Available online at www.sciencedirect.com
ScienceDirect
EconomiA 18 (2017) 98–116
ECONOMIA
www.elsevier.com/locate/econ

Are more resources always the answer? A supply and demand analysis for public health services in Brazilian municipalities

Fabiana Rocha [a], Janete Duarte [b], Plínio Portela de Oliveira [b], Luis Felipe Vital Nunes Pereira [b], Sérgio Ricardo de Brito Gadelha [b,*]

[a] *Department of Economics, Universidade de São Paulo (FEA/USP), Brazil*
[b] *National Treasury Secretariat of Brazil, Brazil*

Received 26 August 2015; received in revised form 18 October 2016; accepted 25 October 2016
Available online 15 December 2016

Abstract

The aim of this paper is to determine whether it is necessary to increase available resources to local governments or if better use of these funds is sufficient. The paper contributes to the literature by looking both at the supply and demand for public health services. If the demand is estimated correctly, one can compare its expected value to actual health expenditures. Even if actual expenditures are lower than the estimated demand, it is not certain that additional spending is necessary. If the efficiency scores (supply side) indicate that local governments can simply "save" resources to make up for the difference, then it is possible to reduce (or bring to zero) new resources only by requiring local governments to efficiently manage their expenditures. Since municipalities in Brazil are very heterogeneous, we estimate their efficiency using the metafrontier approach (O'Donnell et al., 2008), while we estimate the demand through an equation derived from the median voter theorem model. Using 2010 data, we find evidence that efficient management of spending is sufficient to meet excess demand for goods and services in the health sector.

JEL classification: C31; D24; I12; H72

Keywords: Health; Municipalities; Metafrontier; Data envelopment analysis; Median voter; Quantile regression

Palavras-chaves: saúde; municípios; metafronteira; análise envoltória de dados; eleitor mediano; regressão quantílicas

Resumo

O objetivo deste artigo é avaliar se é necessário aumentar os recursos disponíveis para os governos locais, ou se o melhor uso desses recursos é suficiente. Esse estudo contribui à literatura, tanto na oferta quanto na demanda de serviços públicos em saúde. Se a demanda é estimada corretamente, é possível comparar o valor previsto da demanda com o gasto efetivamente realizado em saúde. Se o gasto efetivamente realizado é menor do que o gasto demandado, não haverá, a princípio, necessidade de gastos adicionais. Se os escores de eficiência (lado da oferta) indicarem que os governos locais podem simplesmente "economizar" recursos para compensar a diferença, então é possível reduzir (ou trazer a zero) novos recursos apenas exigindo que os governos locais gerenciem eficientemente suas despesas. Uma vez que os municípios no Brasil são muito heterogêneos, estimamos sua eficiência usando a abordagem de metafronteira (O'Donnell et al., 2008), enquanto estimamos a demanda através de uma equação derivada do modelo

* Corresponding author.
E-mail addresses: frocha@usp.br (F. Rocha), janete.duarte@tesouro.gov.br (J. Duarte), plinio.oliveira@tesouro.gov.br (P.P. de Oliveira), luis.n.pereira@tesouro.gov.br (L.F.V.N. Pereira), sergio.gadelha@tesouro.gov.br (S.R. de Brito Gadelha).

do teorema do eleitor mediano. Usando dados de 2010, encontramos evidências de que uma gestão eficiente dos gastos é suficiente para atender à demanda excessiva de bens e serviços no setor de saúde.

1. Introduction

To enhance service delivery, the Brazilian Constitution of 1988 transferred the responsibility for health and basic education provision to local governments. The federal government was made responsible for establishing health policy guidelines, while states and municipalities became responsible for putting them into practice. Although financing should also be shared, only the federal government's role was established by the transient constitutional provisions, determining that 30% of the social security budget must be allocated to the health sector.

The early 1990s health financing crisis can be explained by the lack of a clear distribution of tasks among federal, state, and local governments, by the lack of strong revenue sources and by the limited commitment of each level of government to expenditure financing. Consequently, some changes in the constitutional text were suggested, giving rise to Constitutional Amendment no. 29/2000, which established that federal, state, and local funds should be allocated to health financing. Thus, the shared responsibility of financing the Unified Health System (Sistema Único de Saúde—SUS) established by the Constitution was strengthened. The federal government should spend in health services the same amount spent in the last year plus 5%. From then on, values should be adjusted according to the nominal variation of the gross domestic product (GDP). States and municipalities should devote a minimum amount to the health sector corresponding to a percentage of their own tax revenue and constitutional and legal transfers. This percentage gradually increased, and since 2004, it has been 12% for states and 15% for municipalities.

Nevertheless, there are recurrent complaints that health spending in Brazil is quite low and should increase. Médici (2011), however, notes that the amount of Brazilian public health expenditures is not so different from that seen in other countries with similar per capita income. Therefore, "saying that Brazil spends little on health would make more sense if it was possible to guarantee that the existing funds are being spent efficiently and that, in spite of this, the population's health needs are not being sufficiently or properly met" (Médici, 2011, p. 46).

The literature on the assessment of efficiency in the health sector can be split into two different strands.

The first derives from the World Health Report 2000. The original report, written by Evans et al. (2000) for the World Health Organization (WHO), proposes the use of a fixed-effect panel data model to create a production frontier in which individual country fixed-effects are used as inefficiency indicators. The following works use the same data but introduce extensions to the model used by WHO, sometimes employing alternative methods for efficiency estimation (Navarro, 2000; Williams, 2001; Hollihgsworth and Wildman, 2003; Gravelle et al., 2003).

The second strand basically seeks to calculate health efficiency scores using non-parametric and parametric techniques and to estimate their main determinants (Gupta et al., 2002; Afonso and St. Aubyn, 2005).

The available evidence for Brazil follows the second strand of the literature. Marinho (2003) assesses the efficiency of municipalities in the state of Rio de Janeiro in providing outpatient and inpatient health services. Brunet et al. (2006) compare the use of funds by Brazilian states and by the Federal District to the supply of products and services (efficiency) and to the observed outcomes (effectiveness). Mattos et al. (2009) calculate efficiency scores for municipalities in the state of São Paulo in order to evaluate the impact of scale variables on the quality of public health services. Souza et al. (2010) also look at municipalities in the state of São Paulo, but they seek to assess the productive efficiency of the hospital sector between 1998 and 2003 and its determinants.

The aim of the current paper is to assess Brazilian local health expenditures in an attempt to determine whether it is necessary to increase the amount of public funds they allocate to the health sector or if the improvement of efficiency in the use of the existing funds is sufficient to meet the population's demand for public health services.[1] To achieve our purpose, we estimate the supply of public health services from the calculus of efficiency scores using the metafrontier methodology and the demand for public health services from an equation based on the median voter model. If the

[1] As noted by Mirmirani et al. (2008), "Health care costs are a financial burden for developing and transition economies, which have experienced a faster growing demand on their health care systems while aiming to improve efficiency".

demand is estimated correctly, it is possible to compare the estimated demand with actual healthcare expenditures. If actual expenditures are lower than the estimated demand, it still could be the case that no additional expenditures are necessary. If the supply analysis indicates that it is possible to "save" enough resources to make up for the difference, municipalities only need to manage their spending in a more efficient way.

Regarding the supply side, we estimate the relative performance of municipalities within a certain group, but we also compare their performance across groups. We then measure municipalities' efficiencies relative to a common metafrontier, i.e., the boundary of an unrestricted technology set. We also calculate efficiency scores in terms of group frontiers, defined as restricted technology set frontiers, where restraints stem from different characteristics of municipalities' production environments. More precisely, efficiency relative to the metafrontier is broken down into a component that measures the distance of an input–output point to the group frontier and another component that measures the distance between the group frontier and the metafrontier. The former component constitutes the usual measures of technical efficiency, while the latter takes into account the restrictive nature of the production environment (O'Donnell et al., 2008).

To estimate and predict health service demand, we use an equation derived from the median voter theoretical approach. As far as we know, there are only two other papers that estimate public expenditure demand in Brazil using this methodology: Mendes and Sousa (2006) and Menezes et al. (2011). Menezes et al. (2011), however, are not directly interested in the demand estimations. Their purpose is to verify whether deviations between the actual basket of goods and that demanded by the median voter affect the probability of electoral success, defined as the vote share or as the probability of being reelected.

The paper is organized as follows. In Section 2, we describe the methodology we use to calculate the efficiency scores in the provision of health services. In Section 3, we present the input and output variables and discuss the results from the supply analysis. In Section 4, we describe the equation we use to estimate the demand for local public health. In Section 5, we discuss the demand analysis results, while in Section 6, we combine the evidence from the supply and demand analysis in order to determine whether there is a shortage or an excess of funds, and in the case of shortage, whether it can be met by only improving efficiency. In Section 7, we check the robustness of our results using an alternative group (homogenous groups defined by the Brazilian Ministry of Health) to estimate the efficiency scores. Finally, in Section 8, we summarize the main conclusions.

