KU LEUVEN

DEPARTMENT OF ECONOMICS

Service level provision in municipalities: a flexible directional distance composite indicator

Giovanna D'INVERNO and Kristof DE WITTE

FACULTY OF ECONOMICS AND BUSINESS

DISCUSSION PAPER SERIES DPS20.11

JUNE 2020

Service Level Provision in Municipalities: A Flexible Directional Distance Composite Indicator☆

Giovanna D'Inverno[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

With increasing decentralization of public activities to the municipalities, it has become imperative to deploy an enhanced service provision analysis at the local level. This paper suggests the innovative use of a composite indicator to measure the multidimensional aspects of the local public provision comprising of several commonly administered municipal tasks. We propose a robust conditional version of a directional distance composite indicator with weight restrictions based on the municipal expenditure composition. Specifically, we deal with the presence of "undesirable" municipal service indicators and with the heterogeneity among the municipalities in their political preferences, priority public activities and operating environment characteristics. To illustrate the applicability of the suggested method, we show the construction of the municipal service provision composite indicator for 307 Flemish municipalities over the year 2006–2011.

Keywords: Data envelopment analysis, Composite indicator, Weight restrictions, Undesirable indicators, Local governments

1. Introduction

Conforming to the subsidiarity principle and the New Public Management theories, a gradual decentralization of the key activities from the national level to the municipal level has enabled the public sector service providers in developing a closer relationship with its citizens and provide services customized to citizens' needs. Accordingly, the pressure on the provision of public goods

☆We are grateful to Wim Moesen, Peter Bogetoft, Tom Van Puyenbroeck, Benny Geys, Ana Camanho, Massimo Riccaboni, Laura Carosi, Riccardo Cambini, three anonymous referees and the participants at the 6th Workshop on Efficiency and Productivity Analysis (Porto, Portugal) and at the 15th European Workshop on Efficiency and Productivity Analysis (London, UK) for their valuable comments and suggestions on a previous version of the paper. 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).

*Corresponding author.

Email addresses: giovanna.dinverno@kuleuven.be (Giovanna D'Inverno), kristof.dewitte@kuleuven.be (Kristof De Witte)

necessitates a more enhanced service level analysis at the local level and calls for suitable tools to measure and monitor local municipal service provision aimed at effective, innovative and sustainable public sector management.

This paper proposes the innovative use of a fully non-parametric approach to assess the local service provision in a dynamic framework. Specifically, this innovative application shows policy makers the potential of operational research for evaluating multidimensional performances of municipalities thereby accounting for several sensitive issues in practice. As the municipal tasks are composed of various multifaceted activities, the used operational research method relies on the construction of a composite indicator. However, there are five issues that are needed to be addressed in the model specification. First, it is necessary to acknowledge that municipalities differ in the activities they develop and do not develop. This decision is often driven by political preferences and depending on the competencies of different municipalities. This variety is reflected both in terms of local government priorities and their peculiar specializations. The municipal budget allocation keeps track of this kind of information as it presents the areas that should be prioritized and allocated more resources. On one hand, budget shares are the result of historical choices made by previous local public administrations. New municipal boards might have to deal with decisions taken in the past. This is to some extent reflected in the budget constraints. In this perspective, the budget shares cannot change very quickly when a new municipal administration takes place. On the contrary, a large part of the budget is constrained due to the choices made in the past, and therefore, are to some extent exogenous to the municipal administration. On the other hand, they reflect the current government preferences over different areas that require municipal intervention depending on municipal characteristics, voter preferences and perceived local needs. Reasonably, the political preferences of the population do not change dramatically such that services that are offered in the past (e.g. police services, health care or nursery) are still needed in the future. Second, the municipalities differ not only in what they are willing to do, but also in what they are able to achieve and the service they can provide. To deal with this kind of variety among the municipalities under assessment, we propose the use of a flexible 'Benefit-of-the-Doubt' (BoD) composite indicator that endogenously determines the weights for each municipal service. In particular, the weights in the composite indicator are assigned in such a way that the municipality under analysis is evaluated with its most favourable weighting. To avoid meaningless weights, the approach is combined with the weight restrictions based on the municipal expenditure composition. In this way we can provide objectively determined and endogenously flexible weights, but at the same time we directly constrain them according to the information contained in the municipal balance sheets. Given the limited leeway of municipal administrations to change the budget allocation dramatically and rapidly when they come to office, there might be some path dependency in the municipal performance. In other words, the current performance evaluation of a municipal board might be negatively affected due to budget allocation decisions taken in the past. Third, "more" is not always "better" along the dimensions we evaluate the municipal service. For instance, the municipalities have the duty to prevent criminality and it is apparent that higher the level of criminality, poorer the level of quality of service that the municipality provides to its citizens in terms of public safety. We deal with this kind of indicators considering them as undesirable features. For this reason, we tailor the suggested BoD approach to a directional distance function as proposed by Zanella et al. (2015b). Fourth, the characteristics of municipal operating environment have a role to play in the public activities delivery as they influence the selection and the importance of the different municipal areas. To avoid the assumption of "separability condition", we perform a conditional analysis of the emerging model, combined with its robust version, to handle the bias stemming from the atypical observations that can be possibly present in the units under analysis. Finally, as time matters, a dynamic component is added in the conditional model (Mastromarco and Simar, 2015) to exploit intertemporal variations in public service provision (Cordero et al., 2017a).

Taking into account the listed issues, we propose an innovative way to evaluate local municipal service provision. More precisely, we advocate a composite indicator built on a directional distance BoD model, including undesirable features and weight restrictions based on the expenditure composition, performing the robust and conditional analysis, within a dynamic framework. The composite indicator is applied to Flemish municipalities to measure municipal service provision of 307 Flemish municipalities over the years 2006–2011.

Despite the fact that in the Operational Research literature there are a large amount of studies focusing on local governments and various aspects related to their efficiency (for an extensive review see Narbón-Perpiñá and De Witte, 2018a,b), the present paper constitutes a step forward in several directions in this area. First, there are no studies that measure the overall municipal service provision comprehensively (see, e.g., Fusco et al., 2018; Karagiannis, 2017). There are studies that have either proposed this approach only for specific municipal functions, such as the waste collection service (Rogge et al., 2017; Fusco et al., 2019), or have focused only on the global output assessment and have not included undesirable features in their evaluation (Afonso and Fernandes, 2008; Yusfany, 2015). With regard to the inclusion of undesirable features in the construction of the composite indicator, we consider alternative specifications relying on state-of-the-art model formulations of Färe et al. (2019) and a variant of the Zanella et al. (2015b) specification. For these recent models, we provide a comparison of their performance. To the best of our knowledge, this is the first paper comparing these recent models. Second, when constructing the composite indicator we incorporate the information on the expenditure composition share for each municipal area in the weight restriction specification and not only as a direct weighting scheme to aggregate the municipal tasks (Bosch et al., 2012; Helland and Sørensen, 2015; D'Inverno et al., 2018). Finally, considering the huge amount of municipal efficiency papers, it was observed that just few of them included the robust and conditional analysis (Asatryan and De Witte, 2015; Cordero et al., 2017a), even if local public services depend on the characteristics of the municipalities and a fair analysis should account for these differences directly in the main model specification. The proposed innovative use of the tool can be easily adapted to other settings and countries. Additionally, one can benefit from using it to look at specific services, such as water services, waste management, social services and public order to name a few. The code is available on request from the authors to facilitate further research and to foster policy applications for overall municipal service measurement in alternative countries and/or for specific areas.

The remainder of this paper is structured as follows. Section 2 discusses the different municipal tasks and it explains the data. In Section 3 the methodological steps leading to the advocated 'robust conditional directional distance BoD with weight restrictions' composite indicator for the municipal service provision assessment are discussed. Section 4 presents the main findings obtained from the empirical application on Flemish municipalities. Lastly, Section 5 presents some final remarks and conclusions.

2. Municipal service level

As suggested by the OECD (2008), when constructing a composite indicator, its theoretical framework should be defined first, to clarify which phenomenon is intended to be measured and what are its subcomponents that represent it as a whole. Accordingly, it is essential to clarify on what dimensions the proposed municipal service composite indicator has been built upon. The services provided by the municipalities vary from country to country depending on several factors, such as for example the location, the geography, the history and the tradition. However, there are several commonly acknowledged functions that represent the main tasks of a municipality (for a general overview, Narbón-Perpiñá and De Witte, 2018a,b; OECD/UCLG, 2016). Although there might be some differences across countries (see CEMR, 2016), these include general administration, culture, education and care services, housing and public safety, road maintenance and environmental management. For the sake of clarity, we refer to Figure 1 which consists of an example list of services grouped by various municipal functions. These different intervention areas can be seen as the broad categories along which municipal services' composite indicator should be assessed.

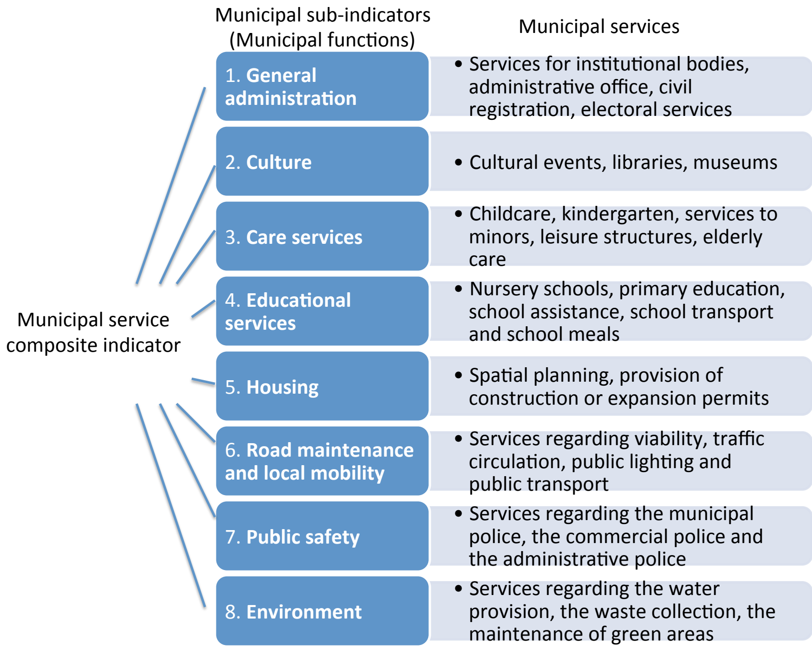

Figure 1: Example of services for each municipal function

Once the sub-indicators are identified with the municipal functions, the representative variables need to be selected consequently, in compliance with the data availability and the output choice in the related local governments efficiency literature (for an extensive review, see Narbón-Perpiñá and De Witte, 2018a). We focus on the Flemish region of Belgium for which we are in possession of exceptionally good data at municipal level (see Appendix A for a discussion). Moreover, in their review of local competences, CEMR (2016) indicates that the categories indicated in Figure 1 also hold for Flanders. The data cover all 307 Flemish municipalities over the period 2006–2011, mostly coinciding with the term of local authorities elected in 2006.[1] Table 1 shows the descriptive statistics of the variables expressed in per capita values. We proxy the general administration by the number of net foreigners and households. The first component, measured as the number of incoming minus leaving migrants, captures the administrative workload related to immigration offices. The variable was also used before by D'Inverno et al. (2018) to proxy the need for welfare institutions. The number of households captures the general administrative, statistical and permit services needed in a municipality. The measure has been used since the pioneering work of Eeckaut et al. (1993), and became more standard since the work of De Borger and Kerstens (1996a,b). Although the combination of the two variables underestimates the full general administration required in a municipality, it is reasonable to assume that other general

[1]There are 308 municipalities in Flanders, but one municipality is excluded because of lack of data.

