

Local government size and service level provision. Evidence from conditional non-parametric analysis.

Giovanna D’Inverno^{a,*}, Wim Moesen^a, Kristof De Witte^{a,b}

^a*KU Leuven, Faculty of Business and Economics (FEB), Leuven, Belgium*

^b*Maastricht University, UNU-MERIT, Maastricht, The Netherlands*

Abstract

Although the local provision of public goods accommodates better to the heterogeneous local preferences and mitigates the fiscal illusion problem, it comes at the cost of potential diseconomies of scale. This paper examines the relationship between municipal size and local service level provision by applying state-of-the-art non-parametric techniques to a unique panel dataset of Flemish data. We measure the service provision level by using an innovative robust conditional ‘Benefit of the Doubt’ model and we estimate its efficiency in relationship with the local expenditures by means of a robust conditional Data Envelopment Analysis model. Overall, the main findings suggest the presence of diseconomies of scale, and provide weak evidence on an optimal size of local public good provision of around 10,000 citizens.

Keywords: Efficiency analysis; Municipal mergers; Municipal expenditure; Local governments; Composite indicator.

JEL Codes: H50; D61; H21

1. Introduction

In the public finance literature on multi-level government, it is well established that local authorities should focus on the provision of local public goods¹ (King, 2016) as they can accommodate to the heterogeneous local preferences and as local provision mitigates the fiscal illusion

*Corresponding author.

Email addresses: giovanna.dinverno@kuleuven.be (Giovanna D’Inverno), wim.moesen@kuleuven.be (Wim Moesen), kristof.dewitte@kuleuven.be (Kristof De Witte)

We gratefully acknowledge insightful comments and suggestions from Chris O’Donnell, Knox Lovell, Ralph Fre, Tom Van Puyenbroeck, Antonio Peyrache, Andreas Themelis, three anonymous referees and seminar participants at University of Leuven, The University of Queensland, and the ‘7th International Workshop on Efficiency in Education, Health and other Public Services’. Giovanna D’Inverno gratefully acknowledges financial support from Research Foundation – Flanders, FWO (Postdoctoral Fellowship 12U0219N). Kristof De Witte acknowledges financial support from Research Foundation – Flanders, FWO (grant G068518N).

¹The local public good or service is defined by the limited size of the benefit region. Typical examples include playgrounds, kindergartens, fire protection, street lighting... This contrasts to national public goods for which the benefit area covers the whole nation (e.g. defense, fundamental research...).

problem. Despite the benefits, local provision comes at the cost of potential diseconomies of scale. This paper provides novel evidence on one of the key-questions in the public finance literature by examining the relationship between local provision and scale. Specifically, we evaluate the municipal performance both in terms of efficiency and in terms of effectiveness.² The former investigates how municipalities use their resources to provide local services and whether they could provide the same level of services with fewer resources. The latter explores to what extent the municipalities meet the citizens need for local services, beyond the available amount of resources (Mbuvi et al., 2012). These two aspects are complementary and together they provide a more comprehensive picture of the municipal performance to be further examined in the light of the scale economies debate.

To do so, we apply state-of-the-art non-parametric techniques to Flemish data. We adapt the novel indicator of D’Inverno and De Witte (2020) to measure the municipal service level provision. Relying on the popular Benefit of the Doubt model (Melyn and Moesen, 1991), the service level indicator explicitly models the heterogeneous preferences of citizens in the provision of e.g. education, transportation, and social and recreational services among others. The model, which is rooted in the approach suggested by Zanella et al. (2015), is sufficiently flexible to capture differences between desirable (e.g. education) and undesirable (e.g. traffic accidents) dimensions of local public good provision. Moreover, to acknowledge the broad range of functions of municipalities, we include assurance region type I weight restrictions that mimic the relative importance of municipal areas in percentage terms using information from the actual budget allocation. Thanks to the insights from robust conditional efficiency analysis (Daraio and Simar, 2005; Simar and Vanhems, 2012; De Witte and Kortelainen, 2013; Daraio and Simar, 2014; Fusco et al., 2019), we immediately control in the construction of the composite indicator for observed heterogeneity coming from contextual elements such as fiscal income, financial debt or unemployment. As such, the environmental context that municipalities are facing, can partly explain their choices in the service level provision. Next, to examine the existence of (dis)economies of scale and to analyze the municipal (cost-)efficiency, using a robust conditional Data Envelopment analysis model, we correlate the estimated service composite indicator to proxies of municipal size, e.g., number of inhabitants, and municipal expenditures per capita.

Our application focuses on Flanders, the northern region of Belgium. Flanders provides an interesting setting as it supplies high quality data on local public services. The issue of the optimal size of municipalities was heavily debated and eventually altered during the last 40 years. Moreover, with a few exceptions, the current size of Flemish municipalities ranges between 5,000

² Cherchye et al. (2019) provide a simple diagnostic tool to compute performance/productivity ratios.

and 50,000 inhabitants, which corresponds to the situation in many other countries (OECD, 2018). This fosters the external empirical validation of our findings.³ Finally, an earlier compulsory municipal amalgamation took place under the principle of a budgetary neutrality, which might encourage the existence of economies of scale.

In sum, the present paper contributes to the public finance literature by estimating local service level provision and linking it to one of the most prominent public finance questions: Is there an optimal scale size for the public good provision? Simultaneously, we contribute to the operations research literature by applying novel non-parametric model specifications to local government efficiency, which is one of the oldest applications in the Data Envelopment Analysis literature (see Section 2). Overall, our results suggest the presence of diseconomies of scale in the local public good provision. More in detail, we provide weak evidence for the presence of an optimal size of local public good provision of around 10,000 citizens.

The remainder of the paper is structured as follows. Section 2 overviews the literature on the optimal size of local authorities. Section 3 elaborates on the push- and pull-experiences in municipal amalgamations in Belgium over the last 40 years. Section 4 presents the applied model to estimate local service level provision and its efficiency, while the data are outlined in section 5. Section 6 discusses the results. Section 7 concludes the paper.

2. Literature review

Advocates for a territorial consolidation of municipalities often rely on scale economies to argue their case. A first subsection overviews the suggested mechanisms on scale (dis)economies in local public good provision. A second subsection discusses the concept and measurement of cost-efficiency.

2.1. Mechanisms for (dis)economies of scale

Similar, but not identical, to industrial production, one expects scale economies in the services delivered by a municipality. Provision of services for a larger community may then be cheaper. When fixed costs are substantial compared to variable costs, economies of scale become more prominent as the average cost per capita decreases. This may lead to the familiar U-shape average cost curve with a minimum point located at the optimal population size (Byrnes and Dollery, 2002). In the same line, Fox and Gurley (2006) argue that scale economies are service specific, and are more likely for infrastructure intensive services such as water and sewerage.

³ To foster applications to different settings, the code is available upon request.

The diseconomies of scale might be explained by two mechanisms. First, larger municipalities deliver public services to their own residents but also to the consumers of the neighbouring autonomous municipalities, i.e. public service spillovers. Typical examples are schooling, health care, cultural events, sport facilities. To the extent that the non-residents are not charged with a fee, they free ride on the provision by the taller center-municipality. Their centrality-function is then expected to put on upward pressure on the spending of the taller municipalities (Fox and Gurley, 2006).

A second underlying mechanism for the more than proportional increase in expenditures in larger municipalities is related to bureaucratic entropy and rent seeking by the local policymakers. As the population size increases more financial resources enable the local governments to increase the number of their administrative employees. This common observation in larger institutions is more visible when flat hierarchy structures tend to get out of date. Intermediate levels with coordinating and controlling functions became then affordable. However, the distance between the front-services and the back offices become also more remote. This upward pressure on local spending can be enhanced by rentseeking behavior of the leading policymakers: both politicians and the higher administrative staff (Oates, 1972). In the context of public budgeting the pivotal decision makers may be tempted to use their discretionary power to privilege the spending-programs of a narrow interest group, lobby or firm. A case in point is urban planning and real estate transactions (De Witte et al., 2018).

2.2. Municipal cost-efficiency

The shape of the cost-curve offers an instructive, but incomplete, picture of the merger debate. The coin has two sides. Not only the input-side (costs) but also the output-side should be taken into account (Cherchye et al., 2019). Standard production function theory relates output to input. A well-known example is the Cobb-Douglas production function, which relates one output to two inputs: labor and capital. A rich variety of parametric, non-parametric and semi-parametric specifications has followed since, and have been applied to local public good provision.

A popular toolbox to analyse municipal (cost-)efficiency is rooted in parametric regression analysis and in non-parametric best practice frontiers. Data envelopment analysis (DEA) paved the way (Charnes et al., 1978; Banker et al., 1984), followed by the Free Disposal Hull (FDH) variant of Deprins et al. (1984). As a parametric exercise Moesen and Vanneste (1983) conducted an early, and in retrospect an elementary, empirical test in 1983 for the Flemish municipalities. They splitted the municipalities into two samples: the small municipalities with a population size below 15,000 inhabitants and the tall municipalities exceeding 15,000 inhabitants up to 100,000. The elasticity of total current expenditures vis--vis population size was 0.95 for the smaller municipalities close to 1, this means that total expenditures are proportional to population size.

This implies that average cost (spending) per capita roughly remains constant over that interval. Examining the larger municipalities, they found an elasticity coefficient of 1.36, suggesting that with a population increase of say 10 percent, the total costs would increase with 13.6 percent, i.e. more than proportional. Bluntly stated: Moesen and Vanneste (1983) observed diseconomies of scale in the Flemish municipalities.

