Chang e Cooper (1996) tenham mostrado que a abordagem de Banker (1984) e o enfoque de Färe, Grosskopf e Lovell (1985) produzem resultados equivalentes, pequenas diferenças persistem entre esses métodos, deixando alguns pontos em aberto nessa discussão, em particular no que tange à computação das economias de escala. Ver, a esse respeito, os trabalhos de Banker, Chang & Cooper (1996), Banker & Thrall (1992), Färe (1997), Golany & Yu (1997), entre outros.

Tabela 5
Retornos de escala e eficiência econômica (métodos FGL e de Banker):
região Nordeste

Classes de população	Nº municípios	Método FGL						Método de Banker					
		Retornos constantes		Retornos decrescentes		Retornos crescentes		Retornos constantes		Retornos decrescentes		Retornos crescentes	
		Nº	%	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
0-4.999	125	1	0,8	4	3,2	120	96,0	1	0,8	5	4,0	119	95,2
5.000-9.999	307	0	0,0	9	2,9	298	97,1	1	0,3	9	2,9	297	96,7
10.000-14.999	319	2	0,6	41	12,8	276	86,5	4	1,3	42	13,7	273	85,6
15.000-19.999	202	1	0,5	90	44,6	111	55,0	0	0,0	90	44,6	112	55,4
20.000-24.999	147	3	2,0	86	58,5	58	39,5	1	0,7	87	59,2	59	40,1
25.000-29.999	94	5	5,3	68	72,3	21	22,3	4	4,3	69	73,4	21	22,3
30.000-34.999	58	1	1,7	43	74,1	14	24,1	1	1,7	42	72,4	15	25,9
>35.000	177	7	4,0	156	88,1	14	7,9	7	4,0	156	88,1	14	7,9
Total	1.429	20	1,4	497	34,8	912	63,8	19	1,3	500	34,8	910	63,4

Tabela 6
Retornos de escala e eficiência econômica (métodos FGL e de Banker):
região Sudeste

Classes de população	Nº municípios	Método FGL						Método de Banker					
		Retornos constantes		Retornos decrescentes		Retornos crescentes		Retornos constantes		Retornos decrescentes		Retornos crescentes	
		Nº	%	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
0-4.999	331	1	0,3	0	0,0	330	99,7	1	0,3	0	0,0	330	99,7
5.000-9.999	319	0	0,0	0	0,0	319	100	0	0,0	0	0,0	319	100
10.000-14.999	208	1	0,5	0	0,0	207	99,5	1	0,5	0	0,0	207	99,5
15.000-19.999	132	1	0,8	4	3,0	127	96,2	1	0,8	4	3,0	127	96,2
20.000-24.999	76	2	2,6	0	0,0	74	97,4	1	1,3	1	1,3	74	97,4
25.000-29.999	48	1	2,1	2	4,2	45	93,8	1	2,1	2	4,2	45	93,8
30.000-34.999	45	1	2,2	8	17,8	36	80,0	1	2,2	8	17,8	36	80,0
>35.000	175	18	10,3	79	45,1	78	44,6	8	4,6	79	45,1	88	44,6
Total	1.334	25	1,9	93	7,0	1.216	91,2	14	1,9	94	7,0	1.226	91,2

Isto ocorre porque, aparentemente, municípios excessivamente pequenos não exploram as economias de escala que caracterizam muitos dos serviços públicos e, portanto, não utilizam de maneira ótima os recursos disponíveis. Este argumento encontra apoio em estudos empíricos recentes. Por exemplo, existe ampla evidência de que, para os serviços educacionais, em virtude de os custos fixos serem consideráveis, os custos operacionais decrescem fortemente com o número de matrículas. Escolas maiores tendem a ser mais eficientes (*cost-efficient*) porque os custos fixos são diluídos entre um número mais elevado de alunos. Cidades muito pequenas, cujas escolas atendem a um pequeno número de alunos, podem, então, estar trabalhando em uma faixa onde os custos médios são demasiado elevados. Caso a cidade fosse maior, seria possível

aumentar o número de matrículas, sem perda significativa de qualidade, e ainda reduzir os custos por aluno.¹⁴

Um raciocínio análogo pode ser usado quando da avaliação de outros tipos de serviços públicos. É o caso dos serviços administrativos. Aqui, também, a existência de custos fixos importantes faz com que, em pequenas municipalidades, o custo por habitante desses serviços seja provavelmente maior que aqueles incorridos pelas cidades de maior porte. Em trabalho recente, Gomes e MacDowell (1997) corroboram esse ponto. Examinando o gasto com pessoal, por habitante, nos municípios brasileiros, esses autores mostram que esses gastos são significativamente maiores em municípios menores. Para os municípios de até 50 mil habitantes, eles encontraram uma curva de custo médio para as despesas de pessoal claramente decrescente.