administrative services are strongly correlated with the number of net foreigners and households. The cultural events measure the cultural activities promoted by the local government and they can be seen as the cultural engagement boosted by a local administration. The number of cultural events in the municipality are measured as all registered events in a national and popular database for cultural events (UIT-databank). Measuring the events by using a count indicator might hide some heterogeneity in the size of the events. Although similar information is not available in Flanders, Narbón-Perpiñá and De Witte (2018a) list four earlier papers that use the expenses for theatres, cinemas, municipal museums and galleries as a proxy for cultural facilities in municipalities. Except for Prieto and Zofio (2001), the literature does not account for quality differences in the facilities. The recipient of the education and care services are respectively the students in primary school and the children in kindergarten together with the residents over 80. Despite the fact that they are not direct measures of the actual services as much as the number of meals per pupil or residential aged care might be, it is reasonable to assume that the chosen groups well represent the recipient of this kind of services. Similar assumptions have also been made in earlier literature. For example, Nikolov and Hrovatin (2013) and D'Inverno et al. (2018) consider the population until 5 years old as a proxy for kindergarten services. Asatryan and De Witte (2015) use the ratio of the number of children at kindergartens to population. The elderly patient population has been used before by Loikkanen and Susiluoto (2005) and Asatryan and De Witte (2015) to measure the supply for social services to elderly, such as geriatric and retirement homes, or the general municipal assistance for the elder. The built-up area is considered for the housing and country planning area, as a variable capturing the effort of a municipality in fostering its urban development. It is closely related to municipal service provision as municipalities with a larger built-up area might have a different cost structure. Earlier papers using this proxy include Lo Storto (2016) and Arcelus et al. (2015). However, as it serves only as an indirect proxy due to the difficulty to quantify the supply of public services, other papers have used the number of households in the local area as a proxy for the demand of the urban services. For the road mobility, we consider the number of accidents as a proxy of the traffic circulation interventions when an accident occurs and the related necessary local road maintenance. This is an innovative way as earlier research typically uses the length of municipal roads (Narbón-Perpiñá and De Witte, 2018a). However, specifically in Flanders, accidents captures better the quality of the public provision. For the police function, we keep track of the local police intervention and prevention activities by means of police records. Specifically, the crime level includes the number of thefts, physical and property crimes. This variable has been used before by Eeckaut et al. (1993) and Moore et al. (2005). It serves as a better proxy for the quality of service provision than the expenditures for police or a dummy variable for municipal policy also used before. Finally, one of the main tasks of a local government is to promote local strategies to develop a sustainable environmental attitude among the citizens (Global Taskforce of Local and Regional Governments, 2016). In this perspective, the energy consumption and the waste production can be good indicators of the local effort in the field of energy efficiency and waste reduction. These are the variables collected for the present application. Nevertheless, the proposed approach is flexible to encompass different performance indicators, as long as they are representative of the municipal service provision, adapting to the data availability and country-specific local competences.

Table 1: Descriptive statistics for the municipal service level

Municipal service (per 1000 inhabitants)	Obs	Mean	St. Dev.	Min	Max
1. Robust BOD CI for general administration	1842	0.85	0.11	0.73	2.92
Net foreigners	1842	7.44	6.31	0.04	120.8
Households	1842	404.5	23.76	351.33	560.83
2. Cultural events	1842	6.53	5.51	0.3	36.05
3. Robust BOD CI for care service	1842	0.8	0.11	0.56	1.54
Children in kindergarten	1842	37.35	7.37	8.38	96.64
Residents over 80	1842	46.26	9.65	20.19	88.28
4. Students in primary school	1842	63.51	13.47	9.81	145.58
5. Built-up area (Km^2)	1842	64.33	21.2	22.34	251.01
6. Accidents	1842	4.44	1.57	0.39	13.37
7. Robust BOD CI for Crime	1842	0.5	0.21	0.11	1.62
Thefts	1842	21.77	12.42	3.44	92.79
Physical crimes	1842	4.91	2.44	0.9	19.56
Property crimes	1842	8.43	3.56	0	34.01
8. Robust BOD CI for environment services	1842	0.61	0.12	0.41	1.46
Waste ($Tonnes$)	1842	141.08	39.51	56.83	362.63
Energy consumption	1842	7.9	1.8	4.68	22.68

Note: Panel for 307 Flemish municipalities over 2006–2011.

A few additional considerations deserve to be mentioned. A recurrent issue in the non-parametric analysis is the curse of dimensionality, which is likely to occur when the model consists of numerous variables. To exploit the data availability and to gather together several aspects of the same phenomenon, the literature proposes a few solutions such as aggregating first in sub-indicators (Afonso et al., 2005) or specifying the weight restrictions on a more aggregate level (Morais and Camanho, 2011). In this line, an overall composite indicator could be considered imposing both weight restrictions at the level of the dimensions and at the level of the underlying tasks for relatively low dimensional analysis. However, given the increasing attention at transparent administration together with the need of monitoring and evaluating the local services, there will be more and more data available that can be used in the proposed municipal composite indicator. Accordingly, to avoid the curse of dimensionality and to keep separated each area, we suggest aggregating first the sub-indicators and then aggregating at the overall level as a reasonable alternative way to approach the multidimensional municipal service provision assessment. Specifically, for some municipal sub-indicators with data covering multiple underlying tasks, a 'robust Benefit-of-Doubt' composite indicator has been constructed aggregating variables belonging to the same municipal function (for the methodological details, please refer to Section 3). [2] This has been done for the following functions: the general administration (composite indicator from Net foreigners and Households), the care services (composite indicator from Children in kindergarten and Residents over 80), the public safety by a crime index (composite indicator from number of Thefts, Physical and Property crimes [3]), the environmental management (composite indicator from Waste and Energy consumption).

Second, not all the variables included in the analysis are strictly direct measures of the services provided to the citizens, but are rather proxies. This procedure is widely accepted both in the composite indicator and in the local governments' efficiency literature to the extent that the selected variables are clearly representative of the intended composite indicator (OECD, 2008; Narbón-Perpiñá and De Witte, 2018a). Moreover, to the best of our knowledge, there are no available data on the quality of the services and on the role of citizens in co production, even though this type of information would add a very interesting dimension to the overall analysis, as pointed out for example by De Witte and Geys (2011, 2013).

Finally, it is worth pointing out that "more is better" is not true for all the municipal sub-indicators, as not all of them represent desirable features. In fact, the municipalities are not merely required to offer the greatest amount of services that they can, but, in some cases, the 'service' that they are supposed to deliver have to be aimed at alleviation of the production of undesirable features, such as the number of accidents, the level of criminality and the level of environmental pollution to the maximum extent possible. To keep the production of these undesirable indicators as low as possible, municipalities have to spend resources that would have otherwise spent in producing other services. In a similar vein, a local government has to spend

[2]To aggregate variables belonging to the same municipal function we use the robust BoD-model, and not the robust and conditional BoD-model (as we do in the macro-aggregation). Although the operating environment influences directly the macro-aggregation across different municipal functions, reflecting different preferences over different municipal functions and therefore different budget allocation, this does not necessarily hold within the dimension. For example, if municipalities provide care services, they have to do it in the best way irrespective of the age and structure of the population. On the contrary, the heterogeneity coming from the operating context must be dealt with to grasp possible synergies and trade-offs across the different dimensions.

[3]Providing a BoD index for crime rather than just adding numbers captures more information: it takes into account the heterogeneity among the different kinds of crime across the municipalities that would be otherwise wiped out.

resources to keep the roads safe so to minimize the number of accidents and they have to pay the public officers for protection and assistance offered by them to the citizens. The introduced "undesirable" outcomes might reflect to some extent the quality of services (e.g., accidents might partly reflect poor-quality roads). In the current analysis we consider three undesirable and five desirable indicators, with the former referring to public safety, road mobility and environmental management functions and the latter referring to general administration, education and care services, culture and housing functions.

In addition to the considerations acknowledged so far, it is necessary to recognize that the operating conditions are also of relevance when constructing a municipal service composite indicator. In fact, the characteristics of the municipalities (e.g., size, income, age composition) affect local public activities and, as a consequence, they also affect the overall assessment of the aggregate indicator. For this reason, in compliance with the variables used in the related literature (for an extensive review, see Narbón-Perpiñá and De Witte, 2018b), three groups of background variables need to be included in the analysis: economic-financial characteristics, socio-demographic structure and the political dimension. We distinguish between variables that can be considered as exogenous only in a relatively short term and variables that can be considered exogenous both in a short and longer term (Rogge et al., 2017). In the former category we include *fiscal income*, *financial debt* and *unemployment* as representative variables of economic-financial characteristics: specifically, they are also informative of the institutional setting in which municipalities have to operate and to some extent influence through policies over time. The fiscal income is defined as the income per capita and it represents citizens' economic level estimated for each municipality. Even if local authorities cannot have direct control over their citizens wealth, they might affect it by fiscal policies in the long run. Financial debt is measured as the excess of expenditures over revenues per capita and it reflects three interconnected aspects: namely the extent of granted loans, the return on investment and the fiscal revenue capacity. As stock of deficit, it might be due to decisions rooted in the past, but still it might influence the current service provision. Accordingly, this variable can be seen as exogenous in the short term, as well as subject to political influence in a longer period. The unemployment variable is defined as the percentage of unemployed residents between 15 and 64 years over the total working population and it can be seen not only as a cost for the municipality in terms of social and housing benefits, but also as a signal of the living conditions in which municipalities have to operate. In this perspective, there is not much leeway for the local authorities other than promoting policies that can somehow lead to a change over the years.

In the latter category, we include characteristics that can be considered exogenous irrespective of the time span, namely the socio-demographic structure and the political aspect. It is reasonable to look at the population composition as given and not under the discretion of the local authorities. Likewise, the political ideology can be seen as rather exogenous (especially if we keep in mind that the period under analysis falls within an electoral cycle). To frame the socio-demographic structure, we consider the *residents over 65*, the *foreigners* and the *population growth*. The residents over 65 years of age represents a share of retired people over the total population and it represents the age composition. To capture the ways in which a municipality is attractive for foreigners and to present its ethnic composition, the share of immigrants has been considered. Population growth is the variation of residents in a municipality over the years. The local governments should adjust their service level in compliance with the growth of its population in order to avoid disruption or excess provision, and ultimately waste of resources. Finally, concerning the political aspect, the "*Ideological Complexion of the local Government*" (ICG) measures the ideological stance of the local government on a Left-Right scale (from 0 to 10): a higher ICG score represents a more right-wing government. We refer to Table 2 for the descriptive statistics of the presented background variables and Appendix A for additional information on data source and description. [4]

Table 2: Descriptive statistics for the municipal background conditions

Background conditions	Obs	Mean	St. Dev.	Min	Max
Economic-financial components					
Fiscal income (€ per capita)	1842	16330.52	1924.97	11055.29	24278.23
Financial debt (€ per capita)	1842	1014.43	581.88	-1497.33	5402.22
Unemployment	1842	5.52	1.8	2.11	15.19
Socio-demographic components					
Residents over 65 (% of total)	1842	0.18	0.02	0.11	0.3
Foreigners (% of total)	1842	0.05	0.06	0	0.48
Population growth	1842	0.64	0.62	-4.51	3.59
Political component					
Ideological Complexion of the local Government (ICG)	1842	5.04	0.71	2.5	6.3

Note: Panel for 307 Flemish municipalities over 2006–2011.

3. Methodology

This paper proposes an innovative way to measure the local municipal service provision by the construction of a composite indicator. Considering the issues that are required to be addressed while specifying the model, the methodological steps leading to the proposed 'robust conditional directional distance BoD model with weight restrictions' are presented in the subsections under-

[4] Some of the background variables are categorized for methodological reasons (for more technical details, see Rogge et al., 2017, and the references therein).

neath.

3.1. The BoD model

As presented in the previous section, the local municipal service provision covers several areas. Accordingly, we look for an aggregating method to group several dimensions into one single composite indicator (CI). As we do not have any prior knowledge of the functional form and understanding of the importance of the different municipal services, we adopt a fully non-parametric way to avoid any kind of specification bias. In particular, we choose a Benefit-of-the-Doubt (BoD) weighting technique, inspired by the Data Envelopment Analysis (DEA) methodology (Charnes et al., 1978) and labelled as such after Melyn and Moesen (1991). The peculiarity of the BoD model is that it assigns the weights for each municipal service endogenously (Van Puyenbroeck, 2018). More specifically, the service provision level of the municipality under analysis is compared in a relative perspective to the service level of all the municipalities in the sample: a higher weight is assigned to a municipal area where the municipality under analysis provides relatively high service level and a lower weight where it provides relatively low service level. The optimal weights are determined in such a way that the composite indicator for the overall level of service provision of the municipality j_0 under analysis is maximized. The weights are obtained by solving for each municipality j_0 the following problem:

$$\begin{array}{lll} CI_{j_0} = \max & \sum_{r=1}^{s} y_{rj_0} w_{rj_0} & \\ \text{s.t.} & \sum_{r=1}^{s} y_{rj} w_{rj_0} \leq 1, & \text{for } j = 1, \ldots, j_0, \ldots, n \\ & w_{rj_0} \geq 0, & \text{for } r = 1, \ldots, s \end{array} \tag{1}$$

where CI_{j_0} refers to the composite indicator optimal value for the evaluated municipality j_0; y_{rj_0} denotes the observed service level for the municipal area r of the evaluated municipality j_0; w_{rj_0} represents the most favourable weight to the municipal area r for the evaluated municipality j_0; y_{rj} denotes the observed service level for the municipal area r of every municipality j in the dataset; n is the number of municipalities under analysis (n=307) and s signifies the number of municipal functions considered in this application (s=8).