Later studies favored the non-parametric approach. Vanden Eeckaut et al. (1993) study the set of 235 municipalities in Wallonia, the French-speaking part of Belgium. Total costs are taken from the municipal accounts (current expenditures) and figure as the input-variable. The output side is covered with 6 indicators (data of 1986). First, total population reflects the basic administrative services provided to the residents (e.g. maintaining the register of births, marriages, and deaths, issuing identity cards, passports and other certificates). Second, they considered the length of roads to be maintained by each municipality. Third, the number of senior citizens (aged 65 and more) reflects the local supply of services like retirement homes, medical aid in public hospitals. Fourth, the number of beneficiaries of subsistence aid provide an indicator of local poverty. Fifth, the number of crimes registered in the municipality services as a proxy for police services. Finally, they include the number of students enrolled in local primary schools. Applying the FDH approach, Vanden Eeckaut et al. (1993) calculate a numerical performance score for each municipality. Next, they regressed this score against other economic and political variables, including returns-to-scale. They observe (p. 314) the presence of diseconomies of scale: a given proportional increase in all the (output) indicators would induce an increase of local expenditures in higher proportion.

In another study the same FDH-approach was applied for the total set of 589 Belgian municipalities with data of 1985 (De Borger et al., 1994). Among the output variables, De Borger et al. (1994) include a proxy variable for the public service spillovers. This variable, which can be thought of as a centrality index, was defined as the logarithm of the number of non-residents working in the municipality divided by the logarithm of total employment in the municipality. One expects, *ceteris paribus*, more input resources to be necessary in case the services provided by e.g. the road system are being used intensively by non-residents. However, taking into account explicitly the centrality-function of some municipalities may obscure the calculation of the proper scores for productive efficiency. In fact, it turned out that in this study the scores were positively related to the population size.

Geys and Moesen (2009) calculate public performance scores for the Flemish municipalities with data of 2000, using the stochastic frontier analysis. The performance scores are then related to several explanatory variables as income, education, home ownership and debt level. For the present paper one explanatory variable is of particular interest. Population size affects the per-

formance scores significantly and with a negative sign. Larger municipalities seem to suffer from an inferior performance than smaller ones, *ceteris paribus*.

Outside Belgium, theoretical models on the optimal municipal size suggest economies of scale for smaller municipalities, while diseconomies of scale for larger municipalities (Breunig and Rocaboy, 2008).⁴ However, the evidence on a similar U-shaped pattern is mixed. Although, Breunig and Rocaboy (2008) in France, Solé-Ollé and Bosch (2005) and Hortas-Rico and Rios (2019) in Spain, Agasisti and Porcelli (2019) in Italy, and Reingewertz (2012) in Israel observe economies of scale until 400 to 10,000 inhabitants, Holcombe and Williams (2009) do not observe significant correlation between local public good provision and scale. Overall, based on a review of about 34 papers, Byrnes and Dollery (2002, p.393) state that: “overall, 29 per cent of the research papers find evidence of U-shaped cost curves, 39 per cent find no statistical relationship between per capita expenditure and size, 8 per cent find economies of scale, and 24 per cent find diseconomies of scale.”

3. Context: Experiences with municipal mergers in Belgium

In 1977 a major compulsory territorial reform was implemented in Belgium. In one stroke some 2,500 municipalities were regrouped into some 600 new municipalities, most of them having now a minimum size of 10,000 inhabitants. This section documents the nature and extents of municipal amalgamation in Belgium since 1977.

3.1. The Push-Reform in Belgium (1977)

Objectives

It is not always easy to discern the true objectives of a political decision of the past. Fortunately, the parliamentary memorandum, accompanying the reform legislation, provides some useful information at least about the ‘stated’ objectives. One statement is particularly instructive: “To provide to the citizens the services which they may require, the municipalities should have adequate financial resources and qualified personnel, together with a minimum population size and geographical surface”.

Clearly the main aim was to achieve better service levels in municipal provision with a more efficient use of available resources. A better service level is a rather vague notion, referring to quality aspects. For the small (rural) municipalities it was widely accepted that service levels were inappropriate. For example, office hours were limited and even senior employees worked only on a part-time basis. Specialization and division of labor seldom occurred. It was expected that

⁴ Tavares (2018) reviews the literature on the consequences of municipal mergers.

a major merger operation would broaden the tax base, supplying sufficient resources to attract full-time qualified personnel. For the cities and larger municipalities the municipal reform aimed to internalize the public service spillovers, which allow non-residents to free-ride on the services of the central municipality. It is clear that for the small municipalities the efficiency motive was dominant, whereas for the larger municipalities more equity in burden sharing was to be achieved.

The 1977 reform aimed at a minimum population and geographical size. Several conjectures point out that (implicitly) a minimum size of 5,000 inhabitants and a surface of 30 a 40 square kilometers was acknowledged, whereas the target level was 10,000–20,000 inhabitants. The rationale behind these targets was primarily based on statistical observations. Data on municipal expenditures before the reform reveal that in the very small municipalities, with less than 1,000 inhabitants, the per capita expenditures exceeded the level of municipalities with 5,000 inhabitants and almost reached the same level as the municipalities with 10,000 to 20,000 inhabitants. If actual expenditures are seen as a proxy for costs then a U-shaped cost curve could be observed.

Mapping the results of the merger operation

In the reform, each municipality was to be evaluated on its own needs and merits. As a political decision process three characteristics can be distinguished: the municipal reform was (i) comprehensive, (ii) coercive and (iii) had to be realized in the short run. The period 1974–1977 was crucial for the development of the idea of a nation-wide municipal reform. Before that time a few municipalities had merged on a purely spontaneous basis and without much coherence. The national plan had to be elaborated and implemented by January 1, 1977.

Let us now briefly review how the reform was realized. Out of 2,663 existing municipalities 596 new configurations were formed in 1977, which were further to be reduced to 589 municipalities in 1983, after the postponed amalgamation of the Antwerp metropolitan area. This is a quite severe cut to 20 percent or 1/5 of the previous municipalities. As a consequence, the new municipality regroups on average 4 to 5 old municipalities. The average population size increased to 16,700 inhabitants per municipality and a surface of 52 square kilometers. Before the reform there were 1,903 municipalities with less than 5,000 inhabitants, i.e. 4/5 of all the municipalities. After the reform only 105 municipalities still showed this little size. This is less than 1/5 of the new municipalities. Undoubtedly a substantial enlargement of the municipal scale had been achieved.

The hypothesis of budgetary neutrality

The municipal amalgamations took place under the principle of a budgetary neutrality. By law, during the first three fiscal years after the merger the total number of municipal employees in each municipality should not exceed the sum of the employees of the member municipalities. In the national parliament the politicians seemed to agree that the introduction of an upper

ceiling on personnel would be helpful to restrict excessive appointments after the reform.

In terms of budgetary outlays no quantitative ceiling was formulated, although the objective of budgetary neutrality was rather explicit. To examine to what extent this objective was realized, we calculate the total current expenditures in constant (of 1971) and real prices in a four year period before and after the reform. In the period before the reform (1972-1976) the average growth rate in local public expenditures was 5.54 percent a year, while in the period after the reform (1977-1981), the growth rate was 5.66 percent. This minor difference is ambiguous: small enough to avoid an ‘explosion’ of spending but also not significant enough to question the economies of scale economies in local provision at the aggregate level.

3.2. The merger of Antwerp (1983)

Antwerp, the biggest city in the region of Flanders, was allowed to delay the implementation of its merger until 1983. This historic city, enjoying a prosperous port, performs as an economic pole of attraction for the whole ‘hinterland’. It was decided that the city itself should merge with 7 neighbouring municipalities, each on their own endowed with a large population. This was an ambitious project which required more time to implement. As it stands now, Antwerp hosts half a million inhabitants, the highest number for any municipality in Belgium. In fact, later on (in 2001) Antwerp adopted an inner redesign with districts. The purpose was and still is to make that huge size more manageable.

3.3. The voluntary mergers of 2019

After a relatively stable period between 1983 and 2010, the Flemish government launched an incentive scheme for voluntary mergers, to be implemented in 2013. Contrary to expectations, only few municipalities showed a weak interest and none was realized.

In 2016 a new decree was enacted, favouring bottom-up mergers with a bonus of 500 euro per inhabitant. This cash hand-out intended to reduce the outstanding municipal debt. This round was somewhat more successful. Some 15 municipalities engaged for a merger into 7 new entities. Out of this 7 amalgamation, 6 formed a couple with 1 other municipality. Nonetheless the average size increased from 17,000 residents before to 27,950 after the merger, which had to be implemented by January 2019.

4. Methodology

Taking into account the insights and experiences from previous research a first subsection presents the approach to construct a composite index for service level provision. A second subsection presents the robust conditional Data Envelopment Analysis model to measure the adequate use of financial resources in terms of efficiency.

4.1. A comprehensive municipal service provision indicator

When it comes to the evaluation of the municipal service provision as a whole, there are several aspects that need to be considered for benchmarking purposes. First of all, we have to tackle its multidimensionality. From daily life we can see that municipalities offer a number of different services in different areas of intervention. From country to country the local competences might change, but certainly they span a wide range of basic services, such as educational and care services, public safety and general administration to name a few. Each municipality has different preferences over these activities both from the demand side and institutional supply. On the one hand, the municipal service provision might be influenced by political historical choices taken in the past or by current preferences represented in the municipal councils. For example, a municipality that invested for years in a certain area might not be able to dramatically change its specialization all of sudden. Or a more left-wing government in power might be more inclined into expanding the public sector than a right-wing (Hortas-Rico and Rios, 2019). On the other hand, the citizens' needs might differ from a municipality to another one, e.g. due to differences in the demographic profile, urging the provision of certain services rather others. As a result, municipalities differ in the type and the quantity of services they provide and a fair assessment should take into account their operating context, such as the economic and financial characteristics among others.