2. Methodology: supply[2]

2.1. Basic concepts

When production possibility frontiers are estimated to evaluate public spending efficiency, the heterogeneity of countries, states, or municipalities is not taken into account.[3]

Estimating a global production frontier for a municipality set thus implies assuming that municipalities share a common healthcare production technology. However, the different environments in which municipalities operate influence mayors' and health managers' capacity and desire to implement technological innovation. Not only do municipalities make choices based on different input and output combinations, but they also have distinct technology sets owing to differences in their physical, human, and financial capital stock, economic infrastructure, fund availability, etc.

Estimating separate frontiers for municipality groups (subsamples) instead of a common frontier for the whole sample does not solve the problem, given that the technical efficiency scores for municipalities in different groups are not directly comparable.

An alternative is to use the metafrontier approach. The metafrontier function is an envelope curve of production points of the most efficient municipalities. Each municipality can operate on a different production possibility frontier segment, according to their available funds, technology adoption and diffusion and economic environment.

[2] As far as we know, Balaguer-Coll et al. (2010) are the only ones to apply the metafrontier approach to analyze municipality efficiency. They evaluate the Spanish public sector as a whole and use the mix of products provided by municipalities, environmental conditions, and size as ranking criteria.

[3] An exception is Gupta et al. (2002), who sort the countries in the sample according to their income levels to verify the impact of differences in their economic development in their efficiency in providing health and education.

Battese and Rao (2002) compare firm technical efficiencies in different groups that might not have the same technology using a stochastic metafrontier production function. They assume there are two different data generating processes, one related to the stochastic frontier, which is estimated using group-specific data, and another related to the metafrontier, which is estimated using data from the whole sample. The resulting technological gap provides information on the ability of firms in a given group to compete with firms from different groups within the same industry. It shows the size of the technology gap for a given firm whose current available technology is inferior to that available to all firms represented by the metafrontier. The drawback of this approach is that the values of the metafrontier production function can be lower than the deterministic components of the group stochastic frontier production functions.

Battese et al. (2004) solve this problem by explaining deviations between observed outputs and group frontiers through a single data-generating process. In addition, they define the metafrontier as a function that envelops deterministic components of the estimated stochastic frontier for several groups. They estimate the metafrontier using the stochastic frontier production model with time-varying inefficient effects.

Finally, O'Donnell et al. (2008) use both non-parametric (Data Envelopment Analysis—DEA) and parametric (several stochastic frontier approaches) to estimate metafrontiers and group frontiers, besides breaking down performance differences across firms into technical efficiency and technology gap effects.

2.2. Metafrontier cost

To estimate the metafrontier and group frontiers, we will use DEA. There are two available types of DEA model. One is the output-oriented model, in which inputs are held constant and the aim is to maximize output-proportional increases. The other is the input-oriented model, in which output is held constant and the aim is to seek maximum input-proportional reductions.[4]

Because we want to verify whether municipalities can use fewer funds to achieve their current healthcare results, we use an input-oriented DEA model. More precisely, we estimate a metafrontier cost function, which is the specific cost frontier envelopes for municipality groups.

The metafrontier estimation follows the method proposed by O'Donnell et al. (2008), which involves the following steps:

1) Classify all municipalities into $S_1, S_2, \ldots, S_k$ sets.
2) Estimate α_i^S, the efficiency scores for each i municipality within its correspondent set.
3) Apply DEA to the whole sample in order to obtain the efficiency score of each municipality in relation to the metafrontier (α_i^M).
4) Calculate α_i^M/α_i^S, which are called technological gap ratios by Battese et al. (2004) and metatechnology ratios by O'Donnell et al. (2008).

The metatechnology ratio (MTR_i) basically assesses the size of the technology gap for a given set of municipalities whose current adopted technology lags behind the technology available for all municipalities, represented by the metafrontier cost function. At a given output level, the metatechnology ratio is defined as the smallest possible cost within the metafrontier divided by the smallest cost in the specific set. Thus, the higher the metatechnology ratio mean value for a given set, the better the production technology adopted by it (O'Donnell et al., 2008).

In practice, the technical efficiency related to metafrontier cost is obtained using the following decomposition:

$$\alpha_i^M = \alpha_i^S \times MTR_i \quad (1)$$

where α_i^S is the conventional technical efficiency, which measures the deviation of the municipality's effective cost from the specific group cost frontier to which it belongs, while MTR_i measures the deviation of the specific group frontier from the metafrontier cost function. The metafrontier cost efficiency score indicates how good the performance of a municipality is in relation to the expected performance from pairs with the best practices and explores the best technology available for all groups.

[4] The two types of model give the same efficiency scores under constant returns to scale technology but different scores under variable returns to scale technology.

DEA will be carried out through a standard linear program. Below, we present the constant returns to scale input-oriented version of the model:

$$\begin{aligned} &\min \alpha, \\ &\alpha, \lambda \\ &s.t. \quad -q_i + Q\lambda \geq 0, \\ &\qquad \alpha x_i - X\lambda \geq 0, \\ &\qquad \lambda \geq 0, \end{aligned} \tag{2}$$

where q_i and x_i are column vectors of the assessed municipality output and input levels, respectively; Q is a matrix formed by column vectors corresponding to the outputs of all municipalities; X is the corresponding input matrix; and λ is the vector of weights assigned to each municipality, resulting from the minimization process. The variable returns to scale version of the model requires an additional restriction in which the sum of weights equals 1.

To calculate the technical efficiency with respect to group frontiers (step 2), we apply the DEA methodology to the municipalities in each group (cluster), while to calculate the technical efficiency with respect to the metafrontier (step 3), we apply the DEA methodology to all municipalities. Once we estimate the municipalities' technical efficiencies with respect to the metafrontier and to the respective group frontier, we can proceed to step 4 and easily calculate the metatechnology ratio.

In January 2001, the World Health Organization (WHO) organized a seminar targeted at obtaining the opinion of a group of renowned experts about their approach to measuring health efficiency and gathering subsidies and suggestions to allow improvements and new developments in the future (Evans et al., 2000). Two agreements were reached at the seminar. The first is that the frontier method is appropriate to measure health efficiency, and the second is that frontiers for groups of countries should be estimated separately.[5]

3. Empirical application: public health services supply

3.1. Data

We use a sample of 5523 of the 5565 Brazilian municipalities in 2010.

At the WHO meeting, there was also agreement that variables that are actual production factors and those that could explain the observed inefficiencies should be made distinguishable. Only variables that are direct production factors (labor and capital in the traditional microeconomic analysis) should theoretically be included in the production process estimation. Variables that could explain differences in efficiency should not be used as production factors but as inefficiency determinants. Education, for example, may be regarded as a direct production factor if taken as available knowledge, or it could be regarded as a proxy for other inputs, such as housing and nutrition, when these data are unavailable. We agree with the input selection made by WHO (Evans et al., 2000) and use two inputs: per capita health expenditures as a summary variable for health inputs and average years of schooling of the adult population as a summary indicator for non-health inputs.

The health expenditure data come from the Brazilian Treasury Department—Finanças do Brasil (FINBRA),[6] while the population data come from the Brazilian Institute of Geography and Statistics (IBGE).[7] Schooling of individuals older than 25 is also from the IBGE.

The literature usually uses infant mortality rates and life expectancy as output measures. We, however, use a broader output measure, the Brazilian National Health System Performance Index (*Índice de Desempenho do Sistema Único de Saúde*—IDSUS), which seeks to measure the performance of each municipality regarding access to and the

[5] WHO jointly estimates efficiency for all countries, assuming that all countries have the same technology available and that the amount of resources is the main limiting factor for its use. Moreover, the adoption of technology does not depend on context. If these hypotheses can actually be regarded as reasonable, efficiency scores for all countries can be estimated together.

[6] Available at http://www.tesouro.fazenda.gov.br/finbra-financas-municipais. Accessed 23 February 2015.

[7] Available at http://www.ibge.gov.br/home/. Accessed 23 February 2015.

effectiveness of the health care system. The index consists of twenty-four indicators (fourteen measure health care access, and ten measure system effectiveness).[8]

Specifically, IDSUS is a set of simple and compound indicators aimed at making a contextual assessment of Unified Health System (SUS) performance at the municipal level, in compliance with its principles and guidelines. It is therefore suitable for use as an output variable.

Associations between variables will be explored using descriptive statistics and a cumulative distribution function approach, which involves calculating frequency distributions for variables using empirical cumulative distribution function (CDF) curves.[9]

Because we are working with groups (clusters), two questions immediately arise. The first regards the criteria to measure similarity. The second regards the choice of clustering method. To deal with the former question, it is necessary to know which variable(s) is(are) better suited to separate the municipalities into different groups. Theoretically, there is no preset criterion that could indicate which environmental conditions have an impact on the municipalities' "production function".