The first constraint in the model formulation is referred to as the "normalization" constraint: the overall municipal composite indicator CIj_0 is maximized subject to an upper bound equal to one. Therefore, the CIj_0 value ranges between zero and one where the higher the value, the higher is the overall service provision level for the evaluated municipality. If CIj_0 <1, it means that, even if evaluating the municipality under analysis by employing its most favourable weighting system, there is at least another municipality providing a higher overall level of service. Hence, it would mean that there is still room for improvement in the service provision, given the observed overall level of provided services across the whole sample. If CIj_0 =1, then it denotes that the municipality under analysis is not outperformed in terms of the overall service provision and it is considered as its own benchmark while using its most favourable weight system. The second constraint imposes the weights' non-negativity.

The advantage of using this approach is twofold: first, it allows to group together several aspects into one single indicator. Second, it ensures the fairness of the comparison, weighting more the municipal areas where higher priority is devoted and, vice versa, weighting less the ones with lower priority. In this way, each evaluated municipality is granted the "Benefit-of-the-Doubt" in the assessment and the fairness of the comparison is ensured (for more details on the BoD approach, see e.g. Cherchye et al., 2007; Verschelde and Rogge, 2012; Rogge et al., 2017).

3.2. The directional distance BoD model

In the depicted BoD framework, a higher indicator level in a certain municipal area denotes a better overall service provision assessment, i.e. the extent to which the indicator can be labelled as "desirable". However, it must be acknowledged that this might not be the case among all the local services. In fact, municipalities might also provide services in areas where the best they can do is to limit the production of the indicator rather than to expand it and for this reason the label "undesirable" is assigned.

The inclusion of undesirable features in the construction of composite indicators is quite recent and it is associated with the performance measurement literature (for an extensive review, see Zanella et al., 2015b; Dakpo et al., 2016). In this study, we propose the model introduced by Zanella et al. (2015b) and advocated by Rogge et al. (2017), namely a directional distance BoD model. This model combines the earlier listed advantages of the BoD approach together with the ones of the directional distance function, introduced by Chung et al. (1997). In fact, the directional distance model allows to simultaneously contract the undesirable indicators and expand the desirable ones along a specified direction vector $g = (-g_b, g_y)$, as shown in its primal formulation (Zanella et al., 2015b, model (7), p.523). However, the multiplier formulation of the directional distance BoD model (Zanella et al., 2015b, model (8), p.523) is preferred as it provides for the inclusion of weight restrictions in the municipal service level assessment and it has to be solved for each j_0 municipality under analysis as follows:

$$
\begin{array}{lll}
\beta_{j_0} = \min & -\sum_{r=1}^{s} y_{rj_0} u_{rj_0} + \sum_{k=1}^{l} b_{kj_0} p_{kj_0} + v_{j_0} & \\
\text{s.t.} & \sum_{r=1}^{s} g_y u_{rj_0} + \sum_{k=1}^{l} g_b p_{kj_0} = 1 & \\
& -\sum_{r=1}^{s} y_{rj} u_{rj_0} + \sum_{k=1}^{l} b_{kj} p_{kj_0} + v_{j_0} \geq 0 & \text{for } j = 1, \ldots, j_0, \ldots, n \\
& u_{rj_0} \geq 0 & \text{for } r = 1, \ldots, s \\
& p_{kj_0} \geq 0 & \text{for } k = 1, \ldots, l \\
& v_{j_0} \in \Re &
\end{array}
\tag{2}
$$

where β_{j_0} denotes the optimal value for the evaluated municipality j_0; y_{rj_0} and b_{kj_0} respectively refer to the observed r desirable and k undesirable indicator of the evaluated municipality j_0; u_{rj_0} and p_{kj_0} respectively represent the most favourable BoD-weights for the r desirable and k undesirable indicator for the evaluated municipality j_0; y_{rj} and b_{kj} respectively refer to the r desirable and k undesirable indicator of every municipality j in the dataset; n is the number of municipalities under analysis (n=307); s and l respectively signify the number of municipal functions linked to desirable and undesirable indicators considered in this application (s=5 and l =3); v_0 comes from the equality constraint in the primal formulation and should not be associated with a variable returns to scale model, but rather to the presence of performance indicators expressed as ratios (Zanella et al., 2015b).

The direction vector choice is important as it might impact the results. Several solutions have been proposed in the literature depending on the objectives pursued (see for example Rogge et al., 2017, for a discussion on different direction values and formats). For the proposed municipal service provision composite indicator we choose $g = (-b_{kj_0}, y_{rj_0})$. In other words, we deploy the municipal service indicators of the evaluated municipality as the direction vector. By specifying this direction, each municipality follows its own specific path for improvements. This ensures a high level of flexibility and a proportional interpretation of the improvements, in addition to preserving directional distance model units invariance (Zanella et al., 2015a; Rogge et al., 2017). The composite indicator for the municipal service provision is obtained as

$$CI_{j_0} = 1/(1 + \beta_{j_0})$$

and it ranges between zero and one, where one denotes the greatest level of service provision in line with the basic BoD model. [5]

3.3. The directional distance BoD model including weight restrictions

In the local service provision assessment, there is another aspect that cannot be ignored, i.e. the political preferences over the different municipal intervention areas. There are two interconnected explanations for this kind of heterogeneity among the municipalities. First, certain municipal functions require higher priorities than others. This phenomenon is not only quite evident looking at the average expenditure composition across the municipalities, but it is also clearly stated in certain national legislative systems (for example in Italy there is the distinction between "fundamental" and other functions). Second, every municipality has its own peculiar vocation which means that, for instance, a municipality might be more aligned towards the tourism

[5] Unlike model (1), the optimal value of model (2) does not range between 0 and 1.

sector, another one on cultural activities, or on some other economic specialization. In this case, the budget allocation reflects this variety across the municipalities under evaluation.

By including weight restrictions in our model formulation, apart from the inclusion of these value judgments, we can also address one common concern related to greater flexibility in the weighting system associated to the BoD approach. In the DEA/BoD literature several types of weight restrictions have been considered (for a review, see for all Sarrico and Dyson, 2004; Cherchye et al., 2007; Zanella et al., 2015b, and references therein). In this context, we suggest the assurance region type I (ARI) weight restrictions as suggested by Zanella et al. (2015b) and advocated by Calabria et al. (2018).[6] By adding this kind of restrictions to the model specification, we can constrain the relative importance of each municipal function indicator within a certain range and express it in percentage terms, as follows:

$$\begin{cases} \phi_r \leq \dfrac{u_{rj_0}\bar{y}_r}{\sum_{r=1}^{s} u_{rj_0}\bar{y}_r + \sum_{k=1}^{l} p_{kj_0}\bar{b}_k} \leq \psi_r & \text{for } r = 1, \ldots, s \\ \phi_k \leq \dfrac{p_{kj_0}\bar{b}_k}{\sum_{r=1}^{s} u_{rj_0}\bar{y}_r + \sum_{k=1}^{l} p_{kj_0}\bar{b}_k} \leq \psi_k & \text{for } k = 1, \ldots, l \end{cases} \tag{3}$$

where s and l constraints are used respectively for each observed r desirable and k undesirable municipal indicator.

The question remains on how to specify the importance of each municipal indicator and accordingly the bounds ϕ and ψ. The innovative way we propose in this paper is to get this kind of information directly from the municipal expenditure allocation across the different services. To the best of our knowledge, in municipal performance assessment the expenditure composition has been included directly in the aggregation process, but not in the weight restrictions (see for example Bosch et al., 2012; Helland and Sørensen, 2015; D'Inverno et al., 2018). In this regard, the expenditure composition of each municipality resembles to a certain extent the rationale of the "budget allocation" approach, as described by Cherchye et al. (2007). In Cherchye et al. (2007) the budget allocation procedure directly asks stakeholders/experts/policy makers to indicate their viewpoint on the task/dimension importance by distributing 100 points and reflects therefore their preferences. In this sense, this is similar to our approach, as the budget constraints provide an indication of the political and local preferences while accounting for some path dependencies.

The proposed method has the advantage to reflect the heterogeneity across municipalities, granting some leeway but, at the same time, leaving an objective order of importance among the municipal services without imposing any kind of external judgement. Specifically, we propose

[6] As suggested by Sarrico and Dyson (2004), this kind of choice might penalize the units with small or large values. However, as emphasized by Zanella et al. (2015b, p. 526), among other weight restriction alternatives the ARI type is "the best option to construct composite indicators and ranks", so to ensure a fair comparison among the units under evaluation.

three sets of restrictions, which vary according to the different specified bounds.

The first one considers the minimum and the maximum share of expenditure in each municipal area across all the municipalities ("*MinMax* restrictions"). In this way, the municipality under evaluation cannot assign lower or greater importance to each municipal indicator than the one recognized among all the municipalities.

In a second way of specifying the restrictions, for each municipal indicator, the average spending share is considered, identifying a lower and an upper bound value equal to its $\pm 50\%$ ("*Average* restrictions"). This kind of restrictions circumscribes the average importance of each municipal area according to the priorities acknowledged among all the municipalities: local governments are given some leeway in deciding their own weights, but at the same time a certain order of importance among the functions is respected.

Finally, rather than confining a municipality within the overall average choice, the third specification of restrictions allows each municipality to set its own weight based on its current spending allocation ("*Municipal-specific* restrictions"). In other words, the lower and the upper bound value of the constrains associated to each municipal indicator is equal to $\pm 50\%$ of each municipal-specific expenditure share. In this third scenario, given the municipal-specific lower and upper bound, each unit is evaluated against a DMU-specific frontier, as it occurs in the virtual weight restrictions case. Specifically, in the virtual weight restrictions the DMU-specific feature enters because of the unit-specific observations, while in the weight restrictions we proposed in the bounds. Even if this system prevents the units to be evaluated against a unique frontier, it is still informative to give the municipalities some more leeway in deciding their own weights while assessing the extent of their possible performance improvement, keeping though as an ultimate reference the two scenarios with the ARI restrictions. Futhermore, as pointed out in Oliveira et al. (2019), as long as the weights are DMU-specific, the DMUs are ultimately not evaluated on a common ground to rank all the units from a global perspective, even if this still allows a distinction between the best-performing units from the ones that are not. For this reason they propose a goal-programming model to identify a common set of weights that can be used to rank the units (the firms) at a more aggregated level (the industry). Following these lines, we suggest as scope for further extension the identification of a common set of weights to properly allow a ranking of the municipalities at a higher government level (e.g. regional or national level).

By construction, the three sets exhibit increasingly binding restrictions. Table 3 presents summary information about the weights for each municipal area.

Interestingly, the optimally chosen weights not only reflect the importance that each municipal indicator has in the overall assessment of service level provision, but they can also be interpreted as normalized shadow prices (Coelli et al., 2005). In our application, a shadow price can present

Table 3: Summary of the weights obtained from the municipal expenditure composition

	Administration	Culture	Care services	Education	Housing	Local mobility	Security	Environment
MIN	0.08	0.02	0.01	0	0.01	0.03	0.04	0.03
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
MAX	0.49	0.31	0.23	0.39	0.16	0.33	0.23	0.22

the way a municipal indicator is affected whenever another indicator varies, or alternatively, the impact of different political choices on the composition of overall service provision (for more technical details, see Grupp and Schubert, 2010; Fusco, 2015). The shadow prices are useful to determine the "budget shares" (Van Puyenbroeck and Rogge, 2017).

3.4. The robust and conditional directional distance BoD model

All the steps discussed so far are imperative for ensuring an increasing level of fairness in the local service provision assessment. However, there is a last aspect that should not be overlooked, namely, the role of the operating context under which the municipalities are required to function. First of all, the background conditions can affect both the supply and the demand side of the service provision level. For example, concerning the supply aspect, a wealthier municipality with a higher level of local revenues should be endowed with more resources to spend. On the other hand, considering the demand side, a municipality experiencing a higher level of unemployment might be required to provide a substantial number of subsidies which would lead to a diversion of resources from the provision of additional services. Moreover, background conditions can also have a significant impact on the political preferences over the municipal functions, influencing the components of the composite indicator to different extents and the way they enter in the synthetic index. Consequently, the composite indicator CI_{j_0} outlined so far has to be adjusted in a manner that it accounts for the differences in the municipal environmental variables. These can be grouped into three categories with respect to the differences in variables, namely, economic-financial characteristics, socio-demographic structure and political dimension.