Concluindo, em virtude da existência de economias crescentes de escala, um aumento proporcional em todos os indicadores de serviços públicos poderia ser financiado através de um aumento proporcionalmente inferior nas despesas correntes. Por conseguinte, poder-se-ia aumentar o tamanho típico do município e prover os serviços públicos necessários sem incorrer em um aumento equivalente nos gastos públicos. Esses resultados se mantêm para as outras regiões da Federação. De fato, um rápido exame dessa questão no âmbito das outras regiões indica a presença de economias de escala para municípios cuja população é inferior a 50 mil habitantes (tabelas A7-A9, do apêndice). Por conseguinte, a proliferação de pequenos municípios decorrente da intensa criação de municipalidades desde 1985 reduz o nível de eficiência da economia e implica desperdício considerável de recursos.

As tabelas 7 e 8 apresentam estimativas para essas perdas. Constata-se que, em todas as metodologias, o desperdício de recursos é substancial em municipalidades com até 10 mil habitantes. Dependendo do método utilizado e da região considerada, esse desperdício varia de 12 a 70% do total dos recursos disponíveis para as despesas correntes. Esse resultado lança sérias dúvidas sobre a viabilidade técnica dos pequenos municípios. Embora seja necessário investigar mais a fundo esse problema, muito provavelmente a atual concentração de cidades em classes com população de até 10 mil habitantes representa um custo adicional não desprezível para o Brasil e indica que o tamanho típico dessas cidades está muito abaixo do exigido para maximizar o nível de serviços públicos postos à disposição da população. Ademais, o trabalho de Gomes e

¹⁴ Com respeito a esse assunto, ver, entre outros, Tan & Mingat (1992) e Sousa (1997).

MacDowell questiona também a viabilidade fiscal/econômica desses municípios. Segue-se, então, que uma utilização mais racional dos recursos públicos deveria considerar não o desmembramento, mas sim o reagrupamento municipal.

Tabela 7
Nordeste: excesso de dispêndio associado às diferentes metodologias

Classes de população	Municípios		Despesa (Cr\$ mil)	Excesso de dispêndio					
				DEA-C		DEA-V		FDH	
	Nº	%		Nº	%	Nº	%	Nº	%
0-4.999	125	8,8	225.678	157.511	69,8	137.996	61,1	28.366	12,6
5.000-9.999	307	21,5	728.300	479.598	65,8	431.907	59,3	95.018	13,0
10.000-14.999	319	22,3	1.041.961	661.652	63,5	635.341	61,0	169.092	16,2
15.000-19.999	202	14,1	842.945	505.205	59,9	495.666	58,8	130.864	15,5
20.000-24.999	147	10,3	835.362	495.997	59,4	489.922	58,6	162.637	19,5
25.000-29.999	94	6,6	512.888	247.085	48,2	242.336	47,2	82.514	16,1
30.000-34.999	58	4,1	384.962	196.925	51,2	192.991	50,1	47.331	12,3
>35.000	177	12,4	1.812.670	758.906	41,9	699.076	38,6	102.240	5,6
Total	1.429	100,0	6.384.766	3.502.880	54,9	3.325.236	52,1	818.061	12,8

Tabela 8
Sudeste: excesso de dispêndio associado às diferentes metodologias

Classes de população	Municípios		Despesa (Cr\$ mil)	Excesso de dispêndio					
				DEA-C		DEA-V		FDH	
	Nº	%		Nº	%	Nº	%	Nº	%
0-4.999	329	24,7	939.867	608.736	64,7	422.484	45,0	138.320	14,7
5.000-9.999	314	23,5	1.149.201	674.229	58,7	547.385	47,6	167.805	14,6
10.000-14.999	204	15,3	1.226.668	708.270	57,7	665.670	54,2	128.781	10,5
15.000-19.999	131	9,8	1.096.382	618.476	56,4	598.754	54,6	123.778	11,3
20.000-24.999	76	5,7	743.366	380.202	51,1	371.894	50,0	54.467	7,3
25.000-29.999	48	3,6	616.776	319.780	51,8	316.256	51,3	90.761	14,7
30.000-34.999	45	3,4	816.437	476.592	58,4	471.799	57,8	204.499	25,0
>35.000	187	14,0	5.402.742	2.799.720	51,8	2.247.301	41,6	624.721	11,6
Total	1.334	100,0	11.991.439	6.586.006	100,0	5.641.544,40	100,0	1.533.134	100,0