Therefore, we initially choose to group the municipalities according their population. We can provide three reasons to justify our choice.

First, articles 48, 48-A and 73-B of the Brazilian Fiscal Responsibility Law[10] use the population size to check the fiscal transparency of local governments. Therefore, we have an exogenous classification of municipalities by size that can be readily used. Because only one variable defines the groups, we also avoid the second question, which involves the choice of a methodology to form the clusters when more than one variable is used to define similarity. To verify whether municipalities are attending their fiscal targets, the Fiscal Responsibility Law split them into three groups: up to 50,000 inhabitants, between 50,000 and 100,000 inhabitants, and over 100,000 inhabitants. Since the third group includes very heterogeneous municipalities, we split it into two subgroups: municipalities with up to 500,000 inhabitants and municipalities with over 500,000 inhabitants.[11]

Second, there is some evidence in the Brazilian literature that scale is an important determinant of municipality (in)efficiency. For example, Sousa et al. (2005) use population density as a scale variable and note that municipalities with small population density spent relatively more. Low population density leads to an increase in the average costs of public health care delivery, which hinders the exploitation of scale economies associated with the production of these services and is, ultimately, associated with suboptimal use of funds.

Finally, the Brazilian health system has a hierarchical organization. Small municipalities are responsible for primary care actions, while the larger municipalities and the states are responsible for providing medium- and high-complexity procedures, which renders the classification by size natural.

As noted by Sousa et al. (2005), "one of the main drawbacks of DEA is that the resulting efficiency scores are rather sensitive to the presence of DMUs that perform extremely well (the so called outliers), which may stem from outstanding practice or may simply be the result of errors in the data. In either case, the results from the remaining DUMs become shifted toward lower efficiency levels, the efficiency distribution becomes highly asymmetric, and the overall efficiency scale becomes nonlinear (p. 290)". There are several methods for outlier detection, for example, those proposed by Wilson (1993, 1995), Simar (2003) and the extension of Wilson's method proposed by Sousa and Stosic (2003). We use the method proposed by Wilson (1993), implemented in the FEAR software (Wilson, 2008).

[8] More details on the national health system index (IDSUS), such as its methodology and composing variables, can be obtained at http://idsus.saude.gov.br/. Accessed 23 February 2015.

[9] An empirical CDF plot is a graph that can be used to evaluate the fit of a distribution to the data, to estimate percentiles and to compare different sample distributions. An empirical CDF plot: (i) plots each unique value versus the percentage of values in the sample that are less than or equal to it and connects the points with a stepped line; (ii) fits a cumulative distribution function (CDF) for the selected distribution so that it is possible to examine how well the distribution fits the data; (iii) displays a table with the distribution parameter estimates and the number of observations (N) for the data. An empirical CDF plot performs a similar function as a probability plot. However, unlike a probability plot, the empirical CDF plot has scales that are not transformed, and the fitted distribution does not form a straight line.

[10] Supplementary Law No. 101/2000. Available at http://www.planalto.gov.br/ccivil_03/leis/LCP/Lcp101.htm.

[11] We also use only three groups, but because the results do not change much, we report only the results using four groups.

Table 1
Descriptive statistics for all municipalities.

	Per capita public health expenditures	Years of schooling	IDSUS
Group 1: Up to 50,000 inhabitants			
Mean	359.0577	3.623721	5.700421
Median	316.9354	3.508310	5.650000
Maximum	2155.031	9.315312	8.370000
Minimum	9.739348	1.047243	2.500000
Std. dev.	160.3198	1.106374	0.822662
Skewness	2.018874	0.443262	0.065716
Kurtosis	10.99646	2.850043	3.019171
Jarque–Bera	16,447.24	165.6909	3.615898
Probability	0.000000	0.000000	0.163990
Observations	4919	4919	4919
Group 2: Over 50,000 inhabitants to 100,000 inhabitants			
Mean	285.0821	4.962451	5.145247
Median	258.6207	5.035421	5.180000
Maximum	1762.569	7.850125	7.280000
Minimum	47.01527	1.508022	3.000000
Std. dev.	143.7163	1.404548	0.844672
Skewness	4.287303	−0.267130	−0.085533
Kurtosis	38.61223	2.285294	2.640001
Jarque–Bera	18,113.69	10.74924	2.144646
Probability	0.000000	0.004633	0.342213
Observations	324	324	324
Group 3: Over 100,000 inhabitants to 500,000 inhabitants			
Mean	327.4974	6.378666	5.422222
Median	304.0821	6.356023	5.400000
Maximum	999.3008	10.01433	8.210000
Minimum	82.88280	3.166306	3.530000
Std. dev.	154.4415	1.102786	0.797194
Skewness	1.447886	0.259576	0.196321
Kurtosis	6.580711	3.853686	2.976680
Jarque–Bera	214.7208	10.10777	1.566460
Probability	0.000000	0.006384	0.456928
Observations	243	243	243
Group 4: Over 500,000 inhabitants			
Mean	420.3226	7.651585	5.580811
Median	414.3095	7.775878	5.580000
Maximum	620.0786	9.134573	6.960000
Minimum	130.9390	6.095313	4.180000
Std. dev.	122.7260	0.730259	0.704247
Skewness	−0.277444	−0.331634	−0.140918
Kurtosis	2.636471	2.582662	2.341031
Jarque–Bera	0.678417	0.946732	0.791911
Probability	0.712334	0.622902	0.673037
Observations	37	37	37

Data compiled by the authors.

3.2. Results

Table 1 summarizes the descriptive statistics. Looking at the four different groups, we can see that there are huge differences among Brazilian municipalities, which seem to employ different amounts of inputs to provide heterogeneous amounts of output, probably due to the different production environment, captured by the population. This evidence gives additional support to the decision to employ the metafrontier model.

Table 2
Technical efficiency estimates and metatechnology ratios—DAE constant returns to scale.

Groups	No. of municipalities	Mean	Minimum	Maximum
Technical efficiency with respect to group frontiers				
Up to 50,000 inhabitants	4909	0.681	0.347	1.000
Over 50,000 inhabitants to 100,000 inhabitants	317	0.738	0.500	1.000
Over 100,000 inhabitants to 500,000 inhabitants	237	0.717	0.462	1.000
Over 500,000 inhabitants	34	0.879	0.699	1.000
Metatechnology ratios				
Up to 50,000 inhabitants	4909	1.000	1.000	1.000
Over 50,000 inhabitants to 100,000 inhabitants	317	0.793	0.769	0.816
Over 100,000 inhabitants to 500,000 inhabitants	237	0.796	0.793	0.841
Over 500,000 inhabitants	34	0.626	0.522	0.781
Technical efficiency with respect to the metafrontier				
Up to 50,000 inhabitants	4909	0.681	0.347	1.000
Over 50,000 inhabitants to 100,000 inhabitants	317	0.585	0.384	0.816
Over 100,000 inhabitants to 500,000 inhabitants	237	0.570	0.367	0.841
Over 500,000 inhabitants	34	0.550	0.365	0.781

Data compiled by the authors.

The IDSUS average value is approximately 5.5 for all groups, with low dispersion according to standard deviation values, and is normally distributed according to the Jarque–Bera (JB) test.[12]

Per capita public health expenditures vary nonlinearly according to population. In group 1, the average per capita health spending is R$ 359.06, with a standard deviation of R$ 160.32; in group 2, it is R$ 285.08, with a standard deviation of R$ 143.72; in group 3, it is R$ 327.50, with a standard deviation of R$ 154.44; and in group 4, the average value is R$ 420.32, with a standard deviation of R$ 122.73. The municipalities in group 1 (very small municipalities) spend more in per capita terms than municipalities in groups 2 and 3. The dispersion is high in all groups, and it is not possible to reject the normality hypothesis only for group 4.

Finally, the average years of schooling variable is quite low for the four groups, revealing that the Brazilian population is, on average, poorly educated. For very small municipalities (group 1), the adult population has, on average, only 3.62 years of schooling. Education increases with population size. The average years of schooling are 4.96, 6.38 and 7.65 for groups 2, 3, and 4, respectively. The Jarque–Bera test indicates that it is not possible to reject the normality null hypothesis for only group 4.

In Appendix A, we also show empirical cumulative distribution functions for the output and input variables.

Table 2 summarizes the technical efficiency estimates and metatechnology ratios under the constant returns to scale hypothesis. We use the term “technology” to facilitate comparison with other studies that usually have firms as decision-making units. However, in fact, what we suggest is that what matters is the effect of municipality size and the production environment that may be associated with it.

The average technical efficiency with respect to group frontiers ranges from 0.681 for smaller municipalities (group 1) to 0.879 for the larger ones (group 4). These figures imply that, on average, small municipalities can reduce their actual expenditures by 32%. On average, large municipalities are closer to the frontier, but inefficiency is still a serious problem, as there is a savings potential of 12%.