Despite the fact that the operating factors are exogenous with respect to the service provision level and they are not under the control of local policy-makers, they do not merely affect the distribution of the composite indicator scores but also their attainable set. For these reasons, the "separability condition" cannot be assumed and one-stage procedure is needed to compute the municipal service provision composite indicator taking environmental factors into consideration simultaneously. In the literature, this approach is referred to as the "conditional" measurement procedure. Moreover, the conditional analysis is performed in adjunction with its robust version to mitigate the influence of outlying observations, arising from, e.g., measurement errors and atypical observations, using the insights from the "order-m" approach. For the sake of brevity, we refer for a more formal and extensive explanation of the procedures to Cazals et al. (2002); Daraio and Simar (2005, 2007); De Witte and Kortelainen (2013); Cordero et al. (2017b), among others.

For the computation of the robust municipal service composite indicator, we execute a Monte-Carlo algorithm performing B computation rounds (where B is large) to lessen the impact of the outlying observations. In each b round ($b = 1, \ldots, B$), first m municipalities are drawn with replacement from the original sample of n units and then the m-sample 'directional distance BoD with ARI restrictions' composite indicator $CI_{j_0}^{b,m}$ is computed. Finally, the robust composite indicator $CI_{j_0}^{m}$ is obtained as the arithmetic average of the B $CI_{j_0}^{b,m}$, as follows:

$$CI_{j_0}^{m} = \frac{1}{B}\sum_{b=1}^{B} CI_{j_0}^{b,m}$$

Due to the subsampling, the municipality under evaluation might not be included in its own reference set and be accounted as super-performing unit (De Witte and Schiltz, 2018). Accordingly, $CI_{j_0}^{m,b}$ and possibly $CI_{j_0}^{m}$ might be larger than one. This can be construed as the municipality j_0 under evaluation is providing a higher service level than the average m municipalities it has been compared with as its reference sample.

Also, to consider the heterogeneity among the municipalities captured by the z background variables, the Monte-Carlo simulation procedure is the same with certain changes with respect to the drawing process. The m municipalities are drawn with replacement and with a particular probability based on an estimated kernel density function. The idea is to draw m municipalities with a higher probability of being similar to the municipality j_0 under evaluation (and lower probability of being dissimilar): in this way, the municipal service provision level composite indicator

$$CI_{j_0}^{m,z} = \frac{1}{B}\sum_{b=1}^{B} CI_{j_0}^{b,m,z}$$

is assessed considering similar background conditions and ensuring a greater level of fairness while comparing different municipalities under analysis. In this case, a $CI_{j_0}^{m,z}$ score larger than one signifies that the municipality j_0 under evaluation is providing a higher service level than the average m municipalities with similar background characteristics, while $CI_{j_0}^{m,z} = 1$ denotes a similar service provision level.

Finally, as we intend to extend the model to a dynamic framework to exploit intertemporal variations in municipal service provision (Cordero et al., 2017a), we further adjust the robust conditional composite indicator $CI_{j_0}^{m,z}$ by including the time dimension, according to the insights of the approach proposed by Mastromarco and Simar (2015). Accordingly, the composite indicator $CI_{j}^{m,z,t}$ is computed over all the combinations of municipality $j = 1 \ldots, j_0, \ldots, n$ and time period $t = 1 \ldots T$ (where $T = 6$ in our application) and time is also included as an additional conditioning variable together with z.

Furthermore, the computation of the robust unconditional and conditional composite indicators provides two more additional useful insights. Through a non-parametric statistical inference, we can detect whether the environmental variables are on average statistically significant with respect to the composite indicator scores and the direction of the influence of the environmental variables on the service provision level assessment can be determined. More specifically, the ratio of the unconditional CI_j^m and conditional $CI_j^{m,z,t}$ composite indicator scores can be non-parametrically regressed on the external variables (Rogge et al., 2017). By construction, the unconditional CI score is at most (less than or equal to) the conditional CI score. Accordingly, a positive coefficient denotes a favourable influence of a contextual variable on the service provision level score: when the variable increases, the conditional CI score gets closer to the unconditional one. For a negative coefficient, the opposite holds: the contextual variable has an unfavourable influence on the service provision level assessment, as the conditional CI score increases when the variable increases. In other words, the background condition acts as an unfavourable context if, as the variable increases, its CI score increases only because evaluated among similar municipalities and the service provision level it can afford is lower compared with the one of municipalities facing different background conditions. For the sake of brevity, we omit further technical and theoretical details: we refer the interested reader to Daraio and Simar (2007); Bădin et al. (2012).

3.5. Alternative model specifications to handle undesirable features

Before moving on to the results, we conclude this methodological section by outlining alternative model specifications to the directional distance BoD model introduced in section 3.2 as recently suggested by other scholars, to deal with the presence of undesirable features in the construction of a composite indicator.

First, we depart from the directional distance formulation and we move along the lines traced by Färe et al. (2019). Specifically, we consider their model (3) p. 396, so that we can account for the municipal expenditure composition by adding the weight restrictions. We adapt the notation to facilitate the comparison of the models.

$$
\begin{aligned}
CI_{j_0} = \quad & \max \textstyle\sum_{r=1}^{s} y_{rj_0} u_{rj_0} - \sum_{k=1}^{l} b_{kj_0} p_{kj_0} \\
\text{s.t.} \quad & \textstyle\sum_{r=1}^{s} y_{rj} u_{rj_0} - \sum_{k=1}^{l} b_{kj} p_{kj_0} \leq 1 \\
& \qquad\qquad\qquad \text{for } j = 1, \ldots, j_0, \ldots, n \\
& u_{rj_0} \geq 0 \quad \text{for } r = 1, \ldots, s \\
& p_{kj_0} \geq 0 \quad \text{for } k = 1, \ldots, l
\end{aligned}
$$

where y_{rj_0} and b_{kj_0}, respectively, refer to the observed r forward (desirable) and k reverse (undesirable) indicator of the evaluated municipality j_0; u_{rj_0} and p_{kj_0}, respectively, represent the most favourable BoD-weights for the r forward and k reverse indicator for the evaluated municipality j_0; y_{rj} and b_{kj} respectively refer to the r forward and k reverse indicator of every municipality j in the dataset; n is the number of municipalities under analysis (n=307); s and l respectively signify the number of municipal functions linked to forward and reverse indicators considered in this application (s=5 and l =3).

Second, we consider a variant of the directional distance formulation. Specifically, we replace the equality constraint $\sum_{j=1}^{n} \lambda_j = 1$ in the CI model (7, p.523) with an inequality constraint$\sum_{j=1}^{n} \lambda_j \leq 1$. As a result, v_{j_0} is not free anymore and the multiplier formulation trivially follows. We refer to Appendix E.1 for a more extensive discussion of these two models and the comparison with the directional distance formulation.

4. Results

In this section, we present the results of the estimated robust conditional municipal service composite indicator, for different weight restriction specifications (*MinMax*, *Average* and *Municipal-specific*) as presented in section 3.3, on a sample of 307 Flemish municipalities over the years 2006–2011. We estimate different conditional models, depending on the group of background variables as introduced in section 2. Model 1 consists of the economic and financial characteristics that might affect the delivery of municipal services, namely the level of fiscal income, the level of financial debt and the unemployment rate. In Model 2, the socio-demographic structure is also added, by the inclusion of the share of elderly people, the share of foreigners and the municipal population growth. The political component is considered in adjunction with the economic and socio-demographic characteristics in Model 3, by using the Ideological Complexion of the local Government (ICG). Moreover, in every model specification, a year dummy is also included to run the analysis in a dynamic framework. For the sake of comparison, the unrestricted unconditional, the unconditional and the robust unconditional models are also estimated. [7] Table 4 presents the descriptive statistics of the estimated composite indicator results. [8]

[7] After a sensitivity analysis for the choice of m (m=10, 20, ..., 100), we choose m=40 for which there is a remarkable decrease of the super-performing municipalities. As for the bootstrap replications, we consider B=2000. For the municipal sub-indicators as introduced in Section 2, the robust BoD composite indicator has been computed for every year separately.

[8] In Appendix B the Spearman's rank correlation coefficients across the different weight restriction specifications and model distributions are reported to further investigate the impact on the ranking and on the best practices among the municipalities under evaluation.

Table 4: Descriptive statistics of the service provision composite indicator scores estimated for 307 municipalities over 2006–2011

	Mean	St. Dev.	Min	Q1	Median	Q3	Max
Unrestricted Uncondiitonal	0.8388	0.0643	0.6357	0.7945	0.8293	0.8751	1.0000
Unconditional							
MinMax restrictions	0.7832	0.0670	0.6302	0.7387	0.7719	0.8196	1.0000
Average restrictions	0.7178	0.0649	0.5883	0.6758	0.7031	0.7427	1.0000
Municipal-specific restrictions	0.7067	0.0723	0.5121	0.6607	0.6996	0.7437	1.0000
Robust Unconditional							
MinMax restrictions	0.9618	0.1098	0.8008	0.8985	0.9336	0.9919	2.0990
Average restrictions	0.8752	0.0936	0.7002	0.8163	0.8536	0.9051	1.5770
Municipal-specific restrictions	0.8662	0.1041	0.6652	0.8045	0.8470	0.9030	1.8767
Robust Conditional Model 1							
MinMax restrictions	0.9753	0.0315	0.8007	0.9608	0.9895	0.9992	1.0013
Average restrictions	0.9215	0.0611	0.7034	0.8788	0.9263	0.9776	1.0004
Municipal-specific restrictions	0.9158	0.0672	0.6571	0.8737	0.9230	0.9756	1.0003
Robust Conditional Model 2							
MinMax restrictions	0.9969	0.0093	0.8886	0.9985	1.0000	1.0000	1.0000
Average restrictions	0.9846	0.0275	0.7769	0.9817	0.9967	0.9999	1.0000
Municipal-specific restrictions	0.9831	0.0312	0.7132	0.9815	0.9965	0.9999	1.0000
Robust Conditional Model 3							
MinMax restrictions	0.9985	0.0069	0.9072	1.0000	1.0000	1.0000	1.0000
Average restrictions	0.9905	0.0248	0.7729	0.9973	1.0000	1.0000	1.0000
Municipal-specific restrictions	0.9892	0.0298	0.6848	0.9966	1.0000	1.0000	1.0000

Note: Unrestricted indicates the absence of weight restrictions. *MinMax* restrictions refer to the minimum and maximum share of expenditure in each municipal area across all the municipalities. *Average* restrictions consider the average spending share (lower and upper bound equal to its ±50%). *Municipal-specific* restrictions are based on the municipal-specific current spending allocation (lower and upper bound equal to the ±50% of each municipal spending share).

Model 1 includes the economic and financial characteristics (Fiscal income, Financial debt and Unemployment). *Model 2* adds to the economic and financial characteristics the socio-demographic structure (Share of elderly people, Share of foreigners and Population growth). *Model 3* adds to the economic-financial and socio-demographic variables the political component (Ideological Complexion of the local Government). In every conditional model specification a year dummy is also included.

The results can be explored along two complementary dimensions. The first one is related to the use of the weight restrictions. It is not surprising that the inclusion of the weight restrictions lowers the values of the composite indicators with respect to the unrestricted model: every municipality under analysis is forced to choose its own optimal system of weights only within a certain range. Moreover, for each model specification the three different sets of weight restrictions lead to a lower average service provision. As pointed in section 3.3, they are by construction increasingly binding. The inclusion of the information on the expenditure composition has a role to play in the composite indicator estimation through alternative weight specifications. In addition, further information can be retrieved from the shadow prices, as they are useful to determine the "budget shares" (Van Puyenbroeck and Rogge, 2017). In particular, it may be observed that imposing weight restrictions provides shadow prices that are closer to the current composition. We refer to Appendix C for the shadow price and the budget share results.