D'Inverno and De Witte (2020) suggest the use of a flexible composite indicator for the municipal service level provision evaluation along different municipal areas, while accounting for different sources of municipal heterogeneity. The main idea behind the construction of this synthetic index is to aggregate several sub-indicators that represent municipal services into a comprehensive indicator that measures the overall service provision level as a whole. Each item enters in the composite indicator according to a data-driven weighting scheme following insights from the 'Benefit-of-the-Doubt' (BoD) approach (Melyn and Moesen, 1991; Cherchye et al., 2007), so that more importance (higher weight) is assigned to what a municipality undertakes to provide the most and less importance in the opposite case. This composite indicator is flexible in so far as it can overcome several shortcomings that would arise in the municipal service provision assessment if using a BoD model in its basic version: 1) the presence of undesirable indicators, 2) the influence of atypical observations, 3) the role of contextual factors, 4) the inclusion of the actual municipal budget allocation across the municipal functions.

As for the first issue, the sub-indicator may have either a desirable or an undesirable connotation. In the first case, the more the municipality provides the better it is (e.g., cultural events). In the second case, the more the municipality restricts the sub-indicator the better it is (e.g., the level of crime). To encompass both desirable and undesirable sub-indicators, the model builds on

a directional distance function as introduced by Zanella et al. (2015, models 7 and 8 p.523) and recently used by other scholars (see for example Rogge et al., 2017; Fusco et al., 2019; Lavigne et al., 2019). As for the second issue, the deterministic nature of the basic BoD model makes each municipality in the sample constitute the performance possibility set. As a consequence, the presence of any municipality with atypical or outlying indicators would affect the estimation, and bias the service provision composite indicator scores. Another source of bias in the service provision evaluation comes from the third issue, namely when benchmarking the municipalities without properly taking into account the background conditions they have to operate in. This would be in addition to the lack of fairness in the comparison procedure, being the municipality judged against others possibly operating in very different conditions (e.g., from an economic perspective).

To mitigate the influence of atypical observations and to include characteristics of the operational environment, the robust conditional directional distance version of the “Benefit of the Doubt” (BoD) model is considered, combining insights introduced by Cazals et al. (2002); Daraio and Simar (2007b); De Witte and Kortelainen (2013), extended to the directional distance framework by Simar and Vanhems (2012); Daraio and Simar (2014) and adapted for the directional distance BoD composite indicator by Fusco et al. (2019). With this model specification, a municipality’s service provision is not evaluated against the service provision of all the other municipalities in the sample, but rather against municipalities with equal or better performance and with similar background conditions or, put another way, *against the expectation of the “best”* (Daraio et al., 2020). Following Fusco et al. (2019) and Lavigne et al. (2019), we can formally define the performance possibility set for the robust conditional BoD model in its directional distance version as follows:

$$\tilde{\Psi}_{m_z} = \bigcup_{j=1}^m \{(\mathbb{1}, y^+, y^-) \in \mathbb{R}_+^{1+d+u} | Y_{j,Z=z}^+ \geq y^+, Y_{j,Z=z}^- \leq y^-\} \quad (1)$$

where d is the number of desirable indicators y^+ and u the number of undesirable indicators y^- , m is the number of j peers the municipality under evaluation is benchmarked against and $\mathbb{1}$ denotes the dummy input equal to one for all observations.

The composite indicator is computed by a Monte-Carlo procedure. For a high number of times B the municipality is compared with a sub-sample of $m < n$ observations drawn with replacement (in the present case, $m=40$ and $B=1000$). In particular, due to the conditional approach, units with background characteristics more similar to the ones of the unit under assessment, $Z = z$, are more likely to be drawn. In each iteration b the municipal service provision score is computed

by solving the convex relaxation of the following problem:

$$\begin{aligned} \tilde{D}_{m_z}^b(\mathbb{1}, y^+, y^-; g^+, g^-) = \sup\{\beta | (\mathbb{1}, y^+ + \beta g^+, y^- - \beta g^-) \in \tilde{\Psi}_{m_z}\}, \\ \forall b = 1, \dots, B \end{aligned} \quad (2)$$

where $g = (g^+, -g^-) = (y^+, -y^-)$ is the direction along which desirable features are expanded and undesirable features contracted simultaneously and it is specified as the municipal indicators of the evaluated municipality; β is the optimal value in terms of maximal expansion and contraction. The associated municipal service provision composite indicator score is obtained as

$$CI_{m_z}^b = 1/(1 + \tilde{D}_{m_z}^b). \quad (3)$$

The robust conditional composite indicator is obtained as the mean of the Monte-Carlo realizations:

$$CI_{m_z} = \mathbb{E}[CI_{m_z}^b] \quad (4)$$

The composite indicator score CI_{m_z} normally ranges between 0 and 1, where 1 denotes the greatest level of municipal service provision. Since the unit under assessment might not be included in the sub-sampling, CI_{m_z} can be greater than 1. In this case the municipality will be considered super-performing with respect to the average m municipalities it is compared with.

In addition to this, the basic version of the BoD model might assign in the construction of the composite indicator zero weights to the municipal areas where they perform the worst or provide the least level of services. We use the assurance region type I (ARI) weight restrictions to avoid this issue and mostly to incorporate the relative importance of municipal areas in percentage terms using actual information from the budget allocation. We consider in place of Eq. (2) the multiplier formulation and add the weight restrictions' specification, as follows:

$$\begin{aligned} \phi_r \leq \frac{w_r \bar{y}_r^+}{\sum_{r=1}^d w_r \bar{y}_r^+ + \sum_{k=1}^u p_k \bar{y}_k^-} \leq \psi_r \quad \text{for } r = 1, \dots, d \\ \phi_k \leq \frac{p_k \bar{y}_k^-}{\sum_{r=1}^d w_r \bar{y}_r^+ + \sum_{k=1}^u p_k \bar{y}_k^-} \leq \psi_k \quad \text{for } k = 1, \dots, u \end{aligned} \quad (5)$$

where r refers to the d indicators linked to desirable features and k to the u undesirable ones (in the present application, $d=5$ and $u=3$, each of them corresponding to a municipal area). ϕ and ψ are the lower and upper bounds and they reflect the relative importance range for each municipal area. D'Inverno and De Witte (2020) link this relative importance to the municipal expenditure shares and discuss three alternative specifications. For the present application, we consider the case where weights are allowed to range $\pm 50\%$ of the average spending share in each municipal area. Appendix A reports the information on the average observed budget shares together with their computed lower and upper bounds. w_r and p_k are the weights endogenously obtained for

each municipal area solving the problem. The derived optimal weights can be seen as shadow prices and can be used to compute the optimal expenditure shares for the municipal functions under analysis.

4.2. Robust Conditional Data Envelopment Analysis

In line with the non-parametric model formulation used for the municipal composite indicator estimation in the previous subsection, we measure the efficiency of local expenditure using a Data Envelopment Analysis (DEA) approach (Charnes et al., 1978; Banker et al., 1984). For each municipality j_0 under analysis, the following input-oriented model with Variable Return to Scale has to be solved:

$$\hat{\theta}_{DEA}(x_0, y_0) = \min\{\theta | y_0 \leq \sum_{j=1}^n \lambda_j Y_j; \theta x_0 \geq \sum_{j=1}^n \lambda_j X_j; \theta > 0; \sum_{j=1}^n \lambda_j = 1; \lambda_j \geq 0; j = 1, \dots, n\} \quad (6)$$

where x_0 and y_0 are respectively the inputs and the outputs of the municipality j_0 under assessment, X_j and Y_j are the inputs and the outputs of the observed n municipalities in the reference sample and λ_j is the intensity variable. θ is the input efficiency score and it can range between 0 and 1, where 1 denotes that the unit constitutes the best practice frontier, otherwise $1 - \theta$ measures the potential reduction in the use of inputs to provide the same service levels. As in the case of the composite indicator, the deterministic nature of DEA makes it sensitive to atypical observations (outliers or measurement errors). As argued for the composite indicator, *mutatis mutandis*, the robust version of the DEA model is considered, following insights from Cazals et al. (2002), Daraio and Simar (2007b) and De Witte and Kortelainen (2013). In this case, $m < n$ municipalities are randomly drawn with replacement among those observations producing at least the same level of output as the municipality j_0 under evaluation. If the unit under analysis does not constitute its own reference set in the drawing, then the efficiency score will be greater than 1. In this case the municipality will be considered super-efficient with respect to the average m municipalities it is compared with. Moreover, to further explore the municipal size aspect, we consider the conditional version of the robust DEA estimator by conditioning the drawing on the municipal size (i.e. the residents). Differently from the robust unconditional case, m units will be drawing with replacement among those units producing at least the same level of output as the municipality under assessment but with a higher probability if operating in a more similar context (in this case the municipal size). Also in the conditional case the efficiency score can be greater than 1, indicating that the evaluating municipality is more efficient than

the m municipalities in the reference sample with similar municipal size.

5. Data

For this analysis we observe all 307 municipalities in the Flemish region of Belgium over the period 2006 and 2011, which mostly covers a local government electoral term.⁵ To discuss the main findings in light of the municipal size debate, we group the municipalities in seven classes as reported in Table 1. As a result of the push- and pull- experiences in Belgium over the last 40 years most of the municipal size ranges between 5,000 to 50,000 inhabitants.

Table 1: Distribution of municipalities in size classes.

Municipal classes	<5,000	5,000-10,000	10,000-15,000	15,000-20,000	20,000-50,000	50,000-150,000	>150,000	Total
Number of municipalities	13	74	83	52	73	10	2	307

Note: 307 municipalities in the Flemish region of Belgium over the period 2006 and 2011.