Although the municipalities in group 4 have the highest efficiency scores among all groups, they seem to adopt poorer production technologies (they show the smallest metatechnology ratios). Conversely, small municipalities operate with better technologies, but at the expense of lower average efficiency scores. Municipalities in group 4 show production costs above and far from their metafrontier cost, and they could cut their costs by up to 37% if they adopt the technology available to all municipalities (metafrontier technology).

[12] If the data come from a normal distribution, the *JB* statistic asymptotically has a chi-squared distribution with two degrees of freedom. The null hypothesis is a joint hypothesis of zero skewness and zero excess kurtosis. Samples from a normal distribution have an expected skewness of 0 and an expected excess kurtosis of 0 (which is the same as a kurtosis equal to 3). As the *JB* definition shows, any deviation from this increases JB statistics.

Table 3
Technical efficiency estimates and metatechnology ratios—DEA variable returns to scale.

Groups	No. of municipalities	Mean	Minimum	Maximum
Technical efficiency with respect to group frontiers				
Up to 50,000 inhabitants	4909	0.765	0.765	1.000
Over 50,000 inhabitants to 100,000 inhabitants	317	0.839	0.699	1.000
Over 100,000 inhabitants to 500,000 inhabitants	237	0.885	0.773	1.000
Over 500,000 inhabitants	34	0.970	0.921	1.000
Metatechnology ratios				
Up to 50,000 inhabitants	4909	1.000	0.713	1.000
Over 50,000 inhabitants to 100,000 inhabitants	317	0.839	0.814	0.907
Over 100,000 inhabitants to 500,000 inhabitants	237	0.743	0.688	0.848
Over 500,000 inhabitants	34	0.644	0.678	0.891
Technical efficiency with respect to metafrontier				
Up to 50,000 inhabitants	4909	0.765	0.545	1.000
Over 50,000 inhabitants to 100,000 inhabitants	317	0.704	0.569	0.970
Over 100,000 inhabitants to 500,000 inhabitants	237	0.657	0.531	0.848
Over 500,000 inhabitants	34	0.625	0.625	0.891

Data compiled by the authors.

In fact, while the municipalities from groups 2, 3 and 4 show remarkable differences between group efficiencies and metafrontier efficiencies, we do not find such a difference for municipalities in group 1. The metatechnology ratios for the municipalities in group 1 are equal to 1. In fact, these municipalities determine the metafrontier, given that the metatechnology ratio is always lower than 1 for the other groups. This is evidence that municipalities with a population over 50,000 inhabitants are part of a more adverse production environment in the delivery of public health services.

The average technical efficiency with respect to the metafrontier of groups 2, 3 and 4 is equal to 0.58, 0.57 and 0.55, respectively, which suggests that their municipalities should try to adopt the potential technology available to all municipalities in order to move their cost frontiers downward.

Table 3 summarizes the technical efficiency estimates and metatechnology ratios under the hypothesis of variable returns to scale. The results are similar to those under the constant returns to scale hypothesis.

As noted by O'Donnell et al. (2008), technical efficiency estimates are ultimately calculated to be used on programs for performance improvement based on municipality management changes. Metatechnology ratios, in turn, can be used on programs that seek to change the characteristics of the environment where production takes place (for instance, infrastructure and financing). Therefore, there seems to be evidence that both the restrictive nature of the production environment and the mismanagement of funds affect municipalities' efficiency in Brazil.

4. Methodology: demand

The literature on public sector efficiency only takes into account the supply side, and it has as its main purpose to obtain efficiency scores in order to rank the units of analysis (countries, states, municipalities). However, we believe it is necessary to evaluate the demand side as well. Although more resources can be made available if municipalities improve their efficiency, the increase cannot be sufficient to meet the demand.

Municipalities are different in terms of economic and social conditions, geography, and other characteristics. As we discussed before, these different environmental aspects may have a strong impact on municipalities' performance. Therefore, some municipalities should be classified as inefficient, or more inefficient than other municipalities, just because they face different local conditions. The metafrontier methodology allows us to adequately compare local governments facing similar environmental conditions.

We then divide municipalities according to size, since the levels of health services each local government must provide depend on the population level.

To make the supply and demand analysis compatible, we should then estimate demand functions for municipalities of up to 50,000 inhabitants, etc. . ., that is, estimate the demand also for groups using size as a criterion.

However, to guarantee the robustness of our results, we should divide municipalities into different groups according to at least a different classification other than size. In fact, in Section 7, we use the grouping of municipalities defined by the Ministry of Health to assess the IDSUS as an alternative criterion to form groups of municipalities. The problem is that whenever we define new groups for the metafrontier analysis, we should estimate the demand equation again for these different groups.

To avoid the need to re-estimate demand every time a different classification of local governments is adopted, and at the same time guarantee that some heterogeneity is taken into account when estimating the demand function, we use a quantile regression approach. If the estimated coefficients are similar across the quantiles, we can suppose that the demand is quite stable among the municipalities, and we do not have to evaluate it each time we consider a new group for the metafrontier.

To estimate health care demand, we use the median voter approach, as originally proposed by Bergstrom and Goodman (1973). If preferences have a single maximum and the public good has only one dimension, the median voter theorem implies that votes can turn individual demands into a collective demand.

Despite various criticisms about the median voter model, some factors justify its choice. Using a survey of Michigan households, Bergstrom et al. (1982) test both the Tiebout hypothesis and the median voter hypothesis. The Tiebout hypothesis states that people choose their residential location according the amount of local public spending, i.e., citizens "vote with their feet". The median voter hypothesis states that the actual spending level corresponds to the choice of the median voter, as that spending level will garner more support in an election. The authors estimate the demand for local public goods using survey data on individual demand for public goods, and they find smaller income and price elasticities than those found in earlier macrolevel studies. The parameter estimates are similar to those found by Bergstrom and Goodman (1973) using overall spending on public goods, indicating that the results are robust to the type of data used.

There is some evidence supporting the median voter model. For example, Aronsson et al. (2000) find evidence of vertical interactions between federal and local expenditures using Swedish panel data from 1981 to 1986. This suggests that it is important to take into account potential national spending effects when estimating interactions between municipalities. Moreover, Dahlberg and Johansson (2000) investigate the dynamic relationship between local government revenues and expenditures using GMM bootstrapping techniques for a panel of 265 Swedish municipalities over the period 1979–1987. They find a one-year lag in the expenditure equation but no dynamics in the own-source revenue and grant equations. Although these results are quite different from those obtained when asymptotic critical values are used, they are well in line with the theoretical explanations given in the literature for dynamic behavior in the local public sector.

Regarding Brazil, Mendes and Sousa (2006) estimate the demand for local public spending of Brazilian municipalities using a median voter framework. The median voter theorem provides a method of aggregating individual voter's demands in an aggregate (local) demand. Besides, in federal systems, voter's preferences are more likely to be reflected at the local level, as public service consumers have a better knowledge of the benefits and costs of local public expenditures. They obtain results that are consistent with the theoretical background and suggest that the median voter hypothesis seems to be adequate to describe the demand for local public goods in Brazil. Quantile regressions allow them to investigate the impacts of conditioning variables on local public expenses across different expenditure classes, that is, to account for heterogeneity across municipalities. They find evidence that the size impact on the quality of club goods has crowding effects. However, marginal congestion slightly decreases with expenditure. This is a rather surprising result, as one is tempted to conclude that congestion effects should be higher for large cities. Yet, a careful analysis can show the drawbacks of such interpretation. Indivisibilities preclude the provision of certain public services in small towns and restrict their provision mainly to larger cities. Hence, the higher expenditures of larger cities reflect not only crowding costs but also the fact that these cities offer a wider range of public services relative to the small ones. Thus, in Brazil, contrary to traditional results, smaller congestion effects along spending classes reflect the predominance of scale effects, measured by population elasticities over price effects.

Menezes et al. (2011) also test the median voter model for the Brazilian municipalities and find evidence that it seems to be valid. They estimate the demand for local public services in order to obtain a measure of misperception of candidates to reelection regarding the median local demand. The expenditures effectively made by the candidates during their first terms are taken as the bundle of public services they offer. They then evaluate whether this misperception affects the electoral performance of incumbents, measured by their vote share or probability of election, using selection models.

Finally, there is evidence that the median voter model provides a better explanation for certain public programs than the interest group models. For example, Congleton and Bennet (1995) explore the extent to which public demand for roads and/or power of special interest groups determines road expenditures at the state level using an extension of the methodology developed in Congleton and Shughart (1990). They use reduced form models of median voter demand, special interest group equilibria, and a combined model and find support for the hypothesis that voting matters for American states. Pure median voter models show a better fit than pure special interest group models. Moreover, based on their combined model, they find evidence that variables from the median-voter model cannot be dropped without significantly reducing the fit of the combined model.