Next, we gradually change the assumptions in the model. Without loss of generality, we provide critical discussion of the results of the *MinMax* specification in what follows. If we look at the overall picture by considering the *Unconditional* specification, it seems that there is room for municipal service improvement on average of 22% (i.e. 1-0.7832). However, when considering the outliers and/or some atypical observations by means of the *Robust* specification, then this room for improvement shrinks up to 4% (i.e. 1-0.9618), although this average is boosted because of the influence of outlying observations. [9] This is even stronger if the *Conditional* specifications are considered, where we compare more 'like with likes'. Accordingly we can see that, when comparing more similar municipalities, there is no major scope of improvement on average: this leads us to the conclusion that each municipality under analysis is almost producing what other similar municipalities are also providing in terms of services. This evidence shows the methodological importance of such an integrated analysis: the atypical observations and the background variables do affect the composite indicator. Specifically, in the present application when considering the operating environment each municipality has to operate in, it is observed that there is no longer much room for improvement left as the mean scores are almost equal to one in the conditional model results. Especially in the Conditional Model 2 and 3, a comparison of the median, 75-percentile and max values might suggest a low discriminatory power of these two

[9] Although the results point at the importance of using a robust approach and the impact of the robustification, we decided not to remove any 'outlying observation'. First, there is no broadly accepted strategy to remove particular outliers (see De Witte and Marques (2010) for an extensive discussion). So, any applied approach would potentially result in some (grounded) critique. Second, removing particular observations might have strong impact on the results. It is unclear whether the new results are more reliable than the robust approach where we mitigate the impact of the outliers. Third, the outlying observations might also be the most interesting observations as they succeed in delivering a high service level.

model specifications due to the inclusion of several environmental variables whose interaction was meant to be investigated in the statistical inference part. To account for this, a separate model for socio-demographic background and one for the ideological complexion are also reported below in Table 5. By providing estimates for these restricted models, the results are more sensible.

Table 5: Descriptive statistics of the service provision composite indicator scores estimated for 307 municipalities over 2006–2011

	Mean	St. Dev.	Min	Q1	Median	Q3	Max
Robust Conditional Model 2 bis							
MinMax restrictions	0.9737	0.0348	0.7290	0.9548	0.9862	0.9999	1.0651
Average restrictions	0.9196	0.0645	0.6701	0.8716	0.9192	0.9845	1.0345
Municipal-specific restrictions	0.9120	0.0708	0.5827	0.8627	0.9140	0.9785	1.0443
Robust Conditional Model 3 bis							
MinMax restrictions	0.9626	0.0930	0.7796	0.9073	0.9438	0.9941	1.8311
Average restrictions	0.8831	0.0882	0.6964	0.8241	0.8680	0.9213	1.4758
Municipal-specific restrictions	0.8729	0.0972	0.6599	0.8115	0.8578	0.9168	1.7012

Note: Unrestricted indicates the absence of weight restrictions. *MinMax* restrictions refer to the minimum and maximum share of expenditure in each municipal area across all the municipalities. *Average* restrictions consider the average spending share (lower and upper bound equal to its ±50%). *Municipal-specific* restrictions are based on the municipal-specific current spending allocation (lower and upper bound equal to the ±50% of each municipal spending share).

Model 2 bis includes the socio-demographic structure (Share of elderly people, Share of foreigners and Population growth). *Model 3 bis* considers the political component (Ideological Complexion of the local Government). In every conditional model specification a year dummy is also included.

On one hand, the evidence shows that this set of variables explains largely the service level provision in municipalities; on the other hand, the conditional analysis helps in identifying the correlation between some municipal characteristics and the service provision level.

The influence of the contextual variables on the municipal service provision level can be detected by looking at the robust unconditional and conditional estimates together, as explained in section 3.4. Table 6 presents the results of the statistical inference. For the sake of brevity, only the results for the MinMax restrictions are presented along the paper: the results are robust across the three weight restriction specifications and we refer to Appendix D for the complete list of results. The direction of the influence of the environmental variables is in line with the main evidence described in the literature on local government's efficiency (see for all Narbón-Perpiñá and De Witte, 2018b).

Table 6: Influence of background conditions on municipal service composite indicator

MinMax weight restrictions

	Model 1			Model 2			Model 3		
	Influence	p-value		Influence	p-value		Influence	p-value	
Economic-financial									
Fiscal income	Unfavourable	0.000	***	Unfavourable	0.000	***	Unfavourable	0.000	***
Financial debt	Unfavourable	0.000	***	Favourable	0.000	***	Favourable	0.000	***
Unemployment	Unfavourable	0.000	***	Unfavourable	0.170		Unfavourable	0.085	*
Socio-demographic									
Residents over 65				Favourable	0.075	*	Favourable	0.000	***
Foreigners				Favourable	0.045	**	Favourable	0.080	*
Population growth				Unfavourable	0.000	***	Unfavourable	0.000	***
Political									
ICG							Unfavourable	0.000	***

Note: The background variable has an *unfavourable influence* on the service provision assessment when the municipal composite indicator score increases only because the municipality under assessment is evaluated among similar municipalities: the service provision it can afford is lower compared with the one of municipalities facing a different context. The opposite holds when a background variable is found to have a *favourable influence*.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Concerning the economic and financial characteristics, the level of fiscal income is observed to play an unfavourable role in the municipal service provision assessment. When local governments have a greater amount of financial resources, the politicians tend to spend in a less prudent way and the citizens are likely to be less motivated to monitor the expenditures. As a result, the overall level of delivered services seems to reduce (D'Inverno et al., 2018; Ashworth et al., 2014). The financial debt shows an unfavourable correlation as well. When the level of local government debt is higher, more resources will be spent on debt interests and amortization payments: therefore less resources will be available and this will bring to an overall lower level of service provision (Cordero et al., 2017a; Da Cruz and Marques, 2014). Also, when there is a higher level of unemployment, a lesser amount of resources is available to provide municipal services as higher spending is devoted to social and housing benefits. Hence, in the overall service provision assessment this variable also plays an unfavourable role (Pérez-López et al., 2015).

The dataset covers the 2006-2011 period, which by and large coincides with the term of local authorities elected in 2006. As in Cordero et al. (2017a), we adopt a dynamic approach to exploit intertemporal variations in public service provision and to observe whether municipalities made some changes during the term in which they were elected. Interestingly, if we consider the time trend looking at the partial plot for the year variable (see Figure 2) combined with the economic characteristics, it may be observed that 2008, the year of the economic crisis, had been the most unfavourable concerning the municipal service provision (the same was observed in each weight restriction specification). However, from 2009 increases in public service provision have been recorded over time. This phenomenon might be linked to the fact that in 2009 a new legislative era began at national level. During this time, one of the main priorities of the government had been to stimulate public service provision at local level in line with the subsidiarity principle (Sadioglu, 2016). Therefore, the negative influence of the crisis might have been balanced by the renewed attention on local service provision.

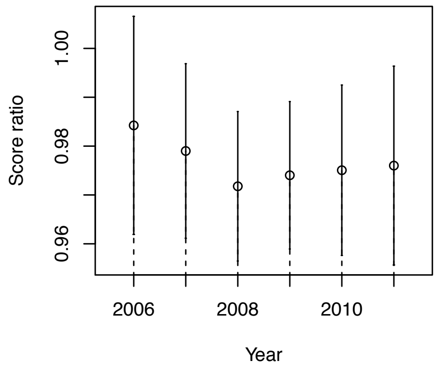

Figure 2: Intertemporal variation of the service provision when considering economic variables. The plot includes bias-corrected bootstrapped nonparametric confidence intervals.

When including also the socio-demographic characteristics, the interactions among the variables lead to contradicting evidence related to the direction of the influence of the level of fiscal income and financial debt, playing as counteracting factors. This could be due to the fact that municipalities are more concerned with the management of public resources when they have a higher level of accountability or when they pay more attention on cost saving due to their financial problems. In addition, the results of Model 2 on the socio-demographic structure show that the share of elderly people and of foreigners has a favourable influence on the municipal service provision assessment, whenever it is statistically significant. These population groups are the recipients of several municipal services provided with the aim of satisfying their needs: the higher the number of people in these categories, the greater the level of scale economies exploitation. Vice versa, when the level of population growth is too high, the municipalities might not be able to completely satisfy overall citizens' demand and therefore, keeping other things constant, it will lead to a lower level of provided services.

Finally, Model 3 also includes the information on the political component, namely the Ideological Complexion of the local Government (ICG), that captures the ideological stance of the municipality on a Left-Right scale. We observe that a low level of municipal service provision is associated with a more right-wing government. In fact, as a common hypothesis a more left-wing coalition is more prone to have a larger public sector.

For completeness, we report also the results of the statistical inference for the Conditional Model 2 and 3 (MinMax restrictions) run including respectively only the socio-demographic background and the Ideological Complexion in Table 7. The results are robust, except for 'Foreign', which turns out to be insignificant, suggesting that ignoring the economic and financial conditions of municipalities results in unobserved heterogeneity, and consequent endogeneity issues.

Table 7: Influence of background conditions on municipal service composite indicator

MinMax weight restrictions						
	Model 2 bis			Model 3 bis		
	Influence	p-value		Influence	p-value	
Economic-financial						
Fiscal income						
Financial debt						
Unemployment						
Socio-demographic						
Residents over 65	Favourable	0.000	***			
Foreigners	Unfavourable	0.570				
Population growth	Unfavourable	0.000	***			
Political						
ICG				Unfavourable	0.000	***

Note: The background variable has an *unfavourable influence* on the service provision assessment when the municipal composite indicator score increases only because the municipality under assessment is evaluated among similar municipalities: the service provision it can afford is lower compared with the one of municipalities facing a different context. The opposite holds when a background variable is found to have a *favourable influence*.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.1. Robustness tests

To conclude this section, it is worth pointing out that further analysis has been performed to check the robustness of the proposed tool and its empirical application in a twofold manner.

Comparative performance of alternative model formulations

First, we consider alternative state-of-the-art model specifications proposed in the literature to handle undesirable features in composite indicators (CIs) and we provide insights about their comparative performance. Specifically, we consider the CI model formulation proposed by Färe et al. (2019) and a variant of the Zanella et al. (2015b) model specification, both outlined in section 3.5. We replicate the empirical analysis and we compare them with the directional distance formulation advocated in this paper (hereafter referred to as the Zanella et al., 2015b model). To keep the discussion concise, in the following we comment the main findings and we refer to Appendix E.1 for more detailed results, along with the main tables and plots.

The CI scores obtained estimating the Zanella et al. (2015b) model are on average greater than the ones obtained in the Färe et al. (2019) model. If we compare these two model findings by scatter plot of ranks, we get a similar ranking when using the weight restrictions and especially in the average restriction case. We consider also the expenditure budget shares obtained from the shadow prices in both model specifications. Similarly to the scatter plot findings, we observe that even using a different model formulation to deal with undesirable indicators, by imposing the weight restrictions we get increasingly similar optimal expenditure shares.

About the comparison between the Zanella et al. (2015b) model and its variant, both the estimated CI scores and the scatter plot of ranks suggest that the equality constraint seems to play a role merely in the unrestricted model. In all other scenarios, the findings are very similar, showing a picture coherent with the one arising from the advocated composite indicator to measure the service level provision in municipalities, that is the flexible directional distance composite indicator with weight restrictions. As for the budget shares, the same reasoning applies straightforwardly.

Alternative sample specification to control for municipal size effect

Second, we address possible criticisms on the municipal size effect, even if the robust analysis should have taken this issue already into account. Accordingly, we perform the analysis excluding from the sample the largest cities (so-called "Centrumsteden"[10]). The main findings are confirmed and the results are presented in Appendix E.2.

5. Conclusion

In this paper we propose an innovative way to measure municipal service provision level. To show the usefulness of the proposed approach, we compute the municipal service provision composite indicator for 307 Flemish municipalities over the year 2006–2011.

The model specification advocated in this paper is fully flexible and has the ability to capture the multifaceted aspects involved in local public goods provision evaluation and, in particular, the heterogeneity among the municipalities in their activities is taken into consideration to provide a fair analysis. Accordingly, local political preferences and municipal characteristics are directly embedded in the model. Overall, the approach ensures an objective way to determine the suitability of each municipal area that is taken for the evaluation, while granting the most favourable aggregating scheme for the units under analysis. In the analysis, we include information on the municipal expenditure composition to determine the weight restrictions. The directional distance function formulation assists with the evaluation even along the undesirable features. The robust conditional version of the model controls for outlying observations, the municipal operating context and the time dimension. The main analysis is robust to different

[10]The "Centrumsteden" indicates 13 Flemish city centres, with relatively high numbers of inhabitants, that play a central role in the employment, care, education, culture and recreational activities. https://nl.wikipedia.org/wiki/Centrumstad

model specifications recently suggested by other scholars to deal with the undesirable features. Specifically, we provide insights about the comparative performance of the advocated flexible directional distance and two other models, that is the model formulation proposed by Färe et al. (2019) and a model variant of the Zanella et al. (2015b) specification.

The proposed composite indicator not only groups together all these components, going a step forward in the existing literature, but it also allows for further investigation. First of all, it can help exploring how municipal characteristics influence overall service provision through statistical inference, detecting whether the background condition inclusion favours the assessment or not. Broadly, the obtained composite indicator can be used to explore the relationship between the provided municipal services and some other relevant issues in municipal management, such as the government size expressed in terms of the tax burden imposed on citizens, who pay the taxes for the local public goods they receive.