6. À Guisa de Conclusão

Este artigo preenche uma lacuna importante no que diz respeito à avaliação da *performance* dos municípios brasileiros. Trata-se de um estudo pioneiro que permite elaborar um instrumental para analisar o desempenho dos governos municipais. Esse instrumental apóia-se solidamente na teoria econômica e é suficientemente flexível para ser operacionalizado e usado como forma de avaliação sistemática e permanente pelos governos locais. Os resultados obtidos poderão futuramente servir para detectar as ineficiências na utilização dos recursos públicos escassos e orientar a elaboração de políticas apropriadas de gastos públicos.

No que diz respeito à mensuração dos níveis de eficiência, é importante enfatizar a natureza exploratória deste trabalho. Os indicadores calculados devem ser usados com cautela levando-se em conta as limitações da base de dados. Ineficiências resultam não somente de incompetência administrativa ou da inexistência de incentivos adequados, mas podem também refletir erros de mensuração tanto no que diz respeito aos indicadores de insumos quanto no tocante aos indicadores de produto. Isto porque uma limitação importante dos modelos não-paramétricos de cálculo de eficiência técnica reside no fato de serem determinísticos. Em consequência, os níveis de eficiência computados dependem, crucialmente, da qualidade dos dados. Variáveis omitidas e outros tipos de discrepâncias estatísticas comprometem a representatividade dos coeficientes e, no extremo, podem invalidar os resultados obtidos. É preciso, pois, garantir a qualidade da base de dados sobre a qual o modelo se assenta, assegurando-se que as variáveis relevantes sejam corretamente identificadas e sua mensuração feita de forma acurada. Conclui-se, pois, que antes que o diagnóstico de ineficiência seja atribuído a uma determinada municipalidade, é importante avaliar, cuidadosamente, o caso, inclusive recorrendo a informações complementares para confirmar tal diagnóstico.

Finalmente, no que concerne às economias de escala, nossos resultados sugerem que a política atual de descentralização, no Brasil, não promove o uso eficiente dos recursos públicos. A proliferação de pequenas municipalidades resultante dessa política concorre para o aumento dos custos médios dos serviços municipais. Isto porque, em virtude do pequeno tamanho, esses municípios não podem se beneficiar das economias de escala inerentes à produção de determinados serviços públicos; o desperdício de recursos daí decorrente reflete-se nos elevados níveis de excesso de dispêndio que caracterizam as cidades menores. Portanto, para reduzir as perdas de eficiência, deve-se reavaliar a política de descentralização municipal, no sentido de controlar e evitar o excessivo desmembramento de municípios.

Referências Bibliográficas

- Aigner, D. J. & Chu, S. F. On estimating the industry production function. *American Economic Review*, 58:826-39, 1968.
- Banker, R. D. Estimating most productive scale size using data envelopment analysis. *European Journal of Operational Research*, 17:35-44, 1984.

_____; Chang, H. & Cooper, W. W. Equivalence and implementation of alternative methods for determining returns to scale in data envelopment analysis. *European Journal of Operational Research*, 89:473-81, 1996.

_____; Charnes, H. & Cooper, W. W. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30:1.078-92, 1984.

_____. & Thrall, R. M. Estimation of returns to scale using data envelopment analysis. *European Journal of Operational Research*, 62:74-84, 1992.

Bauer, P. W. A survey of recent econometric development in frontier estimation. *Journal of Econometrics*, 46:39-56, 1990.

Bergstrom, T. C. & Goodman, R. P. Private demand for public goods. *American Economic Review*, 63:280-96, 1973.

Brasil. MEC/Sediae/Inep. *Estatísticas da educação no Brasil*. Brasília, 1996.

Brasil. MF/STN. *Finanças do Brasil: receita e despesa dos municípios, ano de 1995*. Brasília, DF, 1996. v. 41.

Broke, A.; Kendrick, D. & Meeraus, A. *Gams: a user guide*. USA, The Scientific Press, 1988.

Chang, K. P. & Guh, Y. Y. Linear production functions and the data envelopment analysis. *European Journal of Operational Research*, 52:215-23, 1988.