The standard equation generally used to analyze the demand for local public goods is[13]:

$$\ln G = \kappa + (1 + \beta_1)\left[\ln\left(b_m/b\right)\right] + \beta_2\left(\ln\ \ y\right) + \beta_3\left(\ln\Omega\right) + \beta_4\left(\ln N\right) + \varepsilon \qquad (3)$$

where:

G is the health expenditures per capita of each municipality;

$b = B/N$, in which B is the total tax base and N is the municipality population size;

b_m is the tax base of the median voter;

y is the median voter's income plus the median voter's share in intergovernmental transfers per capita;

Ω is a vector of explanatory variables (control variables);

$(1 + \beta_1)$ is the price elasticity of demand for public health services;

β_2 is the income elasticity of demand for public health services;

β_3 is a parameter vector related to explanatory variables;

β_4 is the population elasticity;

ε is the error term, and its estimate contains information about the inability of local governments to meet the demand of the median voter, as argued by Menezes et al. (2011).

We estimate the median voter model using quantile regressions—QR (Koenker and Bassett, 1978). This method allows us to evaluate the impact of the explanatory variables not only on the dependent variable conditional distribution mean but also on different points along this distribution.

Constant coefficient regression models, such as Ordinary Least Squares (OLS), have been extensively applied in empirical studies, providing, however, only the central distribution measurement of the dependent variable. Unfortunately, these models fail to address the behavior of the dependent variable in the tail regions. To address this issue, various random coefficient models emerged as viable alternatives in the field of statistical application. The Conditional Quantile Regression (QR) Model, proposed by Koenker and Bassett (1978), is one of these alternative models. This approach allows estimating various conditional distribution quantile functions. Among them, the median (0.5th quantile) function is a special case. Each quantile regression characterizes a particular (center or tail) point of the conditional distribution, and putting different quantile regressions together provides a more complete description of the underlying conditional distribution. This analysis is particularly useful when the conditional distribution is heterogeneous and does not have a standard shape, such as in an asymmetric, fat-tailed, or truncated distributions.

Conditional quantile regression can serve as (i) an alternative to least squares when the normality assumption does not hold and (ii) a complement to least squares, allowing one to look beyond mean effects and obtain a more complete picture of the problem (Koenker, 2005). Quantile regression is desired if conditional quantile functions are of interest. One advantage of quantile regression compared to ordinary least squares regression is that quantile regression estimates are more robust against outliers. Different measures of central tendency and statistical dispersion may also be useful to obtain a more comprehensive analysis of the relation between variables. While OLS can be inefficient if errors are highly non-normal, QR is more robust to non-normal errors and outliers. QR also provides a richer data characterization, as it allows one to measure the impact of a covariate on the entire distribution of the dependent variable, not merely on its conditional mean.

Therefore, we use QR to verify the effects of the exogenous variables on different sizes of public health care expenditures per capita. We employ the following quantiles: 0.10 (10th percentile); 0.25 (lower quartile); 0.50 (median); 0.75 (upper quartile); and 0.90 (90th percentile).

[13] Details on how the model is derived can be obtained in Mendes and Sousa (2006) or Menezes et al. (2011), among others.

5. Empirical application: public health services demand

5.1. Data

Per capita health expenditures are from the Brazilian Treasury Department—*Finanças do Brasil* (FINBRA), while the remaining data are from the 2010 Census (IBGE).

Tax share (b_m/b) corresponds to the ratio between median and average income. Median income corresponds to median income + (LTR/OR) × intergovernmental transfers per capita, where (LTR/OR) is the ratio between local tax revenue (LTR) and overall revenues (OR). On the other hand, intergovernmental transfers per capita are the sum of constitutional and legal transfers (from federal and state governments) to municipalities plus conditional transfers to education and health.

The control variables include health care expenditures per capita by neighboring municipalities. The spatial effects from neighboring municipalities' expenditures (spillovers effects) are well known, and the smaller the area in which municipalities are located, the higher the probability of correlation. Because it is not possible to combine quantile regressions and spatial econometrics, we use the average per capita health expenditure by municipalities that belong to the same microregion (as defined by IBGE), except for the per capita health expenditure of the municipality under consideration to take into account spillover effects, following Videira and Mattos (2011). We name this variable neighborhood.

The remaining control variables are: population, the share of the population aged from 0 to 15 years, urbanization rate, population density, and region dummies.

5.2. Results

Table 4 summarizes the results. In the first column, we provide OLS estimates and in the remaining columns the quantile results. The evidence is quite consistent.

Tax share elasticity is negative and statistically significant at 1%, indicating that increases in the price of public health sector goods and services are associated with a decline in demand. It is important to note that the magnitude of the impact of tax share across the spending categories is the same for municipalities in the 0.10 and 0.90 quantiles. This result is, to some extent, surprising. Municipalities with larger expenditures tend to have a more solid and diverse tax base, such that their high tax share elasticities indicate broad diversification and larger integration with the market economy (Mendes and Sousa, 2006).

Total median income positively affects the demand for health services. The local public spending income elasticity is statistically significant at 1% and is lower than one in magnitude, as also observed by Mendes and Sousa (2006), which indicates that these goods are not luxury goods, since they are targeted at the basic needs of the population. In other words, this result implies that public goods and services in the health sector have the same characteristics as normal goods (Menezes et al., 2011). The impact of total median income increases slightly across expenditure categories, indicating that higher development levels lead to increasing pressures on these public goods and services.

The elasticity of the neighborhood variable is positive and statistically significant at 1% and 5%. Therefore, there seems to be a spatial dependence among municipalities, such that high (small) levels of government spending tend to spread to neighboring municipalities due to spillover effects (Mendes and Sousa, 2006). Unlike the results obtained by Mendes and Sousa (2006), the neighborhood variable does not show an uptrend across spending categories, even though it indicates that the spillover effect tends to be a little stronger in municipalities within the 0.10 and 0.50 quantiles.

The larger the population, the higher the demand for public goods and services in the health sector, as found by Mendes and Sousa (2006). However, we do not find an uptrend of the parameter associated with this variable across spending categories as do Menezes et al. (2011).

The elasticity of population to local per capita health expenditure is positive and statistically significant at 1% in the 0.50, 0.75 and 0.90 quantiles. It is negative and statistically significant at 1% in the 0.10 quantile. The supply of public goods and services in the health sector is subject to diseconomies of scale, so that in municipalities with smaller populations, the supply of complex health services is not justifiable. Thus, the demand for health services will eventually be smaller. In addition, there is no evidence of agglomeration economies, as we do not observe an increase in the estimated elasticity across spending categories.

Table 4
Demand determinants: health expenditures per capita, 2010. Quantile regression results.

Regressors	OLS	(0.10)	(0.25)	(0.50)	(0.75)	(0.90)
Tax share	−0.257	−0.238	−0.201	−0.208	−0.213	−0.238
	(0.039)**	(0.046)***	(0.052)***	(0.035)***	(0.037)***	(0.049)***
Total median income	0.482	0.449	0.463	0.473	0.481	0.486
	(0.010)**	(0.011)***	(0.007)***	(0.006)***	(0.007)***	(0.010)***
Neighborhood	0.024	0.024	0.020	0.025	0.019	0.018
	(0.008)**	(0.010)**	(0.006)***	(0.004)***	(0.006)***	(0.009)**
Population	0.015	−0.044	−0.006	0.013	0.031	0.045
	(0.006)**	(0.009)***	(0.006)	(0.004)***	(0.006)***	(0.005)***
Urbanization rate	0.141	0.116	0.070	0.091	0.081	0.175
	(0.040)**	(0.046)**	(0.028)**	(0.033)***	(0.030)***	(0.038)***
Population (0–15 years old)	−0.534	−0.304	−0.424	−0.315	−0.225	−0.385
	(0.153)**	(0.157)*	(0.104)***	(0.114)***	(0.177)	(0.207)*
Population density	−0.017	−0.019	−0.018	−0.012	−0.006	−0.008
	(0.005)**	(0.006)***	(0.005)***	(0.003)***	(0.003)**	(0.005)*
North dummy	−0.073	−0.129	−0.083	−0.049	−0.044	−0.069
	(0.134)**	(0.031)***	(0.025)***	(0.015)***	(0.014)***	(0.023)***
Northeast dummy	2.222	−0.091	−0.051	−0.024	−0.020	−0.021
	(0.027)**	(0.025)***	(0.025)**	(0.010)**	(0.012)*	(0.022)
South dummy	−0.036	−0.117	−0.118	−0.091	−0.078	−0.088
	(0.019)**	(0.021)***	(0.016)***	(0.012)***	(0.012)***	(0.017)***
Southeast dummy	0.027	0.008	0.002	0.020	0.017	−0.004
	(0.017)**	(0.015)	(0.014)	(0.009)**	(0.009)**	(0.016)
Constant	2.258	2.825	2.518	2.288	2.176	2.140
	(0.128)**	(0.166)***	(0.120)***	(0.079)***	(0.125)***	(0.157)***
Pseudo-R2	0.6000	0.3677	0.4336	0.4850	0.5170	0.5278
N	4815	4815	4815	4815	4815	4815

Obs: Standard errors in parentheses.
*** Means significant at 1% ($p < 0.01$).
** Means significant at 5% ($p < 0.05$).
* Means significant at 10% ($p < 0.1$).