In the empirical application, we find an unfavourable influence of the considered economic variables, i.e. level of fiscal income, financial debt and unemployment on the municipal service provision assessment. With respect to the share of elderly people and foreigners, a favourable influence is found, while for the population growth the opposite holds. Finally, a left-wing government favours municipal activities.

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.

Arcelus, F.J., Arocena, P., Cabasés, F., Pascual, P., 2015. On the cost-efficiency of service delivery in small municipalities. Regional studies 49, 1469–1480.

Asatryan, Z., De Witte, K., 2015. Direct democracy and local government efficiency. European Journal of Political Economy 39, 58–66.

Ashworth, J., Geys, B., Heyndels, B., Wille, F., 2014. Competition in the political arena and local government performance. Applied Economics 46, 2264–2276.

Bădin, L., Daraio, C., Simar, L., 2012. How to measure the impact of environmental factors in a nonparametric production model. European Journal of Operational Research 223, 818–833.

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.

Calabria, F.A., Camanho, A.S., Zanella, A., 2018. The use of composite indicators to evaluate the performance of Brazilian hydropower plants. International Transactions in Operational Research 25, 1323–1343.

Cazals, C., Florens, J.P., Simar, L., 2002. Nonparametric frontier estimation: a robust approach. Journal of econometrics 106, 1–25.

CEMR, 2016. Local and Regional Government in Europe. Structure and Competences URL: https://www.ccre.org/docs/Local_and_Regional_Government_in_Europe.EN.pdf.

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., Moesen, W., Rogge, N., Van Puyenbroeck, T., 2007. An introduction to âĂŸbenefit of the doubtâĂŹ composite indicators. Social indicators research 82, 111–145.

Chung, Y.H., Färe, R., Grosskopf, S., 1997. Productivity and undesirable outputs: a directional distance function approach. Journal of Environmental Management 51, 229–240.

Coelli, T.J., Rao, D.S.P., O'Donnell, C.J., Battese, G.E., 2005. An introduction to efficiency and productivity analysis. Springer Science & Business Media.

Cordero, J.M., Pedraja-Chaparro, F., Pisaflores, E.C., Polo, C., 2017a. Efficiency assessment of Portuguese municipalities using a conditional nonparametric approach. Journal of Productivity Analysis 48, 1–24.

Cordero, J.M., Salinas-Jiménez, J., Salinas-Jiménez, M.M., 2017b. Exploring factors affecting the level of happiness across countries: A conditional robust nonparametric frontier analysis. European Journal of Operational Research 256, 663–672.

Da Cruz, N.F., Marques, R.C., 2014. Revisiting the determinants of local government performance. Omega 44, 91–103.

Dakpo, K.H., Jeanneaux, P., Latruffe, L., 2016. Modelling pollution-generating technologies in performance benchmarking: recent developments, limits and future prospects in the nonparametric framework. European Journal of Operational Research 250, 347–359.

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., 2007. Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications. Springer Science & Business Media.

De Borger, B., Kerstens, K., 1996a. Cost efficiency of Belgian local governments: A comparative analysis of FDH, DEA, and econometric approaches. Regional science and urban economics 26, 145–170.

De Borger, B., Kerstens, K., 1996b. Radial and nonradial measures of technical efficiency: an empirical illustration for belgian local governments using an FDH reference technology. Journal of Productivity Analysis 7, 41–62.

De Witte, K., Geys, B., 2011. Evaluating efficient public good provision: Theory and evidence from a generalised conditional efficiency model for public libraries. Journal of Urban Economics 69, 319–327.

De Witte, K., Geys, B., 2013. Citizen coproduction and efficient public good provision: Theory and evidence from local public libraries. European Journal of Operational Research 224, 592–602.

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., Marques, R.C., 2010. Influential observations in frontier models, a robust non-oriented approach to the water sector. Annals of Operations Research 181, 377–392.

De Witte, K., Schiltz, F., 2018. Measuring and explaining organizational effectiveness of school districts: Evidence from a robust and conditional Benefit-of-the-Doubt approach. European Journal of Operational Research 267, 1172–1181.

D'Inverno, G., Carosi, L., Ravagli, L., 2018. Global public spending efficiency in Tuscan municipalities. Socio-Economic Planning Sciences 61, 102–113.

Eeckaut, P., Tulkens, H., Jamar, M.A., 1993. Cost efficiency in Belgian municipalities. The Measurement of Productive Efficiency–Techniques and Applications , 300–334.

Färe, R., Karagiannis, G., Hasannasab, M., Margaritis, D., 2019. A benefit-of-the-doubt model with reverse indicators. European Journal of Operational Research 278, 394–400.

Fusco, E., 2015. Enhancing non-compensatory composite indicators: a directional proposal. European Journal of Operational Research 242, 620–630.

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 .

Fusco, E., Vidoli, F., Sahoo, B.K., 2018. Spatial heterogeneity in composite indicator: A methodological proposal. Omega 77, 1–14.

Global Taskforce of Local and Regional Governments, 2016. Roadmap for localizing the SDGs: Implementation and monitoring at subnational level URL: `https://unhabitat.org/roadmap-for-localizing-the-sdgs-implementation-and-monitoring-at-subnational-level/`.

Grupp, H., Schubert, T., 2010. Review and new evidence on composite innovation indicators for evaluating national performance. Research Policy 39, 67–78.

Helland, L., Sørensen, R.J., 2015. Partisan bias, electoral volatility, and government efficiency. Electoral Studies 39, 117–128.

Karagiannis, G., 2017. On aggregate composite indicators. Journal of the Operational Research Society 68, 741–746.

Lo Storto, C., 2016. The trade-off between cost efficiency and public service quality: A non-parametric frontier analysis of Italian major municipalities. Cities 51, 52–63.

Loikkanen, H.A., Susiluoto, I., 2005. Cost Efficiency of Finnish Municipalities in basic service provision 1994-2002. Urban Public Economics Review , 39–63.

Mastromarco, C., Simar, L., 2015. Effect of FDI and time on catching up: new insights from a conditional nonparametric frontier analysis. Journal of Applied Econometrics 30, 826–847.

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 .

Moore, A., Nolan, J., Segal, G.F., 2005. Putting out the trash: Measuring municipal service efficiency in us cities. Urban Affairs Review 41, 237–259.

Morais, P., Camanho, A.S., 2011. Evaluation of performance of european cities with the aim to promote quality of life improvements. Omega 39, 398–409.

Narbón-Perpiñá, I., De Witte, K., 2018a. Local governments' efficiency: A systematic literature review-âĂŤPart I. International Transactions in Operational Research 25, 431–468.

Narbón-Perpiñá, I., De Witte, K., 2018b. Local governments' efficiency: A systematic literature reviewâĂŤ-Part II. International Transactions in Operational Research 25, 1107–1136.

Nikolov, M., Hrovatin, N., 2013. Cost efficiency of Macedonian municipalities in service delivery: does ethnic fragmentation matter? Lex localis-Journal of Local Self-Government 11.

OECD, 2008. Handbook on constructing composite indicators: methodology and user guide.

OECD/UCLG, 2016. Subnational Governments around the world: Structure and finance.

Oliveira, R., Zanella, A., Camanho, A.S., 2019. The assessment of corporate social responsibility: The construction of an industry ranking and identification of potential for improvement. European Journal of Operational Research 278, 498–513.

Pérez-López, G., Prior, D., Zafra-Gómez, J.L., 2015. Rethinking new public management delivery forms and efficiency: Long-term effects in spanish local government. Journal of Public Administration Research and Theory 25, 1157–1183.

Prieto, A.M., Zofio, J.L., 2001. Evaluating effectiveness in public provision of infrastructure and equipment: The case of Spanish municipalities. Journal of Productivity Analysis 15, 41–58.

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.

Sadioglu, U., 2016. Theoretical Foundations and Discussions on the Reformation Process in Local Governments. IGI Global.

Sarrico, C.S., Dyson, R., 2004. Restricting virtual weights in data envelopment analysis. European Journal of Operational Research 159, 17–34.

Van Puyenbroeck, T., 2018. On the Output Orientation of the Benefit-of-the-Doubt-Model. Social Indicators Research 139, 415–431.

Van Puyenbroeck, T., Rogge, N., 2017. Geometric mean quantity index numbers with benefit-of-the-doubt weights. European Journal of Operational Research 256, 1004–1014.

Verschelde, M., Rogge, N., 2012. An environment-adjusted evaluation of citizen satisfaction with local police effectiveness: Evidence from a conditional Data Envelopment Analysis approach. European Journal of Operational Research 223, 214–225.

Yusfany, A., 2015. The Efficiency of Local Governments and its Influence Factors. International Journal of Technology Enhancements and Emerging Engineering Research 4, 219–241.

Zanella, A., Camanho, A.S., Dias, T.G., 2015a. The assessment of cities? livability integrating human wellbeing and environmental impact. Annals of Operations Research 226, 695–726.

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

Appendix Online - Supplementary material

Appendix A. Data sources and description

Data are available via `http://statistieken.vlaanderen.be/QvAJAXZfc/notoolbar.htm?document=SVR%2FSVR-alle-domeinen.qvw&host=QVS%40cwv100154&anonymous=true`.

In Table A.1, the list of the variables as downloaded is reported with their original Dutch names. In addition to the listed variables, the data linked to *Population Growth* and *Ideological Complexion of the local Government (ICG)* have been provided by De Witte and Geys (2011).

Per capita variables as listed in Table 1 have been obtained using the total number of residents per municipality.

Some of the explanatory variables have been categorized for methodological reasons (for more technical details, see Rogge et al., 2017, and the references therein). Specifically, *Income per capita* and *Financial debt per capita* has been divided into deciles, *Population growth* into quartiles and *Ideological Complexion of the local Government (ICG)* has been split in three parts.

Out of 308 Flemish municipalities, one municipality has not been included in the analysis because of the lack of data.

Table A.1: Downloaded variables used for the empirical application.

DOMEIN	SUBDOMEIN	INDICATOR	INDICATOR
Arbeidsmarkt	Werkloosheid	Werkloosheidsgraad (15-64 jaar) (Steunpunt Werk)	Unemployment rate (15-64) (Centre for Work)
Criminaliteit	Geregistreerde misdrijven	Diefstallen en afpersingen - per 1.000 inwoners	Thefts and extortions - per 1,000 inhabitants
Criminaliteit	Geregistreerde misdrijven	Misdrijven tegen eigendom - per 1.000 inwoners	Crimes against property - per 1,000 inhabitants
Criminaliteit	Geregistreerde misdrijven	Misdrijven tegen lichamelijke integriteit - per 1.000 inwoners	Crimes against physical integrity - per 1,000 inhabitants
Cultuur	Algemeen	Cultuurevenementen: aantal per 1.000 inwoners	Culture Events: number per 1,000 inhabitants
Demografie	Structuur bevolking	Inwoners - totaal	Residents - total
Demografie	Structuur bevolking	Inwoners 65 jaar en ouder - aandeel	Residents age 65 and older - share
Demografie	Structuur bevolking	Ouderen 80 jaar en ouder - aantal	Older people aged 80 and over - Number
Demografie	Huishoudens	Private huishoudens - totaal aantal	Private households - total
Demografie	Migraties	Saldo internationale migraties van vreemdelingen	Net international migration of foreigners
Economie en innovatie	Macro-economie	Belastbaar inkomen per inwoner	Taxable income per capita
Energie	Energieverbruik	Energieverbruik per inwoner door verwarming huishoudens	Energy consumption per capita by household heating
FinanciÃČÂńn & Bestuur	FinanciÃČÂńle schuld gemeenten	FinanciÃČÂńle schuld per inwoner	Financial debt per capita
Inburgering & Integratie	Aanwezigheid	Vreemdelingen - aandeel t.o.v. totale bevolking	Foreigners - share relative to total population
Milieu en natuur	Afval	Restafval in kg per inwoner	Waste in kg per capita
Mobiliteit	Verkeersveiligheid	Ongevallen - aantal	Accidents - number
Onderwijs en vorming	Kleuteronderwijs	Totaal kleuteronderwijs - aantal leerlingen (naar vestigingplaats)	Total kindergarten - number of students (by domicile)
Onderwijs en vorming	Lager onderwijs	Totaal lager onderwijs - aantal leerlingen (naar vestigingplaats)	Total primary education - number of students (by domicile)
Ruimtelijke ontwikkeling	Oppervlakte	Bebouwde oppervlakte	Built up area

Note: The first three columns show the path followed to get the chosen variables. Accordingly, the original Dutch names are reported. The last column presents the English version as translated by the authors.