Charnes, A.; Cooper, W. W. & Rhodes, E. Measuring efficiency of decision making units. *European Journal of Operational Research*, 1:429-44, 1978.

_____. Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through. *Management Science*, 27:668-97, 1981.

Deprins, D.; Simar, L. & Tulkens, H. Measuring labor efficiency in post offices. In: Marchand, M.; Pestieau, P. & Tulkens, H. (eds.). *The performance of public enterprises: concepts and measurements*. Amsterdam, North Holland, 1984.

Färe, R. *Determining returns to scale in DEA*. Southern Illinois University, 1997. (Discussion Paper Series, 97-10.)

_____. Grosskopf, S. & Lovell, C. K. *The measurement of efficiency of production*. Boston-Dordrech, Kluwer-Nijhoff Publishing, 1985.

_____. *Production frontiers*. Cambridge, Cambridge University Press, 1994.

Farrel, M. J. The measurement of productive efficiency. *Journal of the Royal Statistical Society*, 125:252-67, 1957.

Førsund, F.; Lovell, C. K. & Schmidt, P. A survey of frontier production functions and of their relationship to efficiency measurement. *Journal of Econometrics*, 13:5-25, 1980.

Golany, B. & Yu, G. Estimating returns to scale in DEA. *European Journal of Operational Research*, 103:28-37, 1997.

Gomes, G. Maia & MacDowell, C. Os elos frágeis da descentralização: observações sobre as finanças dos municípios brasileiros, 1995. In: Encontro Nacional de Economia, 25. *Anais*. Recife, 1997. p. 645-60.

IBGE. *Censo demográfico de 1991*. Rio de Janeiro, 1994. v. 1, Brasil.

MF/STN. *Finanças do Brasil: receita e despesa dos municípios, ano de 1995*. Brasília, DF, 1996. v. 41.

Seiford, L. M. & Thrall, R. M. Recent developments in DEA: the math programming approach to frontier analysis. *Journal of Econometrics*, 46:7-38, 1990.

Sousa, M. C. Sampaio de. Efficiency and equity aspects of social spending in selected countries of Latin America and East Asia: a comparative approach. In: Encontro Nacional de Economia, 25. *Anais*. Recife, 1997. p. 1.328-47.

_____; Ramos, F. S. & Silva, M. C. Eficiência dos gastos públicos no Brasil: uma análise em nível de municípios. Brasília, Ipea, 1997.

Tan, J. P. & Mingat, A. *Education in Asia: a comparative study of cost and financing*. World Bank Regional and Sectoral Studies, The World Bank, 1992.

Tulkens, H. *The measurement of productive efficiency by FDH frontiers*. Louvain-la-Neuve, Université Catholique de Louvain, 1990.

Vanden, Eeckaut; Tulkens, H. & Jamar, M. A. *A study of cost-efficiency and returns of scale for 235 municipalities in Belgium*. Louvain-la-Neuve, Université Catholique de Louvain, 1991. (Core Discussion Paper, 9.158.)

Apêndice A

Resultados principais para as regiões Norte, Centro-Oeste e Sul

Tabela A1
Caracterização das municipalidades segundo as diferentes metodologias:
região Centro-Oeste

Classes de população	Total de municípios	Municípios eficientes						Municípios ineficientes					
		DEA-C		DEA-V		FDH		DEA-C		DEA-V		FDH	
		Nº	%	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
0-4.999	97	1	1,0	3	3,1	59	60,8	96	98,9	94	96,9	38	39,2
5.000-9.999	109	1	0,9	2	1,8	67	61,5	108	99,1	107	98,2	42	38,5
10.000-14.999	59	0	0,0	0	0,0	41	69,5	59	100	59	100	18	30,5
15.000-19.999	31	0	0,0	0	0,0	26	83,9	31	100	31	100	5	16,1
20.000-24.999	25	0	0,0	0	0,0	20	80,0	25	100	25	100	5	20,0
25.000-29.999	10	0	0,0	0	0,0	9	90,0	10	100	10	100	1	10,0
30.000-34.999	13	0	0,0	0	0,0	10	76,9	13	100	13	100	3	23,1
>35.000	24	2	8,3	10	41,7	24	100	22	91,7	14	58,3	0	0,0
Total	368	4	1,1	15	4,1	256	69,6	364	98,9	353	95,9	112	30,4

Tabela A2
Caracterização das municipalidades segundo as diferentes metodologias:
região Norte