In line with the Flemish institutional setting, the service level provision is measured along eight municipal areas: general administration, culture, care services, educational services, housing, road maintenance and local mobility, public safety and environment. For each area, the sub-indicators are chosen in line with the local government efficiency literature (for an extensive review see Narbón-Perpiñá and De Witte, 2018a) and with the data availability.⁶

Before discussing the variables, it is relevant to stress that, despite the Flemish municipalities have sufficient leeway in developing their own policy targets and setting their own goals, the homogeneity in service quality is enforced by the regional level (i.e., the Flemish government). In fact, for all policy domains studied in the paper, the municipality is the main actor to initiate the policy and set the size of the facilities. This guarantees that the facilities are targeted to the needs of the population. However, the Flemish government is always co-funding part of these facilities (e.g. in education, in care services, in road maintenance, in public safety, in environment, in culture and in housing). Thanks to this co-financing, it can set rigorous and homogeneous

⁵Data are available at <https://statistieken.vlaanderen.be/QvAJAXZfc/notoolbar.htm?document=SVR%20FSVR-alle-domeinen.qvw&host=QVS%40cww100154&anonymous=true>. In Flanders there are 308 municipalities, but there is lack of data for one of them. For more details, please refer to D’Inverno and De Witte (2020).

⁶In the process of data collection, we assemble as many indicators as possible to measure the municipal service provision, so to grasp different nuances and to offer a richer picture of the municipal services. To overcome the curse of dimensionality, we aggregate the sub-indicators at municipal area using a robust Benefit of the Doubt composite indicator, so to be consistent with the non-parametric formulation adopted in the rest of the paper (for technical details the reader can refer for example to Fusco et al., 2019).

quality standards across municipalities, and enforces them through visits of inspectorate teams. This makes sure that we can consider the quality across municipalities as rather homogeneous.

The service provision of a municipality can consist of offering the most of desirable features or reducing the most of undesirable ones. These two aspects are equally important and subject to the control of the municipality. For example, the municipalities are equally in control of the effort to curb the level of criminality as much as to promote cultural events giving their patronage or to support urban development by issuing building permits, to name a few. In particular, as desirable indicators we include the net number of foreigners and the number of households for the administrative services, as proxy respectively for the workload and the services provided in the immigration offices and in the registry offices. For the municipal cultural function, we look at the number of cultural events officially reported in the national and local database and considered a proxy of the municipal effort to boost the cultural engagement of the local community by granting patronage and authorizations. For the care services, we consider the main recipients benefiting from these services, namely the number of children in kindergarten and the elderly people, reasonably assuming that they are strongly correlated to the issued services. Similarly, we look at the number of students in primary school as a municipal indicator for the education function. For the housing function we consider the built-up area as a proxy of the municipal effort in promoting the urban development and the workload to issue building permits. As undesirable indicators, we consider the number of accidents as proxy for the road interventions. In this sense, the services provided in this municipal area can be intended to contain undesirable features by means of circulation interventions and local road maintenance. For the local security, we look at the level of thefts, physical and property crimes that a municipality should combat as much as possible by means of local police interventions as a service to the local community. Finally, for the environmental function the waste and energy consumption have been considered as proxy of the municipal effort in promoting a sustainable environmental attitude among the citizens. Similarly to the previous two cases, the service is as more effective as this indicator is more contained. Note that among the listed variables there are no indicators directly measuring the quality of the provided services. Nevertheless, we can reasonably see some of them as quality indicators. For example, the number of accidents can reflect the quality of the services provided within the local mobility function: a higher number of accidents can reflect to some extent a poorer road surface maintenance. Similarly, a higher level of criminality can reveal a low quality in the municipal efforts in the local security.

For the conditional model three economic variables are considered as proxy of the operating background (in line with the review of Narbón-Perpiñá and De Witte, 2018b): the fiscal income, defined as the income per capita recorded in a municipality and economic index of the citizens'

wealth; the financial debt, measured as the excess of expenditures over revenues per capita and associated with the granted loans and the fiscal capacity; the unemployment rate, defined as the ratio between the unemployed residents in the range of 15-64 years and the total working population and indicator to some extent of the living conditions experienced in a municipality.⁷ The descriptive statistics of the variables encompassed in the composite indicator are presented in Appendix A, together with the information on the average spending share for the weight restrictions of each municipal function.

6. Results

This section discusses the results of the analysis, and links the findings to the municipal amalgamations discussed in section 3. We proceed in four steps. First, we present the estimates of the service level provision, and link them to the municipal size. Second, we provide insights on the expenditure composition and descriptive statistics of the local public expenditures, and also relate them to the municipal size. Third, we combine both expenditures and service level provision in an efficiency analysis. Finally, we exploit the differences between the conditional and unconditional efficiency scores and link them non-parametrically to the municipal size. In Appendix B we provide some robustness checks to gauge the validity of the main findings.

6.1. Service level provision and municipal size

Table 2 presents the descriptive statistics of the robust municipal service provision composite indicator conditioned on fiscal income, financial debt and unemployment for 307 Flemish municipalities in the period 2006-2011. The overall mean value suggests that, on average, there is little room for municipal service level improvement. Nevertheless, the minimum values indicate the presence of some municipalities providing meagre services.

To explore the association between municipal size and the local service level, the estimates in Table 2 are presented according to the municipal size. From the mean and minimum values, a U-shaped relationship emerges, suggesting that small and larger municipalities provide a greater service level than the municipalities between 10,000 and 20,000 inhabitants. In this perspective, we might conclude that the reform started almost 40 years ago (see section 3) was effective, insofar as relatively smaller municipalities merged and relatively larger municipalities have managed to internalized the geographical spillovers with the result of reaching higher level of service provision in both cases.

⁷In D’Inverno and De Witte (2020) the socio-demographic structure and political ideology have been also included for alternative conditional model specifications. For the present case, we focus only on the composite indicator conditioned on economic variables to preserve discriminatory power in the model specification.

Table 2: Descriptive statistics of the comprehensive municipal service provision Composite Indicator (CI) for 307 municipalities over 2006–2011.

Municipal service provision CI	Obs	Mean	St. Dev.	Min	Max
Overall	307	0.9893	0.0247	0.8173	1.0000
<i>By municipal size classes (residents)</i>					
< 5,000	13	0.9998	0.0005	0.9984	1.0000
5,000 – 10,000	74	0.9941	0.0132	0.9289	1.0000
10,000 – 15,000	83	0.9904	0.0237	0.8584	1.0000
15,000 – 20,000	52	0.9811	0.0312	0.8770	1.0000
20,000 – 50,000	73	0.9853	0.0310	0.8173	1.0000
50,000 – 150,000	10	0.9996	0.0012	0.9962	1.0000
> 150,000	2	1.0000	0.0000	1.0000	1.0000

Note: Average for 307 municipalities in the Flemish region of Belgium over the period 2006 and 2011. Municipal service composite indicator conditioned on fiscal income, financial debt and unemployment, with weight restrictions.

6.2. Local public expenditures and municipal size

The estimated municipal service provision brings additional elements useful to shed lights on the role of local public expenditures, namely the optimal expenditure shares across the municipal functions. In this perspective, we explore the expenditure composition by municipal area and municipal size, looking at the differences between the averaged observed and optimal expenditure shares to possibly identify systematic scale deviations. We graphically display the deviation from the observed expenditure share, computed as the difference between the averaged optimal (shadow) share and the averaged observed one, for each municipal function. Figure 1 shows the deviation by municipal size class separately for each function, while Figure 2 reports the deviation by municipal function separately for each municipal class, presented along with the 95% confidence interval. For a given function, a positive deviation detects the groups of municipalities that give more importance to that function than the one recognized from the actual spending, while the opposite holds for a negative deviation. To put it another way, we can interpret the deviations from the averaged observed expenditure shares as revealed preferences over the different municipal functions, whether influenced by the specialization of a municipality or its historical institutional setting, besides generally acknowledged structural priorities. The two figures help to grasp complementary elements and systematic aspects, either by size or by function. Figure 1 offers interesting insights while exploring the deviations by municipal size class separately for each function. First, we can distinguish a group of functions reporting positive deviations in almost all the municipal classes except few exceptions and this is the case for the general administration, the culture, the care services and the local mobility (and the opposite holds for

the remaining ones). For the interpretation of the results we remind that each municipality has been benchmarked against municipalities performing at least as good as the one under scrutiny and working in similar operating context. In this perspective, the municipalities give more importance than the one acknowledged in the actual municipal balance sheets to functions whose attention is mostly oriented in supporting the citizen and its social interaction within the community. On the contrary, negative deviations are tracked in municipal areas whose services aim at creating an environment where the interactions can take place, like the security, the housing and the environmental function. On average, these functions attract less importance, once the municipality ensures basic needs in terms of “environmental” conditions at large. Figure 2 helps to detect some patterns with respect to the municipal size. For example, the municipal classes with 5,000 to 10,000 residents and with 10,000 to 15,000 share similar deviations, within the framework outlined above and the same applies also for the municipalities with 15,000 to 20,000 residents and 20,000 to 50,000 residents. The municipal class with 50,000 to 150,000 residents differ from the two preceding ones in the care services function, being attached less importance, and in the security and in the environmental function, reporting a lower negative deviation: as soon as the municipal size increases, the attention to the individual shrinks in favour of the context that hosts the community. When looking at the distribution of the deviations for the largest municipalities we can observe that the optimal and the observed composition mostly coincides.

Beyond the considerations about the municipal service provision and the expenditure composition, the level itself of the local expenditure across the municipal classes deserves attention.