Appendix B. Spearman correlation tables

Table B.1: Spearman correlation among model specifications

MinMax weight restrictions

	Unconditional	Rob. Uncond.	Rob. Cond. 1	Rob. Cond. 2	Rob. Cond. 3
Unconditional	1.0000				
Rob. Uncond.	0.9553	1.0000			
Rob. Cond. 1	0.6708	0.7372	1.0000		
Rob. Cond. 2	0.2151	0.2475	0.3313	1.0000	
Rob. Cond. 3	0.1331	0.1592	0.2292	0.2930	1.0000

Average weight restrictions

	Unconditional	Rob. Uncond.	Rob. Cond. 1	Rob. Cond. 2	Rob. Cond. 3
Unconditional	1.0000				
Rob. Uncond.	0.9862	1.0000			
Rob. Cond. 1	0.7454	0.7643	1.0000		
Rob. Cond. 2	0.2118	0.2192	0.2914	1.0000	
Rob. Cond. 3	0.1129	0.1228	0.1978	0.3651	1.0000

Municipal-specific weight restrictions

	Unconditional	Rob. Uncond.	Rob. Cond. 1	Rob. Cond. 2	Rob. Cond. 3
Unconditional	1.0000				
Rob. Uncond.	0.9142	1.0000			
Rob. Cond. 1	0.6660	0.7548	1.0000		
Rob. Cond. 2	0.1682	0.2162	0.2986	1.0000	
Rob. Cond. 3	0.1040	0.1416	0.2152	0.3477	1.0000

Note: MinMax restrictions refer to the minimum and maximum share of expenditure in each municipal area across all the municipalities. *Average* restrictions consider the average spending share (lower and upper bound equal to its ±50%). *Municipal-specific* restrictions are based on the municipal-specific current spending allocation (lower and upper bound equal to the ±50% of each municipal spending share).

Unconditional refers to the baseline Benefit-of-the-Doubt Directional Distance function model, without correction for outlying observations, as it is instead in the *Robust* specification. *Model 1* includes the economic and financial characteristics (Fiscal income, Financial debt and Unemployment). *Model 2* adds to the economic and financial characteristics the socio-demographic structure (Share of elderly people, Share of foreigners and Population growth). *Model 3* adds to the economic-financial and socio-demographic variables also the political component (Ideological Complexion of the local Government). In every conditional model specification a year dummy is also included.

Table B.2: Spearman correlation among weight restriction models

Unconditional

	MinMax	Average	Municipal
MinMax restrictions	1.0000		
Average restrictions	0.8755	1.0000	
Municipal-specific restrictions	0.6796	0.7973	1.0000

Robust Unconditional

	MinMax	Average	Municipal
MinMax restrictions	1.0000		
Average restrictions	0.9202	1.0000	
Municipal-specific restrictions	0.8235	0.8913	1.0000

Robust Conditional Model 1

	MinMax	Average	Municipal
MinMax restrictions	1.0000		
Average restrictions	0.8834	1.0000	
Municipal-specific restrictions	0.8365	0.9074	1.0000

Robust Conditional Model 2

	MinMax	Average	Municipal
MinMax restrictions	1.0000		
Average restrictions	0.8052	1.0000	
Municipal-specific restrictions	0.7928	0.9423	1.0000

Robust Conditional Model 3

	MinMax	Average	Municipal
MinMax restrictions	1.0000		
Average restrictions	0.7629	1.0000	
Municipal-specific restrictions	0.7466	0.9279	1.0000

Note: MinMax restrictions refer to the minimum and maximum share of expenditure in each municipal area across all the municipalities. *Average* restrictions consider the average spending share (lower and upper bound equal to its ±50%). *Municipal-specific* restrictions are based on the municipal-specific current spending allocation (lower and upper bound equal to the ±50% of each municipal spending share).

Unconditional refers to the baseline Benefit-of-the-Doubt Directional Distance function model, without correction for outlying observations, as it is instead in the *Robust* specification. *Model 1* includes the economic and financial characteristics (Fiscal income, Financial debt and Unemployment). *Model 2* adds to the economic and financial characteristics the socio-demographic structure (Share of elderly people, Share of foreigners and Population growth). *Model 3* adds to the economic-financial and socio-demographic variables also the political component (Ideological Complexion of the local Government). In every conditional model specification a year dummy is also included.

Appendix C. Shadow prices results

For the sake of brevity, only the shadow prices' descriptive statistics comparing the unrestricted unconditional and the robust conditional analysis for the three weight restriction models are presented below. Then, the budget shares obtained from the computed shadow prices are presented, again comparing the Unrestricted Unconditional and the robust conditional analysis for the three weight restriction models. For the sake of brevity, we show in the following only the estimates for the conditional model including the economic variables (Fiscal income, Financial debt and Unemployment). The results for the conditional model encompassing also Socio-demographic and Political components are available upon request from the authors.

For the sake of comparison, we recall below the average expenditure composition across the eight municipal functions considered in the present analysis.

Table C.1: Summary of the weights obtained from the municipal expenditure composition

	Administration	Culture	Care services	Education	Housing	Local mobility	Security	Environment
MIN	0.08	0.02	0.01	0	0.01	0.03	0.04	0.03
AVERAGE	0.18	0.14	0.12	0.10	0.05	0.15	0.11	0.10
MAX	0.49	0.31	0.23	0.39	0.16	0.33	0.23	0.22

We can observe that imposing weight restrictions provides shadow prices closer with the current municipal composition. Budget allocation cannot be changed drastically, as otherwise suggested by the unrestricted results, and a minimum expenditure share is granted to each function in this way, avoiding zero weights whenever a minimum amount of investment is required in a certain municipal function.

For completeness, also the descriptive statistics related to the free variable v contained in the model introduced in Section 3 (formula 2) are presented.

Table C.2: Descriptive statistics of the shadow prices for each municipal function (Unrestricted unconditional and Robust conditional estimates)

Unrestricted unconditional estimates

	Administration	Culture	Care services	Education	Housing	Local mobility	Security	Environment	v
Mean	0.0906	0.0038	0.0667	0.0012	0.0004	0.0214	0.2042	0.9926	-0.1917
SD	0.0476	0.0042	0.1229	0.001	0.0006	0.0257	0.2882	0.3988	0.3611
Min	0	0	0	0	0	0	0	0	-0.7904
Max	0.2033	0.0166	0.6791	0.005	0.0028	0.1582	2.6924	1.8997	1.0795
N	307	307	307	307	307	307	307	307	307

MinMax weight restrictions

	Administration	Culture	Care services	Education	Housing	Local mobility	Security	Environment	v
Mean	0.4134	0.0093	0.1211	0.0016	0.0008	0.0277	0.2085	0.1832	0.4395
SD	0.1219	0.0069	0.0794	0.0015	0.0007	0.0183	0.1098	0.1013	0.1592
Min	0.1101	0.0025	0.0062	0	0.0001	0.0052	0.0549	0.0277	-0.0422
Max	0.6581	0.0348	0.3118	0.0065	0.0026	0.0994	0.5476	0.4386	0.8127
N	307	307	307	307	307	307	307	307	307

Average weight restrictions

	Administration	Culture	Care services	Education	Housing	Local mobility	Security	Environment	v
Mean	0.296	0.0146	0.1746	0.0016	0.0008	0.0335	0.2252	0.1975	0.3927
SD	0.0555	0.0041	0.0455	0.0005	0.0003	0.0118	0.0807	0.0567	0.1094
Min	0.1318	0.0079	0.0709	0.0005	0.0003	0.012	0.0628	0.0577	0.0295
Max	0.4483	0.0275	0.2675	0.0027	0.0016	0.0655	0.4308	0.3214	0.6256
N	307	307	307	307	307	307	307	307	307

Municipal-specific weight restrictions

	Administration	Culture	Care services	Education	Housing	Local mobility	Security	Environment	v
Mean	0.2948	0.0148	0.1709	0.0015	0.0008	0.0345	0.2229	0.2037	0.3816
SD	0.0992	0.006	0.0658	0.0012	0.0006	0.0186	0.0847	0.0899	0.1479
Min	0.0891	0.0029	0.0073	0	0.0001	0.0026	0.0704	0.0366	-0.0548
Max	0.7595	0.0592	0.3511	0.0067	0.0032	0.0981	0.6727	0.4956	0.7499
N	307	307	307	307	307	307	307	307	307

Table C.3: Descriptive statistics of the budget shares across the municipal functions (Unrestricted unconditional and Robust conditional estimates)

Unrestricted unconditional estimates

	Administration	Culture	Care services	Education	Housing	Local mobility	Security	Environment
Mean	7%	3%	6%	7%	3%	8%	9%	56%
SD	4%	3%	13%	7%	4%	9%	11%	21%
Min	0%	0%	0%	0%	0%	0%	0%	0%
Max	18%	19%	67%	47%	23%	51%	61%	89%
N	307	307	307	307	307	307	307	307

MinMax weight restrictions

	Administration	Culture	Care services	Education	Housing	Local mobility	Security	Environment
Mean	35%	6%	10%	10%	5%	12%	10%	11%
SD	9%	5%	6%	9%	5%	7%	5%	5%
Min	9%	2%	1%	0%	1%	3%	4%	3%
Max	49%	26%	23%	39%	16%	33%	23%	22%
N	307	307	307	307	307	307	307	307

Average weight restrictions

	Administration	Culture	Care services	Education	Housing	Local mobility	Security	Environment
Mean	25%	10%	14%	10%	5%	14%	11%	12%
SD	3%	3%	3%	3%	2%	4%	3%	2%
Min	10%	7%	6%	5%	3%	7%	5%	5%
Max	27%	21%	17%	14%	8%	22%	16%	15%
N	307	307	307	307	307	307	307	307

Municipal-specific weight restrictions

	Administration	Culture	Care services	Education	Housing	Local mobility	Security	Environment
Mean	25%	10%	14%	9%	5%	15%	11%	12%
SD	7%	4%	5%	8%	4%	7%	4%	5%
Min	9%	2%	1%	0%	1%	2%	4%	3%
Max	65%	34%	29%	42%	20%	34%	31%	29%
N	307	307	307	307	307	307	307	307

Appendix D. Statistical inference results

In the following, the statistical inference results are reported for each weight restriction specification:

1. *MinMax* restrictions refer to the minimum and maximum share of expenditure in each municipal area across all the municipalities;
2. *Average* restrictions consider the average spending share (lower and upper bound equal to its $\pm 50\%$);
3. *Municipal-specific* restrictions are based on the municipal-specific current spending allocation (lower and upper bound equal to the $\pm 50\%$ of each municipal spending share).

For the conditional model specification, three groups of background variables are considered (in every conditional model specification a year dummy is also included):

1. *Model 1* includes the economic and financial characteristics (Fiscal income, Financial debt and Unemployment);
2. *Model 2* adds to the economic and financial characteristics the socio-demographic structure (Share of elderly people, Share of foreigners and Population growth);
3. *Model 3* adds to the economic-financial and socio-demographic variables the political component (Ideological Complexion of the local Government).

For the sake of clarity, we say that the background variable has an unfavourable influence on the service provision assessment when the municipal composite indicator score increases only because the municipality under assessment is evaluated among similar municipalities: the service provision it can afford is lower compared with the one of municipalities facing a different context. The opposite holds when a background variable is found to have a favourable influence.