Classes de população	Total de municípios	Municípios eficientes						Municípios ineficientes					
		DEA-C		DEA-V		FDH		DEA-C		DEA-V		FDH	
		Nº	%	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
0-4.999	13	2	15,4	3	23,1	12	92,3	11	84,6	10	76,9	1	7,7
5.000-9.999	29	1	3,4	1	3,4	19	61,6	28	96,5	28	96,5	10	34,5
10.000-14.999	27	0	0,0	0	0,0	19	70,4	27	100	27	100	8	29,6
15.000-19.999	32	2	6,3	2	6,2	17	53,1	30	93,7	30	93,7	15	46,7
20.000-24.999	20	0	0,0	1	5,0	16	80,0	20	100	19	95,0	4	20,0
25.000-29.999	11	2	18,2	2	18,2	8	72,4	9	81,8	9	81,8	3	27,3
30.000-34.999	13	2	15,4	2	15,4	12	92,3	11	84,6	11	84,6	1	7,7
>35.000	46	2	4,4	13	28,3	39	84,8	44	95,6	33	71,7	7	15,2
Total	191	11	5,8	24	12,6	142	74,3	180	94,2	167	87,4	49	25,6

Tabela A3
Caracterização das municipalidades segundo as diferentes metodologias:
região Sul

Classes de população	Total de municípios	Municípios eficientes						Municípios ineficientes					
		DEA-C		DEA-V		FDH		DEA-C		DEA-V		FDH	
		Nº	%	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
0-4.999	153	1	0,6	3	1,9	101	66,0	152	99,3	150	98,0	52	34,0
5.000-9.999	252	0	0,0	1	0,4	153	60,7	252	100	251	99,6	99	39,3
10.000-14.999	150	0	0,0	0	0,0	91	60,7	150	100	150	100	59	39,3
15.000-19.999	92	0	0,0	0	0,0	55	59,8	92	100	92	100	37	40,2
20.000-24.999	46	1	2,2	1	2,2	30	65,2	45	97,8	45	97,8	16	34,7
25.000-29.999	38	2	5,2	2	5,3	28	73,7	36	94,7	36	94,7	10	36,3
30.000-34.999	12	0	0,0	0	0,0	7	58,3	12	100	12	100	5	41,7
>35.000	92	8	8,9	8	8,9	68	73,9	84	91,3	84	91,3	24	26,1
Total	835	12	1,4	15	1,8	533	63,8	823	98,6	820	98,2	302	36,2

Tabela A4
Caracterização das municipalidades eficientes (método FDH):
região Norte

Classes de população	Total de municípios	Municípios eficientes	Municípios eficientes e dominantes		Municípios eficientes por <i>default</i>	
			Nº	%	Nº	%
0-4.999	13	12	1	8,3	11	91,7
5.000-9.999	29	19	3	15,8	16	84,2
10.000-14.999	27	19	3	15,8	16	84,2
15.000-19.999	32	17	6	35,3	11	64,7
20.000-24.999	20	16	8	50,0	8	50,0
25.000-29.999	11	8	3	37,5	5	62,5
30.000-34.999	13	12	3	25,0	9	75,0
>35.000	46	39	9	23,1	30	76,9
Total	191	142	36	25,4	106	74,6

Tabela A5
Caracterização das municipalidades eficientes (método FDH):
região Centro-Oeste

Classes de população	Total de municípios	Municípios eficientes	Municípios eficientes e dominantes		Municípios eficientes por <i>default</i>	
			Nº	%	Nº	%
0-4.999	97	59	24	40,7	35	59,3
5.000-9.999	109	67	32	47,8	35	52,2
10.000-14.999	59	41	18	43,9	23	56,1
15.000-19.999	31	26	16	61,5	10	38,5
20.000-24.999	25	20	7	35,0	13	65,0
25.000-29.999	10	9	2	22,2	7	77,8
30.000-34.999	13	10	1	10,0	9	90,0
>35.000	24	24	2	8,3	23	91,7
Total	368	256	102	39,8	154	60,2

Tabela A6
Caracterização das municipalidades eficientes (método FDH):
região Sul

Classes de população	Total de municípios	Municípios eficientes	Municípios eficientes e dominantes		Municípios eficientes por <i>default</i>	
			Nº	%	Nº	%
0-4.999	153	100	31	31,0	69	69,0
5.000-9.999	252	153	87	56,9	66	43,1
10.000-14.999	150	91	52	57,1	39	42,9
15.000-19.999	92	55	30	54,5	25	45,5
20.000-24.999	46	30	19	63,3	11	36,7
25.000-29.999	38	28	14	50,0	14	50,0
30.000-34.999	12	7	3	42,9	4	57,1
>35.000	92	68	19	27,9	49	72,1
Total	835	532	255	47,9	277	52,1