If the reforms were successful from the service provision side, the same does not hold for the expenditure level (cost side). The distribution of the total expenditure per size classes, presented in the top panel of Table 3, suggests that we end up drawing the same considerations that pushed the legislator to reform the expenditure side back to almost 40 years ago. The per capita expenditure of municipalities with less than 5,000 residents exceeds the expenditure level of municipalities with 5,000 to 15,000 residents and almost reaches the level of municipalities with 15,000 to 20,000 residents. Moreover, the per capita expenditure of the two largest municipalities (Ghent and Antwerp) almost doubles the average per capita expenditure of all the other classes.

More in general, we can observe a U-shaped relationship between the population size and the expenditure level found also in other countries (see section 2). As an underlying mechanism, small municipalities face substantial fixed costs in the provision of minimum service levels, while large municipalities reach a level of complexity in the service provision that usually requires a higher number of staff, so that the cost of bureaucracy is higher than in smaller municipalities (Suzuki, 2016; Tavares, 2018). This mechanism emphasises the need of complementing the measurement of service provision levels with an efficiency evaluation of the financial resource use.

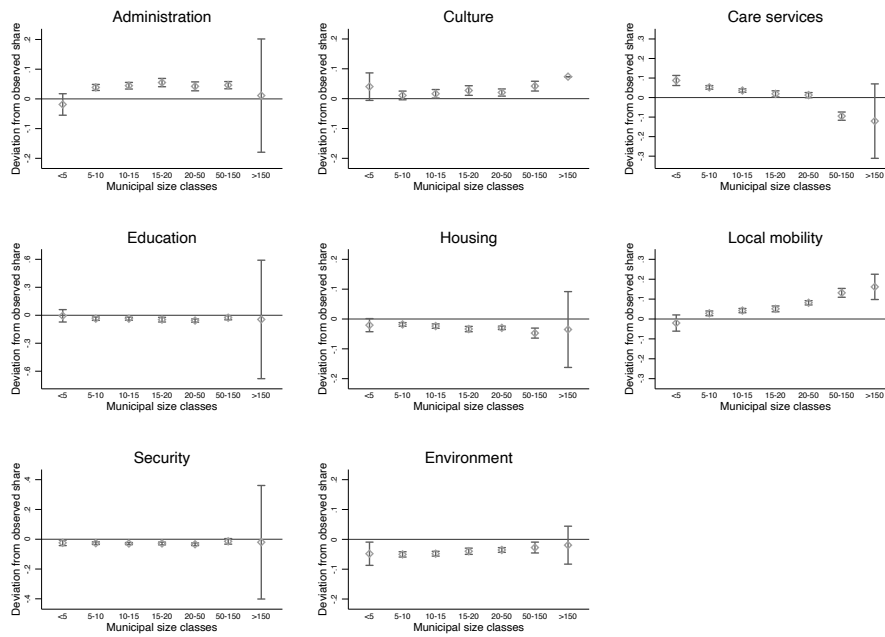


Figure 1: Deviations from averaged observed expenditure share by municipal size class for each function (computed as the difference between the average optimal and observed expenditure share). *Note: Vertical bars denote 95% confidence intervals.*

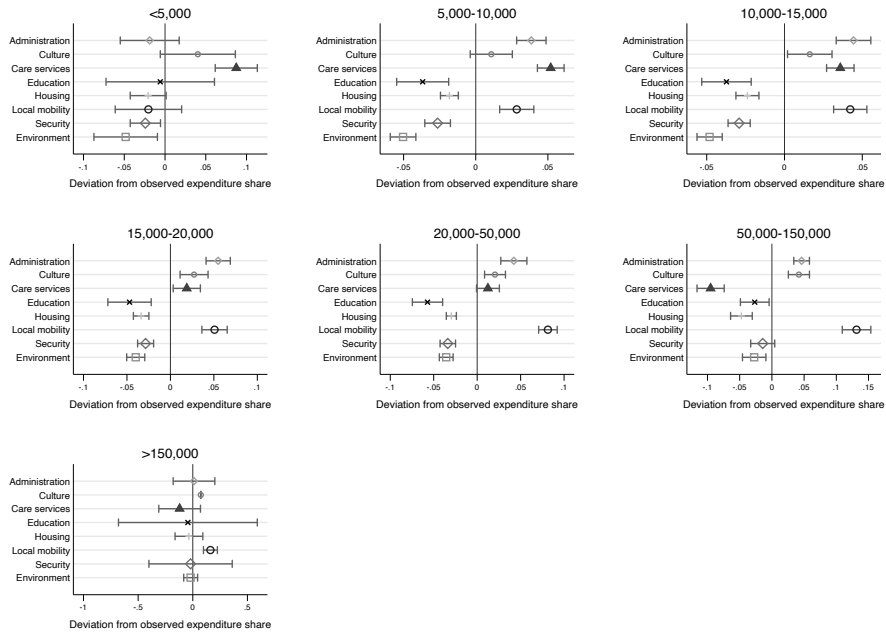


Figure 2: Deviations from averaged observed expenditure share by municipal function for each size class (computed as the difference between the average optimal and observed expenditure share). *Note: Horizontal bars denote 95% confidence intervals.*

6.3. An efficiency analysis for the municipal service provision and expenditure

Next, using the robust conditional DEA model outlined in Section 4, we relate input and outputs. In particular, in line with a stream of the literature on local government efficiency analysis (see for example Afonso and Fernandes, 2008; Borge et al., 2008; Bosch et al., 2012 and for an extensive review Narbón-Perpiñá and De Witte, 2018a), to measure the efficiency of the municipal expenditure we consider the total expenditure per capita as input and the municipal service composite indicator described in the previous section as output. A similar input-output construction has been taken before in the literature. In particular, several authors have proposed a composite local government indicator of municipal performance, but usually either they impose the same weights to aggregate different municipal areas (as for example Afonso et al., 2005; Afonso and Fernandes, 2008) or the weights are taken from the budget allocation but not included in the form of weight restrictions (as for example in Borge et al., 2008; Bosch et al., 2012). The total output indicator computed as such is then included in the efficiency analysis.⁸ Table 3 summarises the descriptive statistics of the above-mentioned variables together with the number of residents considered as proxy for the municipal size.

Table 3: Descriptive statistics of the input, output and control variables

	Obs	Mean	St. Dev.	Min	Max
<i>INPUT</i>					
Total expenditure (€ per capita)	307	1338.40	350.40	840.32	3255.27
<i>By municipal size classes (residents)</i>					
< 5,000	13	1317.79	247.54	879.47	1649.96
5,000 – 10,000	74	1224.74	178.41	840.32	1818.37
10,000 – 15,000	83	1267.55	305.14	895.08	2788.03
15,000 – 20,000	52	1334.03	349.60	915.73	2838.09
20,000 – 50,000	73	1420.74	362.73	908.29	3255.27
50,000 – 150,000	10	1863.10	172.39	1641.66	2219.05
> 150,000	2	3103.40	213.47	2952.46	3254.34
<i>OUTPUT</i>					
Municipal service provision composite score	307	0.9893	0.0247	0.8173	1.0000
<i>CONTROL</i>					
Total number of Residents	307	20,154.53	32,377.18	969.17	475,788.00

Note: Average for 307 municipalities in the Flemish region of Belgium over the period 2006 and 2011. Municipal service composite indicator conditioned on fiscal income, financial debt and unemployment, with weight restrictions.

⁸By doing this, we have the opportunity to analyze separately the performance level and mostly to avoid the curse of dimensionality in the efficiency analysis part carrying as much heterogeneity as possible. Moreover, this can provide policy makers with an attractive and intuitive tool to interpret and communicate the results. The municipal composite indicator and the local expenditure can be represented in an easy graphical two dimensional plot, to seize the relationship between the input and the output and give the idea of the best practice frontier.

Table 4 provides the results of the efficiency analysis. The overall mean input efficiency scores range between 0.74 in the unconditional efficiency model and 0.76 in the conditional efficiency model, suggesting an average potential reduction in the expenditure per capita in the order of 24% - 26%. A back-of-the-envelope analysis indicates that the potential efficiency gain is quite sizeable, corresponding to a reduction of 321 euro - 348 euro per capita. The distribution of the means along the municipal size classes follows an inverse U-shaped relationship in the case of the unconditional estimation and in the conditional case as well. The largest municipalities have the lowest efficiency scores, which provides evidence for the presence of remarkable diseconomies of scale. The same holds when conditioned on the municipal size, albeit to a lesser extent. Following the mechanism discussed in section 2, a possible explanation has to be linked to the public service spillovers from the center municipalities to their neighbours.

Table 4: Descriptive statistics of the estimated municipal expenditure efficiency scores for 307 municipalities over 2006–2011

Municipal expenditure efficiency	Obs	Mean	St. Dev.	Min	Max
Robust Unconditional	307	0.7440	0.1432	0.2941	1.1048
<i>By municipal size classes (residents)</i>					
< 5.000	13	0.7500	0.1514	0.5686	1.0648
5.000 – 10.000	74	0.7903	0.1071	0.5184	1.1048
10.000 – 15.000	83	0.7785	0.1305	0.3445	1.0716
15.000 – 20.000	52	0.7457	0.1517	0.3278	1.0393
20.000 – 50.000	73	0.6985	0.1328	0.2941	1.0329
50.000 – 150.000	10	0.5171	0.0442	0.4321	0.5845
> 150.000	2	0.3097	0.0216	0.2944	0.3249
Robust Conditional	307	0.7644	0.1344	0.3086	1.0990
<i>By municipal size classes (residents)</i>					
< 5.000	13	0.7599	0.1527	0.5356	1.0038
5.000 – 10.000	74	0.8206	0.1095	0.5384	1.0990
10.000 – 15.000	83	0.7585	0.1293	0.3288	1.0272
15.000 – 20.000	52	0.7543	0.1533	0.3305	1.0095
20.000 – 50.000	73	0.7326	0.1379	0.3086	1.0503
50.000 – 150.000	10	0.7258	0.0610	0.6039	0.8209
> 150.000	2	0.5771	0.0395	0.5492	0.6050

Note: For 307 municipalities in the Flemish region of Belgium over the period 2006 and 2011. Conditional efficiency scores conditioned on the total number of residents.