The elasticity of the population share aged between 0 and 15 years has a negative sign and is statistically significant at 1% in the 0.25 and 0.50 quantiles and at 10% in the 0.10 and 0.90 quantiles. This suggests that a higher proportion of young individuals is associated with smaller per capita health expenditures per capita at the local level, as expected. Menezes et al. (2011) obtain the same result, allowing us to conclude that the larger this group, the smaller the economically active group and, therefore, the smaller the tax collection and spending.

The population density elasticity is negative and statistically significant at 1%, 5% and 10%, indicating that the higher the population density, the smaller the local public health expenditures (Mendes and Sousa, 2006; Sousa et al., 2005). The larger health care expenditures in municipalities with low population density may be explained by increasing returns to scale at the local level, especially in small-sized municipalities. As remarked by Mendes and Sousa (2006), a scattered population increases the average costs of public goods and services and does not allow smaller municipalities to explore the scale economies that characterize the production of these public goods and services. Consequently, small municipalities end up not using their funds efficiently. Thus, the negative sign of the elasticity of population density may be due to the existence of scale economies in the delivery of public goods (Menezes et al., 2011). On the other hand, a higher population density reduces the supply costs of public goods and services, thereby contributing to the reduction in local health care expenditures (Mendes and Sousa, 2006). The negative effect of population density decreases slightly across spending categories. Scale economies in Brazil seem to affect smaller municipalities slightly more, as they are unable to afford the fixed costs involved in the production of some local public goods and services. When municipalities reach a certain size, agglomeration benefits tend to disappear, while diseconomies of scale quickly increase (Mendes and Sousa, 2006).

Mendes and Sousa (2006) suggest a theoretical basis for this result from the use of congestion functions—crowding functions. According to this approach, the impact of population on per capita spending depends on public goods

Table 5
DEA and median voter model results.

	Population of less than 50,000 inhabitants	Population between 50,000 and 100,00 inhabitants	Population between 100,000 and 500,000 inhabitants	Population over 500,000 inhabitants	Total
"Excess" (Finbra—demand) US$ millions	(397.34)	(34.55)	(289.09)	(1095.82)	(1747.70)
Waste DEA—CRS US$ millions	3964.28	966.43	2665.54	1182.07	8778.28
Waste DEA—VRS US$ millions	2966.28	628.89	1209.85	328.49	5131.51
Average efficiency DEA—CRS	0.68	0.72	0.70	0.81	
Average efficiency DEA—VRS	0.76	0.82	0.86	0.89	
Total no. of municipalities	4919	324	243	37	5523

Data compiled by the authors, using average exchange rates in 2010 (R$/US$ = 1.7594).

production technology (or private property provided by the public sector). Due to the existence of scale economies, it is possible that the public per capita expenditure may be inversely related to population and population density. For Mendes and Sousa (2006), the suboptimal size of most Brazilian municipalities restricts the exploitation of scale economies that characterize public services production.

The elasticity of the urbanization rate is positive and statistically significant at 1% and 5%, indicating that the demand for health services is greater in municipalities with better infrastructure for the provision of public goods and services.

The results indicate that the median voter model seems to be adequate to estimate the demand for public health, in line with the current evidence for Brazilian municipalities. All the variables are statistically significant and present adequate signs. More importantly, the coefficient estimates are not statistically different across the quantiles, and we can therefore use any one of them to calculate the demand for public goods, regardless of the group we use to calculate the possible resource gains from an improvement in efficiency.[14]

6. Combining public health services supply and demand

Under the hypothesis that demand is correctly estimated, it is possible to assess whether there are imbalances between supply and demand in the delivery of health services and to check whether the eventual lacking resources can be made available through efficiency gains. Table 5 shows that the estimated demand is approximately US$ 1.7 billion higher than the actual expenditure in 2010. This implies that by taking into account the current efficiency levels, it would be necessary to have an additional expense of US$ 1.7 billion to meet the median voter's demand. The largest part of the demand takes place in municipalities with over 500,000 inhabitants. The results also show that for municipalities with 50,000–100,000 inhabitants, additional expenses would not be necessary, but there should be a redistribution of funds among these municipalities.

Nevertheless, taking into account the efficiency scores in the input-oriented DEA model, the current output level (IDSUS) could be obtained with considerably lower expenditures. According to the constant returns to scale (CRS) model, US$ 8.8 billion is wasted, while according to the variable returns to scale (VRS) model, approximately US$ 5.1 billion is wasted. In other words, public policies aimed at improving municipalities' efficiency in the use of health funds would produce estimated potential savings of between US$ 5.1 billion and US$ 8.8 billion. These funds would be more than enough to meet the excess demand for public goods and services in the health sector initially observed.

Some considerations are necessary, though. The amount of resources wasted here cannot be compared to those obtained by other Brazilian government agencies, such as the Federal Court of Accounts (Tribunal de Contas da União—TCU) or the Comptroller General of the Union (Controladoria Geral da União—CGU). The CGU, for example, only has information for decentralized programs (federal programs that are executed by municipalities) and for some municipalities (those chosen in the lottery). The waste of resources we obtain comes from the combination of supply

[14] Here, we use the estimated coefficients for the median. However, we also compute the demand and, consequently, the surplus/deficit of resources, using the coefficients for the other quantiles. The results are obviously only marginally different, as the estimates are not statistically different across quantiles, as mentioned before.

Table 6
Technical efficiency estimates and metatechnology ratios—DEA constant returns to scale.

Groups	No. of municipalities	Mean	Minimum	Maximum
Technical efficiency with respect to group frontiers				
Group 1—SUS	29	0.901	0.666	1.000
Group 2—SUS	94	0.830	0.617	1.000
Group 3—SUS	629	0.672	0.404	1.000
Group 4—SUS	587	0.724	0.456	1.000
Group 5—SUS	2029	0.685	0.363	1.000
Group 6—SUS	2183	0.689	0.347	1.000
Metatechnology ratios				
Group 1—SUS	29	0.605	0.469	0.633
Group 2—SUS	94	0.694	0.363	0.782
Group 3—SUS	629	0.877	0.462	0.920
Group 4—SUS	587	0.830	0.339	0.918
Group 5—SUS	2029	0.989	0.950	1.954
Group 6—SUS	2183	0.958	0.569	0.993
Technical efficiency with respect to the metafrontier				
Group 1—SUS	29	0.544	0.401	0.633
Group 2—SUS	94	0.575	0.363	0.782
Group 3—SUS	629	0.588	0.358	0.920
Group 4—SUS	587	0.600	0.339	0.918
Group 5—SUS	2029	0.678	0.363	1.000
Group 6—SUS	2183	0.660	0.315	0.993

Data compiled by the authors.

and demand for public health estimates and is subject to all restrictive assumptions associated with the methodological approaches we use.

7. Robustness check: alternative municipality groups

Because the supply results can vary according to the clustering criteria, we use the homogenous groups of municipalities defined by the Brazilian Ministry of Health to assess and compare the quality of health as a robustness check. The Ministry of Health classifies municipalities into six groups according to three indices: the socioeconomic development index (SDI), the health status index (HSI) and the health system structure index (HSSI). Each index consists of a set of indicators. SDI, for instance, consists of local GDP per capita, HSI includes infant mortality rate and HSSI includes the average ratio of primary health care and health surveillance professionals.[15]

Tables 6 and 7 show the technical efficiency estimates and metatechnology ratios based on DEA with constant returns to scale and variable returns to scale, respectively.

Group 6 has municipalities with low SDI, low HSI and no MHC (moderately to highly complex health services); group 5 has municipalities with average SDI, average HSI and no MHC; group 4 has municipalities with low SDI, low HSI and few MHC; group 3 has municipalities with average SDI, average HSI and few MHC; group 2 has municipalities with high SDI, average HSI and average MHC; and group 1 has municipalities high SDI, average HSI and many MHC.

Under the hypothesis of constant returns to scale, the efficiencies are higher regardless of the group. The results are quite consistent with those obtained when municipalities are classified according to their population size, which is unsurprising, as there is a large correlation between size and the indicators used by the Ministry of Health to form the groups, especially for groups located at the extremes (groups 5 and 6; group 1).