Tabela A7
Retornos de escala e eficiência econômica (métodos FGL e de Banker):
região Norte

Classes de população	Nº municípios	Método FGL						Método de Banker					
		Retornos constantes		Retornos decrescentes		Retornos crescentes		Retornos constantes		Retornos decrescentes		Retornos crescentes	
		Nº	%	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
0-4.999	13	2	15,4	1	7,7	10	76,9	2	15,4	1	7,7	10	76,9
5.000-9.999	29	1	3,4	0	0,0	28	96,6	1	3,4	0	0,0	28	96,6
10.000-14.999	27	0	0,0	0	0,0	27	100,0	0	0,0	0	0,0	27	100,0
15.000-19.999	32	2	6,2	1	3,1	29	90,6	2	6,2	1	3,1	29	90,6
20.000-24.999	20	0	0,0	8	40,0	12	90,0	1	5,0	8	40,0	11	55,0
25.000-29.999	11	2	18,2	2	18,2	7	63,6	2	18,2	2	18,2	7	63,6
30.000-34.999	13	2	15,4	4	30,8	7	53,0	2	15,4	4	30,8	7	53,8
>35.000	46	2	4,3	43	93,5	1	2,2	2	4,3	43	93,5	1	2,2
Total	191	11	5,8	59	30,9	121	63,4	12	6,3	59	30,9	120	63,4

Tabela A8
Retornos de escala e eficiência econômica (métodos FGL e de Banker):
região Centro-Oeste

Classes de população	Nº municípios	Método FGL						Método de Banker					
		Retornos constantes		Retornos decrescentes		Retornos crescentes		Retornos constantes		Retornos decrescentes		Retornos crescentes	
		Nº	%	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
0-4.999	97	2	2,1	0	0,0	95	97,9	1	1,0	0	0,0	95	97,9
5.000-9.999	109	1	0,9	1	0,9	107	98,2	1	0,9	1	0,9	107	98,2
10.000-14.999	59	0	0,0	4	6,8	55	93,2	0	0,0	4	6,8	55	93,2
15.000-19.999	31	1	3,2	10	32,3	20	64,5	0	0,0	10	32,3	21	67,7
20.000-24.999	25	0	0,0	14	56,0	11	44,0	0	0,0	14	56,0	11	44,0
25.000-29.999	10	0	0,0	5	50,0	5	50,0	0	0,0	5	50,0	5	50,0
30.000-34.999	13	0	0,0	10	76,9	3	23,1	0	0,0	10	76,9	3	23,1
>35.000	24	2	8,3	22	91,6	0	0,0	2	8,3	22	91,6	0	0,0
Total	368	6	1,6	66	17,9	296	80,4	4	1,1	66	17,9	297	80,7

Tabela A9
Retornos de escala e eficiência econômica (métodos FGL e de Banker):
região Sul

Classes de população	Nº municípios	Método FGL						Método de Banker					
		Retornos constantes		Retornos decrescentes		Retornos crescentes		Retornos constantes		Retornos decrescentes		Retornos crescentes	
		Nº	%	Nº	%	Nº	%	Nº	%	Nº	%	Nº	%
0-4.999	153	1	0,6	0	0	152	99,3	1	0,6	0	0,0	152	99,3
5.000-9.999	252	1	0,4	0	0	251	99,6	0	0,0	0	0,0	252	100,0
10.000-14.999	150	0	0,0	0	0	150	100,0	0	0,0	0	0,0	150	100,0
15.000-19.999	92	1	1,1	0	0	91	98,9	0	0,0	1	1,1	91	98,9
20.000-24.999	46	8	17,4	0	0	38	82,6	1	2,2	7	15,2	38	82,6
25.000-29.999	38	8	21,0	0	0	30	79,0	2	5,3	6	15,8	30	79,6
30.000-34.999	12	5	41,7	0	0	7	58,3	0	0,0	5	41,7	7	58,3
>35.000	92	73	79,4	0	0	19	20,6	3	3,2	70	76,1	19	20,6
Total	835	97	11,6	0	0	738	88,4	7	0,8	89	10,7	738	88,5