6.4. Non-parametric analysis of municipal size

As a final step, to draw a conclusion on the existence of (dis)economies of scale in local public good provision, we explain the difference between the conditional and unconditional efficiency

scores through the municipal size. Therefore, we nonparametrically regress the ratio between the unconditional and the conditional efficiency score on the municipal size (as suggested by Li and Racine, 2007; Daraio and Simar, 2007a; De Witte and Kortelainen, 2013). A positive slope in the visualization of the nonparametric regression denotes a positive relationship between the detected level of efficiency and the municipal size and the opposite holds for a negative slope.⁹ Figure 3 displays a negative relationship between the municipal size and the municipal performance, pointing at the presence of diseconomies of scale mostly at any level. The pattern holds if accounting for the uncertainty in the data, represented by the 95% confidence intervals.

Interestingly, we observe a puzzling finding for the smallest municipalities (with less than 5,000 inhabitants), which seem to perform more efficient than the class of 5,000 to 10,000 inhabitants. Municipal statistics provide insight in a potential underlying mechanism. Particularly in the smallest municipalities, and at odd with the general findings in the literature (Breunig and Rocaboy, 2008; Suzuki, 2016; Tavares, 2018), the educational background of employees in the smallest municipalities is significantly higher than employees in larger municipalities. In fact, official statistics reveal a negative correlation between municipal size and the share of highly qualified employees.¹⁰

The evidence from this plot enriches the considerations outlined in the previous discussion of the results (where we look at the distribution of the estimated municipal service levels and the distribution of the efficiency scores). The overall picture only weakly supports evidence for the presence of an optimal size in the adequate use of the financial resources around 10,000-15,000 citizens (also found for example in Italy by Agasisti and Porcelli, 2019 and in Spain by Hortas-Rico and Rios, 2019). It strongly points at the presence of diseconomies of scale, especially for municipalities with more than 50,000 residents. This conclusion is in line with what has been anticipated about Antwerp in Section 3.2, about problems of congestion, denoting the need to cut down on red tape. More in general, a key to interpret the presence of diseconomies of scale might be found in the merger process that lead to the actual municipal size framework in Flanders. As described in Section 3, most of the mergers were implemented on a compulsory basis rather than on a voluntary one. Following Swianiewicz (2018) and Tavares (2018), this might be a possible explanation why the increase in the municipal size has not fully achieved the expected goal.

⁹Actually Daraio and Simar (2007a) suggest the use of the ratio between the conditional and the unconditional efficiency scores. We use the inverse of this advocated ratio to facilitate the graphical interpretation of the relationship between the ratio and the control variable, inspired by other papers *mutatis mutandis* (see for example Rogge et al., 2017).

¹⁰<https://lokaalbestuur.vlaanderen.be/sites/default/files/public/publicaties/VlaamseProfielschets2017.pdf>

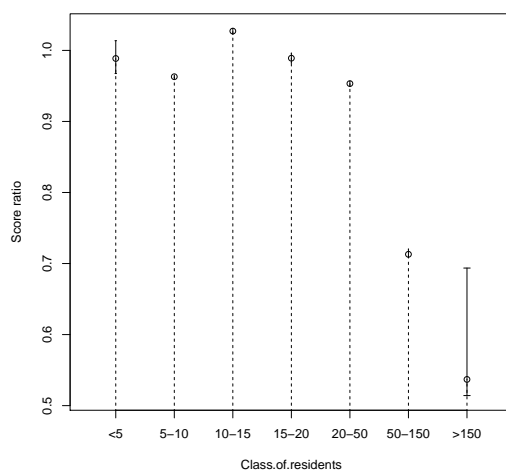


Figure 3: Visualization of the relationship between the municipal class size and the estimated municipal efficiency. *Note:* Municipal service composite indicator conditioned on economic variables with weight restrictions. Observation period: 2006-2011. The bars represent the 95 percent confidence intervals around the point estimates.

7. Conclusion

In this concluding section we outline the main findings of the present study together with the policy implications and future avenues of research.

7.1. Summary of core findings

The role of local government size has been extensively investigated in the literature, both in relationship with the provided service levels and the adequate use of financial resources. Moreover, particular attention has been given to the presence of (dis)economies of scale and the existence of an optimal scale size. However, conclusive evidence is lacking.

In this paper, we provide new evidence on the local government size debate by focusing on Flanders, the northern region of Belgium. Flanders makes an interesting study case for its continuous reform process about municipal size over the last 40 years and the strong quality control from the regional government. We measure the service level provision by adapting a novel composite indicator as proposed in D’Inverno and De Witte (2020) to 307 Flemish municipalities in the period 2006–2011. The indicator, rooted in the approach suggested by Zanella et al. (2015), is flexible to account for different municipal areas granting the ‘Benefit of the Doubt’ in the benchmarking procedure (i.e. attaching higher weights to what municipalities provide the most) even if constrained to the relative importance that each area has in the budget allocation. Moreover, the indicator deals with both desirable and undesirable features and adjusts for the heterogeneous operating context. Next, we estimate the efficiency of the municipal expenditures

by performing a conditional robust Data Envelopment Analysis model. The main results of this integrated non-parametric analysis indicate the presence of diseconomies of scale, with a weak evidence that there is an optimal scale approximately at 10,000 residents.

7.2. Policy conclusions and future research

The paper induces three avenues for policy conclusions. First, from a methodological perspective, the suggested composite indicator allows for differentiated benchmarking. The differentiation originates from the endogenous weights, accounting for the operational environment, and mitigating the uncertainty. The endogenous weights are attractive for policy purposes as it ensures that there is no set of weights that is more benevolent for the evaluated municipality. This contrasts the current autocratic weighting systems, which will always (dis)favor municipalities with strong performances in the services that are higher (lower) weighted. Thanks to controlling for the environmental variables only municipalities operating in a similar environment are compared, such that a more fair benchmarking arises. The robust service level scores signal the inherent uncertainty in performance measurement. The robust scores allow for the computation of confidence intervals (we leave this as a straightforward extension) such that it becomes clear that there are overlaps in the ranking.

A second avenue for policy conclusions is deduced from the empirical application. As we showed clear diseconomies of scale once a threshold of 10,000 inhabitants is reached, it becomes obvious that in the provision of municipal services scale economies should not be taken for granted. This contrasts the traditional Taylorism vision, which predicts scale economies in production sectors. However, we argued in the paper that municipal service provision is a multidimensional concept and includes to a large extent the provision of traditional services where scale economies are not apparent. Despite the absence of scale economies in the overall service level provision, we leave room to examine how inter-municipal collaboration can be used to exploit the scale economies at the specific production of some material pools such as waste disposal, incineration or recycling.

A final avenue for policy conclusions emerges from the use of benchmarking for the individual municipalities. The suggested methodology allows policy makers to calculate a fair performance score, which can be instrumental for benchmarking purposes. From a policy perspective, the estimated service level score should not be used for ‘naming and shaming’ (Cabus and De Witte, 2012) purposes or for sunshine regulation (De Witte and Saal, 2010), but rather for identifying municipalities that are lagging behind. Municipalities with low service level provision scores should be stimulated to exchange practices with the best in class. Thanks to the effective use of the service level scores, the provision of collective goods will be provided in a more efficient way, such that value for tax money will increase. In a similar vein, one may imagine that the municipal performance score could be included as an additional criterion in the fiscal equalization

scheme for the municipalities. In Flanders, for instance, the regional government supplies a block grant to each municipality applying an elaborate allocation based on multiple criteria. Typical examples are a handicap in tax capacity or specific needs as for coastal municipalities.

Finally, the paper provides various lines for future research. First, the presented findings are correlational evidence. Recent studies have evaluated the impact of reforms about the municipal fragmentation or amalgamation on various indicators, mostly using Difference-in-Differences techniques (for a review see for example Tavares, 2018; Swianiewicz, 2018). Following an emerging stream of literature (see for example Santín and Sicilia, 2017; D’Inverno et al., 2019), future lines of research might combine the toolbox of performance and efficiency analysis applied in this paper with policy evaluation techniques so to provide quasi-experimental evidence and therefore causal interpretation of the findings in the municipal size debate. Next, future research could explore cost and allocative efficiencies, and relate these to the service level provision.

References

- Afonso, A., Fernandes, S., 2008. Assessing and explaining the relative efficiency of local government. *The Journal of Socio-Economics* 37, 1946–1979.
- Afonso, A., Schuknecht, L., Tanzi, V., 2005. Public sector efficiency: an international comparison. *Public choice* 123, 321–347.
- Agasisti, T., Porcelli, F., 2019. Local governments efficiency and its heterogeneity empirical evidence from a Stochastic Frontier Analysis of Italian municipalities 2010-2015. Working papers 81. Societ Italiana di Economia Pubblica.
- Banker, R.D., Charnes, A., Cooper, W.W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science* 30, 1078–1092.
- Borge, L.E., Falch, T., Tovmo, P., 2008. Public sector efficiency: the roles of political and budgetary institutions, fiscal capacity, and democratic participation. *Public Choice* 136, 475–495.
- Bosch, N., Espasa, M., Mora, T., 2012. Citizen control and the efficiency of local public services. *Environment and Planning C: Government and Policy* 30, 248–266.
- Breunig, R., Rocaboy, Y., 2008. Per-capita public expenditures and population size: a non-parametric analysis using French data. *Public choice* 136, 429–445.
- Byrnes, J., Dollery, B., 2002. Do economies of scale exist in Australian Local Government? A review of the research evidence. *Urban Policy and Research* 20, 391–414.