Table 8 shows, in the first row, how the estimated resource surplus (difference between the estimated demand and actual expenditures in 2010) is distributed across the new groups. The results indicate that groups 2, 4 and 6 do not

[15] For a detailed description of the indicators that made up the indices and of the method of aggregation, see http://portal.saude.gov.br/portal/saude/area.cfm?id_area=1080.

Table 7
Technical efficiency estimates and metatechnology ratios—DEA variable returns to scale.

Groups	No. of municipalities	Mean	Minimum	Maximum
Technical efficiency with respect to group frontiers				
Group 1—SUS	29	0.975	0.926	1.000
Group 2—SUS	94	0.938	0.867	1.000
Group 3—SUS	629	0.852	0.713	1.000
Group 4—SUS	587	0.829	0.665	1.000
Group 5—SUS	2029	0.828	0.638	1.000
Group 6—SUS	2183	0.790	0.603	1.000
Metatechnology ratios				
Group 1—SUS	29	0.610	0.540	0.651
Group 2—SUS	94	0.674	0.508	0.833
Group 3—SUS	629	0.775	0.583	0.921
Group 4—SUS	587	0.878	0.599	0.938
Group 5—SUS	2029	0.884	0.784	1.000
Group 6—SUS	2183	0.966	0.671	1.000
Technical efficiency with respect to the metafrontier				
Group 1—SUS	29	0.595	0.540	0.651
Group 2—SUS	94	0.633	0.508	0.833
Group 3—SUS	629	0.661	0.528	0.921
Group 4—SUS	587	0.727	0.599	0.938
Group 5—SUS	2029	0.733	0.545	1.000
Group 6—SUS	2183	0.762	0.602	1.000

Data compiled by the authors.

Table 8
DEA and median voter model results: homogenous groups.

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Total
"Excess" (Finbra—demand) US$ millions	(1181.69)	6.37	(817.02)	607.62	(451.74)	88.77	(1747.70)
Waste DEA—CRS US$ millions	1359.68	1428.50	3010.63	1258.43	1444.00	1191.07	9719.33
Waste DEA—VRS US$ millions	387.48	494.67	1397.12	847.35	783.50	805.23	4715.37
Average efficiency DEA—CRS	0.90	0.82	0.67	0.72	0.68	0.68	–
Average efficiency DEA—VRS	0.97	0.93	0.85	0.82	0.82	0.78	–
Total no. of municipalities	28	94	632	587	2038	2183	5562

Data compiled by the authors, using average exchange rate in 2010 (R$/US$ = 1.7594).

need additional funds, while groups 1, 3 and 5 have an excess demand and require additional resources to cover their gaps.

The second and third rows show the waste of resources, according to the DEA, with constant and variable returns to scale, respectively. A better use of resources would produce estimated potential savings of US$ 4.7 billion and US$ 9.7 billion. These funds would be more than enough to meet the excess demand for public goods and services in the health sector estimated by the median voter model. When groups were analyzed individually, an occasional fund shortage would be noticed only in group 1. In this group, municipalities provided moderately to highly complex services, luring patients from other municipalities and states. Thus, although they have a large capacity to generate their own funds, they may end up with a fiscal gap as a result of the remarkable demand for more-complex public health care services.

8. Conclusions and policy implications

The aim of this paper is to evaluate public health spending at the local level in order to ascertain whether it is necessary to increase the amount of resources available to local governments or, alternatively, if all it takes it is to guarantee that they use efficiently the resources they already have.

In fact, the paper has two complementary goals. The first is to assess the efficiency of health spending, and in case municipalities are inefficient, to determine savings when the municipalities' efficiency is enhanced. The second goal is to estimate the demand for health services and compare it with actual health spending. If there is a deficit (the demand is higher than the effective expenditure) and the savings initially estimated are enough to cover the gap, municipalities do not need additional resources, they only need to spend the resources they have in their hands more efficiently. They can provide the same amount of health services using a smaller amount of resources.

Unlike most studies that evaluate municipality efficiency in public provision (supply), we take into account their heterogeneity and use the metafrontier methodology. We follow O'Donnell et al. (2008) and calculate efficiency scores for all municipalities, taking into account the metafrontier; efficiency scores for each group of municipalities, taking into account that they are subject to different technologies (group frontiers); and metatechnology ratios, which determine how far group frontiers are from the metafrontier. We initially group municipalities according to their population size. Then, to check the robustness of the results, we also use the homogenous groups defined by the Brazilian Ministry of Health, based on municipalities' socioeconomic, health, and infrastructure conditions.

We estimate the demand for health services using an equation derived from the median voter model. The results indicate that most of the variables are statistically significant and have the expected sign. Using quantile regression, we show that the coefficient estimates are the same across the spending levels.

When we compare the estimated demand to actual health expenditures, we conclude that there is no gap (excess demand) for some groups of municipalities. For those groups with excess demand, however, saving of funds through the efficient management of local health expenditures is sufficient to meet the excess demand for public health goods and services.

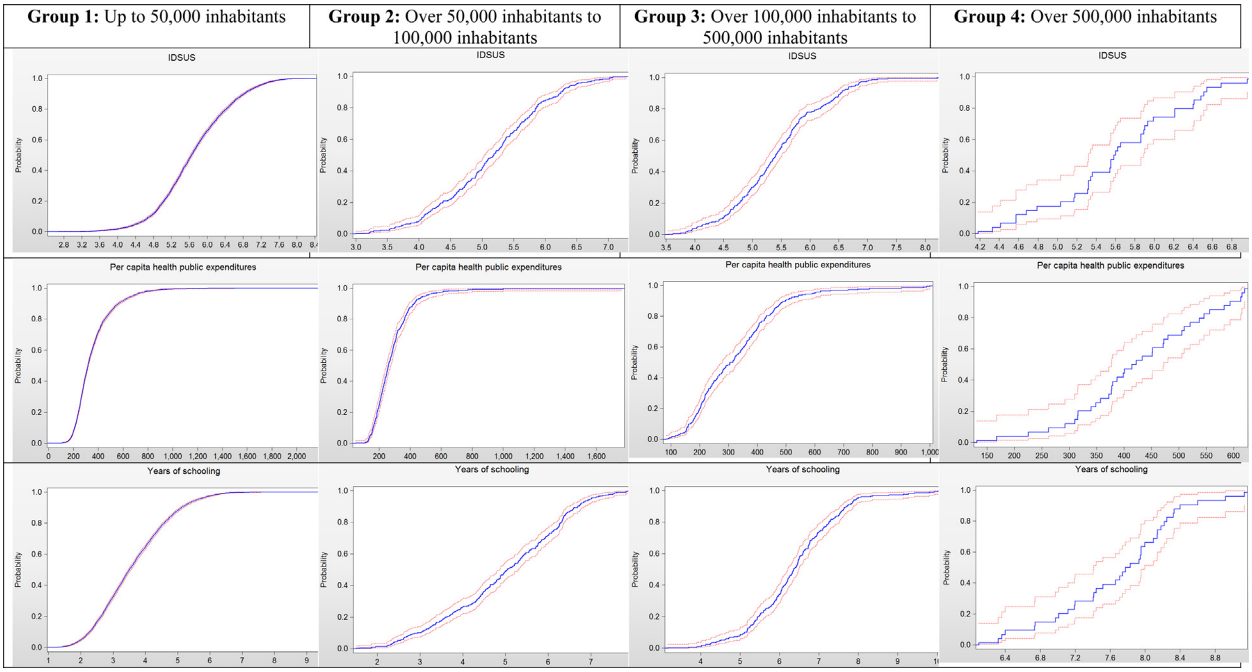
The results should be analyzed carefully. We use aggregate data at the municipal level, but it would be interesting to use disaggregated data involving hospitals in order to, for example, make detailed public policy statements. An example in this direction is the World Bank study about the Brazilian Unified Health System—SUS (Gragnolati et al., 2013). When discussing the role of efficiency in public spending, they argue that *"Greater public spending on health will undoubtedly help to fund more resources to the health system, such as facilities, equipment, human resources, medical supplies and services. However, there is also a need to introduce management techniques and systematic monitoring mechanisms for activities, so that an effort of this nature would not result in worsening system inefficiency indicatives. (...) the lack of resources and supplies, in many cases, is not the main obstacle to increase access and improve the quality of services. The health system could clearly produce better health services and results with the same level of resources by facing some inefficiency factors identified. For example, significant gains could be obtained with better alignment between hospital capacity and specialized medical services provision, by investing in the improvement of hospital technical efficiency and by reducing wastage and inappropriate use of resources, among other initiatives. In the other direction, there would also be gains if public spending distribution was prioritized based on more robust processes, through development and management of existing and new technologies. There are no simple solutions to deal with such complex issues, but there is a lot of international experience on these issues that could certainly benefit Brazil. At the same time, it is noted that even with efficiency increases, spending pressures will probably not decrease in the coming decades. As Brazil continues to grow and develop, the combination of unmet needs, both in primary and specialist care, the incorporation of new technologies and the growing demand for medical care associated with non-communicable diseases, in addition to the increasing demand for health services associated with an older population, will exert significant pressure on public health spending (p. 12–13)."*

Acknowledgments

The authors would like to thank the participants in the 40th Annual Meeting of the National (Brazilian) Association of Postgraduate Centers in Economics (ANPEC, 2012) for helpful discussion and comments. We would also like to thank an anonymous referee for careful reading of the manuscript and very helpful comments and suggestions that have served to improve the paper. The remaining errors are the sole responsibility of the authors. Opinions expressed in this paper are the responsibility of the authors and do not necessarily express the views and position of the Brazilian National Treasury Secretariat.