- Cabus, S.J., De Witte, K., 2012. Naming and shaming in a fairway. on disentangling the influence of policy in observed outcomes. *Journal of Policy Modeling* 34, 767–787.
- Cazals, C., Florens, J.P., Simar, L., 2002. Nonparametric frontier estimation: a robust approach. *Journal of Econometrics* 106, 1–25.
- Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research* 2, 429–444.
- Cherchye, L., De Witte, K., Perelman, S., 2019. A unified productivity-performance approach applied to secondary schools. *Journal of the Operational Research Society* , 1–16.
- Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., 2007. An introduction to benefit of the doubt composite indicators. *Social indicators research* 82, 111–145.
- Daraio, C., Simar, L., 2005. Introducing environmental variables in nonparametric frontier models: a probabilistic approach. *Journal of Productivity Analysis* 24, 93–121.
- Daraio, C., Simar, L., 2007a. Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications. Springer Science & Business Media.
- Daraio, C., Simar, L., 2007b. Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach. *Journal of Productivity Analysis* 28, 13–32.
- Daraio, C., Simar, L., 2014. Directional distances and their robust versions: Computational and testing issues. *European Journal of Operational Research* 237, 358–369.
- Daraio, C., Simar, L., Wilson, P.W., 2020. Fast and efficient computation of directional distance estimators. *Annals of Operations Research* 288, 805–835.
- De Borger, B., Kerstens, K., Moesen, W., Vanneste, J., 1994. Explaining differences in productive efficiency: An application to Belgian municipalities. *Public Choice* 80, 339–358.
- De Witte, K., Geys, B., Schönhage, N.L., 2018. Strategic public policy around population thresholds. *Journal of Urban Economics* 106, 46–58.
- De Witte, K., Kortelainen, M., 2013. What explains the performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables. *Applied Economics* 45, 2401–2412.
- De Witte, K., Saal, D.S., 2010. Is a little sunshine all we need? on the impact of sunshine regulation on profits, productivity and prices in the dutch drinking water sector. *Journal of Regulatory Economics* 37, 219–242.

- Depriens, D., Simar, L., Tulkens, H., 1984. Measuring labor inefficiency in post offices. The Performance of Public Enterprises: Concepts and measurements. M. Marchand, P. Pestieau and H. Tulkens (eds.), Amsterdam, North-Holland , 243–267.
- D’Inverno, G., De Witte, K., 2020. Service level provision in municipalities: A flexible directional distance composite indicator. *European Journal of Operational Research*, *In press* .
- D’Inverno, G., Smet, M., De Witte, K., 2019. Impact evaluation in a multi-input multi-output setting: Evidence on the effect of additional resources for schools. Mimeo .
- Fox, W.F., Gurley, T., 2006. Will consolidation improve sub-national governments? The World Bank.
- Fusco, E., Vidoli, F., Rogge, N., 2019. Spatial directional robust benefit of the doubt approach in presence of undesirable output: An application to Italian waste sector. *Omega* .
- Geys, B., Moesen, W., 2009. Measuring local government technical (in)efficiency: an application and comparison of FDH, DEA, and econometric approaches. *Public Performance & Management Review* 32, 499–513.
- Holcombe, R.G., Williams, D.W., 2009. Are there economies of scale in municipal government expenditures? *Public Finance and Management* 9, 416.
- Hortas-Rico, M., Rios, V., 2019. Is there an optimal size for local governments? A spatial panel data model approach. *Regional Studies* , 1–16.
- King, D., 2016. *Fiscal Tiers (Routledge Revivals): The Economics of Multi-Level Government*. Routledge.
- Lavigne, C., De Jaeger, S., Rogge, N., 2019. Identifying the most relevant peers for benchmarking waste management performance: A conditional directional distance Benefit-of-the-Doubt approach. *Waste Management* 89, 418–429.
- Li, Q., Racine, J.S., 2007. *Nonparametric econometrics: theory and practice*. Princeton University Press.
- Mbuvi, D., De Witte, K., Perelman, S., 2012. Urban water sector performance in Africa: A step-wise bias-corrected efficiency and effectiveness analysis. *Utilities Policy* 22, 31–40.
- Melyn, W., Moesen, W., 1991. Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available. Center for Economic Studies, KU Leuven, 17 .

- Moesen, W., Vanneste, J., 1983. Public choice, local budgets and municipalities as open or closed economies. *SESO Papers* , 1–28.
- Narbón-Perpiñá, I., De Witte, K., 2018a. Local governments' efficiency: A systematic literature review-Part I. *International Transactions in Operational Research* 25, 431–468.
- Narbón-Perpiñá, I., De Witte, K., 2018b. Local governments' efficiency: A systematic literature review-Part II. *International Transactions in Operational Research* 25, 1107–1136.
- Oates, W.E., 1972. *Fiscal federalism*. New York: Harcourt Brace Jovanovich.
- OECD, 2018. *Subnational governments in OECD countries: Key data (brochure)*. OECD, Paris .
- Reingewertz, Y., 2012. Do municipal amalgamations work? Evidence from municipalities in Israel. *Journal of Urban Economics* 72, 240–251.
- Rogge, N., De Jaeger, S., Lavigne, C., 2017. Waste Performance of NUTS 2-regions in the EU: A Conditional Directional Distance Benefit-of-the-Doubt Model. *Ecological Economics* 139, 19–32.
- Santín, D., Sicilia, G., 2017. Impact evaluation and frontier methods in education: a step forward. *Handbook of Contemporary Education Economics* , 211–245.
- Simar, L., Vanhems, A., 2012. Probabilistic characterization of directional distances and their robust versions. *Journal of Econometrics* 166, 342–354.
- Solé-Ollé, A., Bosch, N., 2005. On the relationship between authority size and the costs of providing local services: lessons for the design of intergovernmental transfers in Spain. *Public Finance Review* 33, 343–384.
- Suzuki, K., 2016. Politics of municipal consolidation. *Global Encyclopedia of Public Administration, Public Policy, and Governance* , 1–7.
- Swianiewicz, P., 2018. If territorial fragmentation is a problem, is amalgamation a solution?—Ten years later. *Local Government Studies* 44, 1–10.
- Tavares, A.F., 2018. Municipal amalgamations and their effects: A literature review. *Miscellanea Geographica – Regional Studies on Development* 22, 5–15.
- Vanden Eeckaut, P., Tulkens, H., Jamar, M.A., 1993. Cost efficiency in Belgian municipalities. *The Measurement of Productive Efficiency—Techniques and Applications* , 300–334.

Zanella, A., Camanho, A.S., Dias, T.G., 2015. Undesirable outputs and weighting schemes in composite indicators based on data envelopment analysis. *European Journal of Operational Research* 245, 517–530.

Appendix A. Descriptive statistics

Table A.1: Descriptive statistics of the municipal service sub-indicators and economic context variables

	Obs	Mean	St. Dev.	Min	Max
<i>Municipal sub-indicators (per 1000 inhabitants)</i>					
1. Robust BOD CI for general administration	307	0.85	0.1	0.75	2.19
Net Foreigners	307	7.44	4.87	1.23	61.35
Households	307	404.5	23.12	355.32	508.82
2. Cultural Events	307	6.53	5.28	0.46	27.83
3. Robust BOD CI for care service	307	0.8	0.11	0.58	1.39
Children in kindergarten	307	37.35	7.16	9.98	86.54
Residents over 80	307	46.26	9.15	22.06	84.23
4. Students in primary school	307	63.51	13.28	11.22	139.51
5. Built-up area (Km2)	307	64.33	21.22	22.76	248.09
6. Accidents	307	4.44	1.42	1.33	10.11
7. Robust BOD CI for Crime	307	0.5	0.2	0.17	1.42
Thefts	307	21.77	12.08	5.92	83.4
Physical crimes	307	4.91	2.25	2.03	17.44
Property crimes	307	8.43	3.22	2.54	23.38
8. Robust BOD CI for environment services	307	0.61	0.12	0.44	1.43
Waste (Tonnes)	307	141.08	37.63	70.75	330.87
Energy consumption	307	7.9	1.7	5.69	21.15
<i>Economic context variables</i>					
Fiscal income (€ per capita)	307	16330.52	1757.94	11743.37	22799.69
Financial debt (€ per capita)	307	1014.43	557.79	9.64	5232.58
Unemployment	307	5.52	1.71	2.74	13.86

Note: Average for 307 municipalities in the Flemish region of Belgium over the period 2006 and 2011. For the municipal areas represented by more than one sub-indicators, a robust Benefit of the Doubt (BoD) composite indicator has been computed (for technical details Fusco et al., 2019) to reduce the dimensionality of the model specification while carrying the most of the available information.

Table A.2: Summary of the weights obtained from the municipal expenditure composition

	Administration	Culture	Care services	Education	Housing	Local mobility	Security	Environment
LOWER BOUND	0.09	0.07	0.06	0.05	0.03	0.07	0.05	0.05
AVERAGE	0.18	0.14	0.12	0.10	0.05	0.15	0.11	0.10
UPPER BOUND	0.27	0.21	0.17	0.14	0.08	0.22	0.16	0.15

Appendix B. Robustness check results

In our main model specification, the municipality under evaluation is benchmarked against a group of municipalities performing at least as good as the municipality itself (and not against the whole sample). For the data at hand, imposing such a condition on the drawing does not allow for much variation in the municipal service provision CI scores. Consequently, in the following we propose the same analysis allowing the drawing with replacement from the whole sample (as for example executed in Rogge et al., 2017; Lavigne et al., 2019). Table B.1 shows the estimated municipal service provision CI scores and Table B.2 reports the results of the efficiency analysis including the alternative CI scores. Figure B.1 and Figure B.2 show the deviations from the observed expenditure composition and Figure B.3 displays the influence of the municipal size on the municipal performance accounting for this alternative aggregate measure of municipal service provision. From a policy perspective the results are robust to this alternative specification and confirm the presence of diseconomies of scale for large municipalities, providing weak evidence for the presence of an optimal size of local public good provision of around 10,000 citizens.

Reasonably, once the municipalities are benchmarked against the whole sample rather than against those who are performing at least as good as they are, the revealed preferences that emerge from the expenditure composition slightly change for some functions, while keeping other constant. For example, the deviations from the observed expenditure share display similar patterns in the administration and the care services case. For other municipal functions, the evidence is overturned, as for example in the case of the cultural area. When compared against municipalities whose provision is at least as good as the one of the municipality under evaluation, the cultural function plays an important role in the overall evaluation, as it has been attached a greater (positive deviation between the optimal and the observed expenditure share) and increasing importance with respect to the municipal size. When it comes to the comparison of a municipality against all the other municipalities, then the value acknowledged to the cultural function is lower than its observed one across all the municipal size classes but for the biggest municipalities, giving more importance to other functions, such as the local security.

Table B.1: Descriptive statistics of the comprehensive municipal service provision Composite Indicator (CI) for 307 municipalities over 2006–2011.

Municipal service provision CI	Obs	Mean	St. Dev.	Min	Max
Overall	307	0.9215	0.0524	0.7775	1.0000
<i>By municipal size classes (residents)</i>					
< 5,000	13	0.9556	0.0554	0.8348	1.0000
5,000 – 10,000	74	0.9294	0.0470	0.8086	0.9990
10,000 – 15,000	83	0.9126	0.0499	0.7775	0.9991
15,000 – 20,000	52	0.9080	0.0514	0.7894	0.9970
20,000 – 50,000	73	0.9163	0.0539	0.7881	0.9993
50,000 – 150,000	10	0.9862	0.0162	0.9557	0.9987
> 150,000	2	0.9967	0.0006	0.9963	0.9972

Note: Average for 307 municipalities in the Flemish region of Belgium over the period 2006 and 2011. Municipal service composite indicator conditioned on fiscal income, financial debt and unemployment, with weight restrictions. (Source: Authors' elaboration based on D'Inverno and De Witte, 2020)

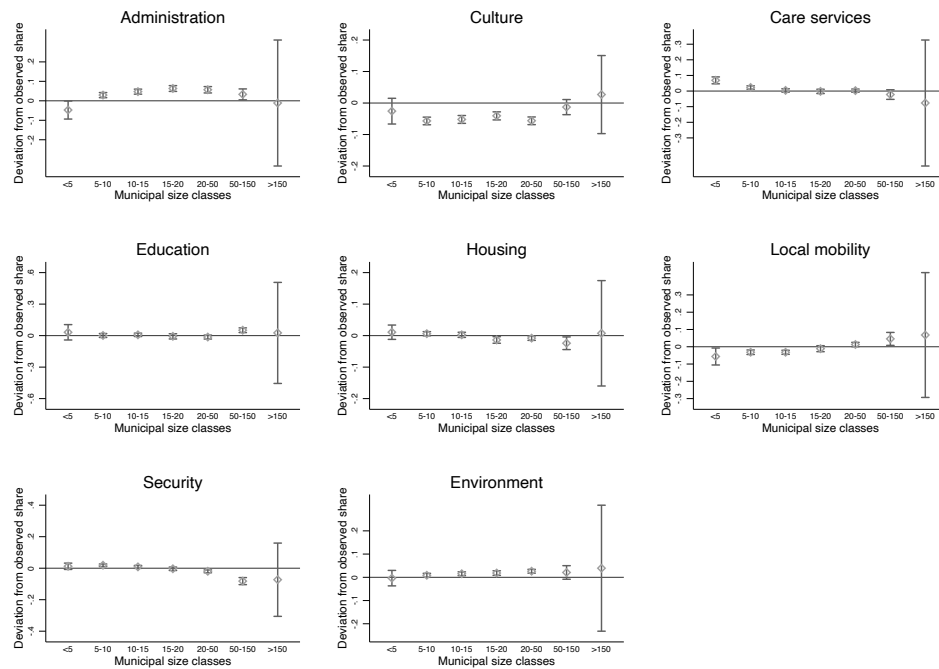


Figure B.1: Deviations from averaged observed expenditure share by municipal size class for each function (computed as the difference between the average optimal and observed expenditure share). *Note:* Vertical bars denote 95% confidence intervals.

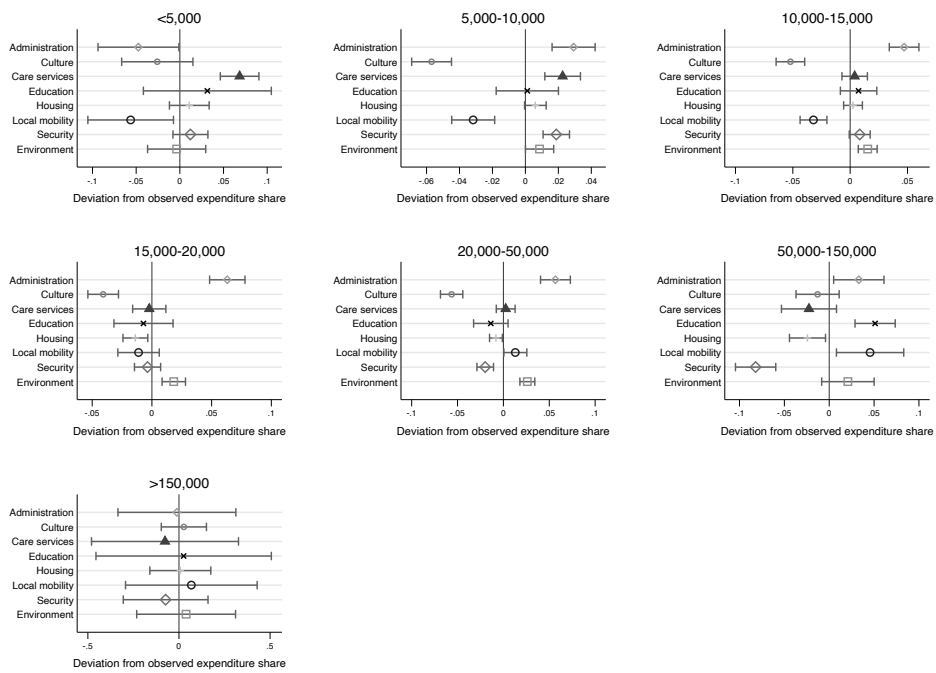


Figure B.2: Deviations from averaged observed expenditure share by municipal function for each size class (computed as the difference between the average optimal and observed expenditure share). *Note: Horizontal bars denote 95% confidence intervals.*

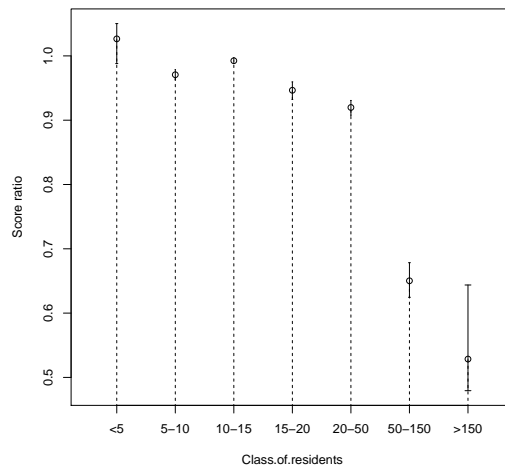


Figure B.3: Visualization of the relationship between the municipal class size and the estimated municipal efficiency. *Note: Municipal service composite indicator conditioned on economic variables with weight restrictions. Observation period: 2006-2011. The bars represent the 95 percent confidence intervals around the point estimates.*

Table B.2: Descriptive statistics of the estimated municipal expenditure efficiency scores for 307 municipalities over 2006–2011

Municipal expenditure efficiency	Obs	Mean	St. Dev.	Min	Max
Robust Unconditional	307	0.7472	0.1444	0.2885	1.0991
<i>By municipal size classes (residents)</i>					
< 5.000	13	0.7840	0.1661	0.5605	1.0288
5.000 – 10.000	74	0.7941	0.1166	0.5257	1.0991
10.000 – 15.000	83	0.7758	0.1364	0.3376	1.0542
15.000 – 20.000	52	0.7434	0.1513	0.3270	1.0287
20.000 – 50.000	73	0.7005	0.1330	0.2885	1.0246
50.000 – 150.000	10	0.5534	0.0405	0.5017	0.6364
> 150.000	2	0.3658	0.0372	0.3395	0.3921
Robust Conditional	307	0.7883	0.1386	0.3039	1.0851
<i>By municipal size classes (residents)</i>					
< 5.000	13	0.7659	0.1675	0.5354	1.0001
5.000 – 10.000	74	0.8192	0.1170	0.5385	1.0851
10.000 – 15.000	83	0.7814	0.1348	0.3494	1.0249
15.000 – 20.000	52	0.7876	0.1598	0.3313	1.0186
20.000 – 50.000	73	0.7629	0.1463	0.3039	1.0595
50.000 – 150.000	10	0.8542	0.0468	0.7811	0.9404
> 150.000	2	0.6942	0.0105	0.6867	0.7016

Note: For 307 municipalities in the Flemish region of Belgium over the period 2006 and 2011. Conditional efficiency scores conditioned on the total number of residents.