Appendix A. Histogram and empirical cumulative distribution function results

See Graph 1.



Graph 1. Empirical cumulative distribution function (CDF).
Data compiled by the authors. Quantile method: Rankit (Cleveland). Confidence interval: 0.95.

References

Afonso, A., St. Aubyn, M., 2005. Non-parametric approaches to education and health efficiency in OECD countries. J. Appl. Econ. 8 (2), 227–246.

Aronsson, T., Lundberg, J., Wikström, M., 2000. The impact of regional public expenditures on the local decision to spend. Reg. Sci. Urban Econ. 30 (2), 185–202.

Balaguer-Coll, M., Prior, D., Tortosa-Ausina, E., 2010. Output Complexity, Environmental Conditions and the Efficiency of Municipalities: A Robust Approach. Working Paper WP-EC 2010-02. Instituto Valenciano de Investigaciones Economicas, Available at: http://www.ivie.es/downloads/docs/wpasec/wpasec-2010-02.pdf (Accessed 28 February 2015).

Battese, G.E., Rao, D.S.P., 2002. Technology gap, efficiency, and a stochastic metafrontier function. Int. J. Bus. Econ. 1 (2), 87–93.

Battese, G.E., Rao, D.S.P., O'Donnell, C.J., 2004. A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies. J. Prod. Anal. 21 (1), 91–103.

Bergstrom, T.C., Goodman, R.P., 1973. Private demand for public goods. Am. Econ. Rev. 63 (3), 280–296.

Bergstrom, T.C., Rubinfeld, D.L., Shapiro, P., 1982. Micro-based estimates of demand functions for local school expenditures. Econometrica 50 (5), 1183–1205.

Brunet, J.F.G., Borges, C.B., Bertê, A.M.A., Bussato, L.M., Brunet, L., 2006. Estados comparados por funções do orçamento: uma avaliação da eficiência e efetividade dos gastos públicos estaduais. In: Prêmio IPEA-Caixa. Menção Honrosa I. Brasília. Available at: http://www.esaf.fazenda.gov.br/premios/premios-1/premios-realizados/pasta-premio-ipea-caixa/monografias-premiadas. (Accessed 28 February 2015).

Congleton, R.D., Bennet, R.W., 1995. On the political economy of state highway expenditures: some evidence of the relative performance of alternative public choice models. Public Choice 84, 1–24.

Congleton, R.D., Shughart, W.F., 1990. The growth of social security: electoral push or political pull? Econ. Inq. 28, 109–132.

Dahlberg, M., Johansson, E., 2000. An examination of the dynamic behavior of local governments using GMM bootstrapping methods. J. Appl. Econom. 15 (4), 401–416.

Evans, D.B., Tandom, A., Murray, C.J.L., Lauer, J.A., 2000. The Comparative Efficiency of National Health Systems in Producing Health: An Analysis of 191 Countries. GPE Discussion Paper Series 29. World Health Organization, Geneva, Available at: http://www.who.int/healthinfo/paper29.pdf. (Accessed 28 February 2015).

Gragnolati, M., Lindelow, M., Couttolenc, B., 2013. Twenty Years of Health System Reform in Brazil: An Assessment of the Sistema Único de Saúde. Directions in Development. World Bank, Washington, DC, http://dx.doi.org/10.1596/978-0-8213-9843-2 (License: Creative Commons Attribution CC BY 3.0).

Gravelle, H., Jacobs, R., Jones, M.A., Street, A., 2003. Comparing the efficiency of national health systems: a sensitivity analysis of the WHO approach. Appl. Health Econ. Health Policy 2 (3), 1–7.

Gupta, S., Verhoeven, M., Tiongson, E.R., 2002. The effectiveness of government spending on education and health care in developing and transition economies. Eur. J. Political Econ. 18 (4), 717–737.

Hollihgsworth, B., Wildman, J., 2003. The efficiency of health production: re-estimating the WHO panel data using parametric and non-parametric approaches to provide additional information. Health Econ. 12 (6), 493–504.

Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46 (1), 33–50, Available at: http://web.stanford.edu/~doubleh/otherpapers/koenker.pdf. (Accessed 28 February 2015).

Koenker, R., 2005. Quantile Regression. Cambridge University Press, New York.

Mirmirani, S., Li, H.C., Ilacqua, J.A., 2008. Health care efficiency in transition economies: an application of data envelopment analysis. Int. Bus. Econ. Res. J. 7 (February (2)).

Marinho, A., 2003. Avaliação da eficiência técnica nos serviços de saúde nos municípios do Estado do Rio de Janeiro. Rev. Bras. Econ. 57 (3), 515–534, Available at: http://www.scielo.br/pdf/rbe/v57n3/a02v57n3. pdf. (Accessed 28 February 2015).

Mattos, E., Rocha, F., Novaes, L., Arvate, P., Orellano, V., 2009. Economias de escala na oferta de serviços públicos de saúde: um estudo para os municípios paulistas. Economia 10 (2), 357–386.

Médici, A., 2011. Propostas para melhorar a cobertura, a eficiência e a qualidade no setor saúde. In: Bacha, E., Simon, S. (Eds.), Brasil: A Nova Agenda Social. LTC, Rio de Janeiro, pp. 23–93, Available at: http://www.schwartzman.org.br/simon/agenda1.pdf. (Accessed 28 February 2015).

Mendes, C.C., Sousa, M.C.S., 2006. Demand for locally provided public services within the median voter's framework: the case of Brazilian municipalities. Appl. Econ. 38 (3), 239–251.

Menezes, R.T., Saiani, C.C., Zoghbi, A.C.P., 2011. Demanda mediana por serviços públicos e desempenho eleitoral: evidências do modelo do eleitor mediano para os municípios brasileiros. Estud. Econ. 41 (1), 25–57, Available at: http://www.revistas.usp.br/ee/article/view/36034/38752. (Accessed 28 February 2015).

Navarro, V., 2000. Assessment of the World Health Report 2000. Lancet 356, 1598–1601, Available at: http://www.eurohex.eu/bibliography/pdf/0857844774/Navarro_2001_Lancet.pdf. (Accessed 28 February 2015).

O'Donnell, C.J., Rao, D.S.P., Battese, G.E., 2008. Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. Empir. Econ. 34 (2), 231–255.

Simar, L., 2003. Detecting outliers in frontier models: a simple approach. J. Prod. Anal. 20, 391–424, Available at: http://www.ersa.org/IMG/pdf/Simar_03-Detecting-Outliers-in-Frontier-Models–A-Simple-Approach.pdf. (Accessed 28 February 2015).

Sousa, M.C.S., Stosic, B.D., 2003. Technical efficiency of the Brazilian municipalities: correcting non-parametric frontier measurements for outliers. J. Prod. Anal. 24 (2), 157–181.

Sousa, M.C.S., Cribari-Neto, F., Stosic, D., 2005. Explaining DEA technical efficiency scores in an outlier corrected environment: the case of public services in Brazilian municipalities. Braz. Rev. Econom. 25 (2), 287–313, Available at: http://www.sbe.org.br/dated/2005_nov06.pdf. (Accessed 28 February 2015).

Souza, I.V., Nishijima, M., Rocha, F., 2010. Eficiência do setor hospitalar nos municípios paulistas. Econ. Apl. 14 (1), 51–66, Available at: http://www.scielo.br/pdf/ecoa/v14n1/a04v14n1.pdf. (Accessed 28 February 2015).

Videira, R.A., Mattos, E., 2011. Ciclos politicos eleitorais e a interação especial de políticas fiscais entre os municípios brasileiros. Economia Aplicada, Ribeirão Preto ? SP 15 (2), 259–286.

Williams, A., 2001. Science of marketing at WHO? A commentary on World Health 2000. Health Econ. 10 (2), 93–100.

Wilson, P.W., 1993. Detecting influential observations in deterministic non-parametric frontiers models with multiple outputs. J. Bus. Econ. Stat. 19 (3), 319–323.

Wilson, P.W., 1995. Detecting influential observations in data envelopment analysis. J. Prod. Anal. 6 (1), 27–45.

Wilson, P.W., 2008. FEAR 1.0: a software package for frontier efficiency analysis with R. Socio-Econ. Plann. Sci. 42, 247–254.