Table D.1: Influence of background conditions on municipal service composite indicator

MinMax weight restrictions

	Model 1			Model 2			Model 3		
	Influence	p-value		Influence	p-value		Influence	p-value	
Economic-financial									
Fiscal income	Unfavourable	0.000	***	Unfavourable	0.000	***	Unfavourable	0.000	***
Financial debt	Unfavourable	0.000	***	Favourable	0.000	***	Favourable	0.000	***
Unemployment	Unfavourable	0.000	***	Unfavourable	0.170		Unfavourable	0.085	*
Socio-demographic									
Residents over 65				Favourable	0.075	*	Favourable	0.000	***
Foreigners				Favourable	0.045	**	Favourable	0.080	*
Population growth				Unfavourable	0.000	***	Unfavourable	0.000	***
Political									
ICG							Unfavourable	0.000	***

Average weight restrictions

	Model 1			Model 2			Model 3		
	Influence	p-value		Influence	p-value		Influence	p-value	
Economic-financial									
Fiscal income	Unfavourable	0.000	***	Unfavourable	0.000	***	Unfavourable	0.000	***
Financial debt	Unfavourable	0.000	***	Unfavourable	0.000	***	Favourable	0.000	***
Unemployment	Unfavourable	0.000	***	Unfavourable	0.000	***	Unfavourable	0.000	***
Socio-demographic									
Residents over 65				Favourable	0.000	***	Favourable	0.210	
Foreigners				Favourable	0.100		Favourable	0.030	**
Population growth				Unfavourable	0.000	***	Unfavourable	0.000	***
Political									
ICG							Unfavourable	0.000	***

Municipal-specific weight restrictions

	Model 1			Model 2			Model 3		
	Influence	p-value		Influence	p-value		Influence	p-value	
Economic-financial									
Fiscal income	Unfavourable	0.000	***	Favourable	0.000	***	Favourable	0.000	***
Financial debt	Unfavourable	0.000	***	Unfavourable	0.000	***	Unfavourable	0.000	***
Unemployment	Unfavourable	0.000	***	Unfavourable	0.750		Unfavourable	0.005	***
Socio-demographic									
Residents over 65				Favourable	0.000	***	Favourable	0.000	***
Foreigners				Favourable	0.000	***	Favourable	0.035	**
Population growth				Unfavourable	0.000	***	Unfavourable	0.000	***
Political									
ICG							Unfavourable	0.000	***

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix E. Robustness test results

Appendix E.1. Alternative model specification to handle undesirable features

In the following, the main results of the robustness tests concerning the model specification are reported. The municipal scores are obtained using alternative state-of-the-art model specifications proposed in the literature to handle undesirable features/reverse indicators in composite indicators (CIs).
Specifically, we consider first the CI model formulation proposed by Färe et al. (2019) and then a variant model of the Zanella et al. (2015b) specification and we replicate the empirical analysis accordingly. Estimating both models suggests in our application robust findings.

The Färe et al. (2019) model
First, we depart from the directional distance formulation and we move along the lines traced by Färe et al. (2019). Specifically, we considered their model (3) p. 396, so that we can account for the municipal expenditure composition by adding the weight restrictions. To facilitate the comparison between the model proposed by Zanella et al. (2015b) and the one by Färe et al. (2019), we report both of them in the following, adapting the notation with the one used in Section 3.

Zanella et al. (2015, model 8 p.523)

$$
\begin{aligned}
\beta_{j_0} = \quad & \min -\textstyle\sum_{r=1}^{s} y_{rj_0}u_{rj_0} + \sum_{k=1}^{l} b_{kj_0}p_{kj_0} + v_{j_0} \\
\text{s.t.} \quad & \textstyle\sum_{r=1}^{s} g_y u_{rj_0} + \sum_{k=1}^{l} g_b p_{kj_0} = 1 \\
& -\textstyle\sum_{r=1}^{s} y_{rj}u_{rj_0} + \sum_{k=1}^{l} b_{kj}p_{kj_0} + v_{j_0} \geq 0 \\
& \qquad \text{for } j = 1, \ldots, j_0, \ldots, n \\
& u_{rj_0} \geq 0 \quad \text{for } r = 1, \ldots, s \\
& p_{kj_0} \geq 0 \quad \text{for } k = 1, \ldots, l \\
& v_{j_0} \in \Re \\
& \phi_r \leq \frac{u_{rj_0}\bar{y}_r}{\sum_{r=1}^{s} u_{rj_0}\bar{y}_r + \sum_{k=1}^{l} p_{kj_0}\bar{b}_k} \leq \psi_r \\
& \qquad \text{for } r = 1, \ldots, s \\
& \phi_k \leq \frac{p_{kj_0}\bar{b}_k}{\sum_{r=1}^{s} u_{rj_0}\bar{y}_r + \sum_{k=1}^{l} p_{kj_0}\bar{b}_k} \leq \psi_k \\
& \qquad \text{for } k = 1, \ldots, l \\
\text{where} \quad & CI_{j_0} = 1/(1+\beta_{j_0})
\end{aligned}
$$

Fare et al. (2019, model 3 p.396)

$$
\begin{aligned}
CI_{j_0} = \quad & \max \textstyle\sum_{r=1}^{s} y_{rj_0}u_{rj_0} - \sum_{k=1}^{l} b_{kj_0}p_{kj_0} \\
\text{s.t.} \quad & \textstyle\sum_{r=1}^{s} y_{rj}u_{rj_0} - \sum_{k=1}^{l} b_{kj}p_{kj_0} \leq 1 \\
& \qquad \text{for } j = 1, \ldots, j_0, \ldots, n \\
& u_{rj_0} \geq 0 \quad \text{for } r = 1, \ldots, s \\
& p_{kj_0} \geq 0 \quad \text{for } k = 1, \ldots, l \\
& \phi_r \leq \frac{u_{rj_0}\bar{y}_r}{\sum_{r=1}^{s} u_{rj_0}\bar{y}_r + \sum_{k=1}^{l} p_{kj_0}\bar{b}_k} \leq \psi_r \\
& \qquad \text{for } r = 1, \ldots, s \\
& \phi_k \leq \frac{p_{kj_0}\bar{b}_k}{\sum_{r=1}^{s} u_{rj_0}\bar{y}_r + \sum_{k=1}^{l} p_{kj_0}\bar{b}_k} \leq \psi_k \\
& \qquad \text{for } k = 1, \ldots, l
\end{aligned}
$$

where y_{rj_0} and b_{kj_0} respectively refer to the observed r desirable and k undesirable indicator of the evaluated municipality j_0; u_{rj_0} and p_{kj_0} respectively represent the most favourable BoD-weights for the r desirable and k undesirable indicator for the evaluated municipality j_0; y_{rj} and b_{kj} respectively refer to the r desirable and k undesirable indicator of every municipality j in the dataset; n is the number of municipalities under analysis (n=307); s and l respectively signify the number of municipal functions linked to desirable and undesirable indicators considered in this application (s=5 and l =3).

Table E.1 compares the descriptive statistics of the municipal CI scores obtained estimating both models. Specifically, we replicate Table 4 in Section 4: we report the unrestricted and three restricted model specifications, combined with the unconditional, robust and robust conditional estimates (to avoid redundancy, we report the CI scores for only one conditional model, namely the Robust Conditional Model 1 including the economic variables - Fiscal income, Financial debt and Unemployment).

Table E.1: Descriptive statistics of the service provision composite indicator scores estimated for 307 municipalities over 2006-2011

	Zanella et al. (2015)				Fare et al. (2019)			
	Mean	sd	Min	Max	Mean	sd	Min	Max
Unrestricted Unconditional	0.8388	0.0643	0.6357	1.0000	0.6679	0.1085	0.4653	1.0000
Unconditional								
MinMax restrictions	0.7832	0.0670	0.6302	1.0000	0.5990	0.1148	0.3166	1.0000
Average restrictions	0.7178	0.0649	0.5883	1.0000	0.4766	0.1205	0.2337	1.0000
Municipal-specific restrictions	0.7067	0.0723	0.5121	1.0000	0.4512	0.1392	0.0904	1.0000
Robust Unconditional								
MinMax restrictions	0.9618	0.1098	0.8008	2.0990	0.8917	0.1360	0.6179	2.0252
Average restrictions	0.8752	0.0936	0.7002	1.5770	0.7702	0.2620	0.3446	3.5385
Municipal-specific restrictions	0.8662	0.1041	0.6652	1.8767	0.7833	0.6344	0.1398	9.4509
Robust Conditional Model 1								
MinMax restrictions	0.9753	0.0315	0.8007	1.0013	0.9537	0.0584	0.6178	1.0006
Average restrictions	0.9215	0.0611	0.7034	1.0004	0.8290	0.1294	0.3834	1.0015
Municipal-specific restrictions	0.9158	0.0672	0.6571	1.0003	0.8056	0.1564	0.2040	1.0173

Note: Unrestricted indicates the absence of weight restrictions. *MinMax* restrictions refer to the minimum and maximum share of expenditure in each municipal area across all the municipalities. *Average* restrictions consider the average spending share (lower and upper bound equal to its ±50%). *Municipal-specific* restrictions are based on the municipal-specific current spending allocation (lower and upper bound equal to the ±50% of each municipal spending share). *Model 1* includes the economic and financial characteristics (Fiscal income, Financial debt and Unemployment) and a year dummy.

Next to the descriptive statistics of the municipal service provision composite indicator scores in Table E.1, we compare the two model findings by scatter plot of ranks. Interestingly, we can observe that even using a different model formulation to deal with undesirable output, by imposing the weight restrictions we get similar ranking, especially in the Average restriction case.

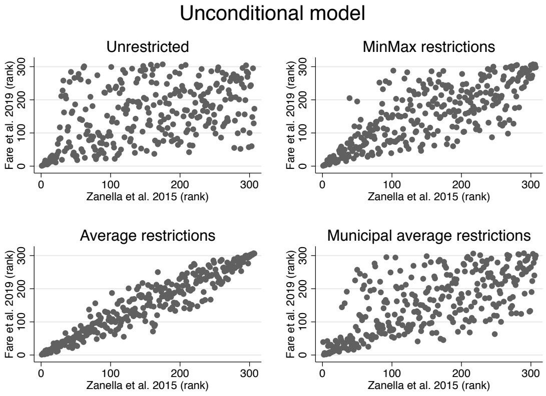

Figure E.1: Zanella et al. (2015) vs Fare et al. (2019). Service provision CI scores aggregated at municipal level

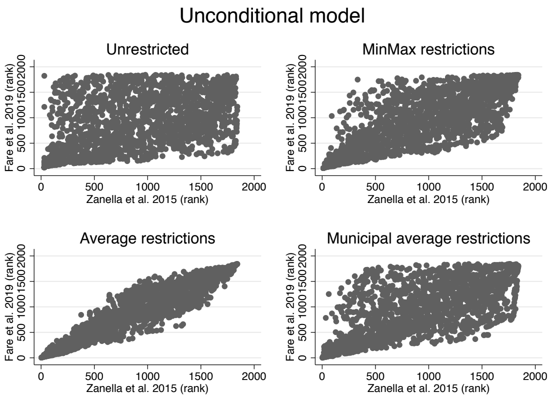

Figure E.2: Zanella et al. (2015) vs Fare et al. (2019).

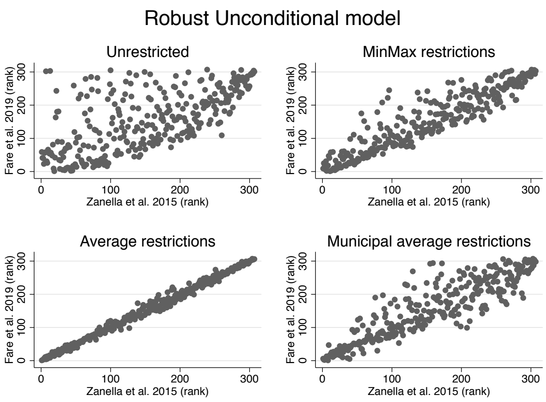

Figure E.3: Zanella et al. (2015) vs Fare et al. (2019). Service provision CI scores aggregated at municipal level

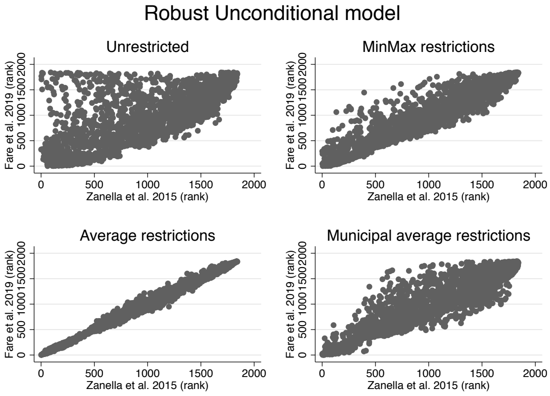

Figure E.4: Zanella et al. (2015) vs Fare et al. (2019).

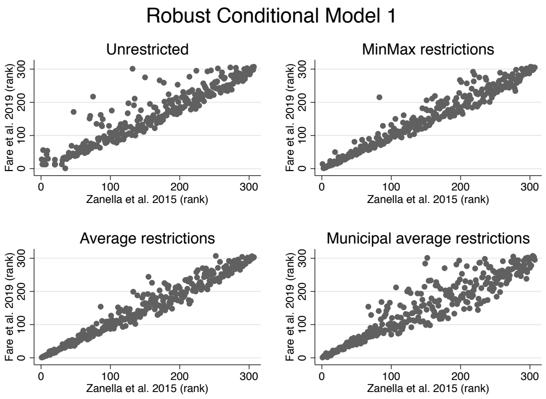

Figure E.5: Zanella et al. (2015) vs Fare et al. (2019). Service provision CI scores aggregated at municipal level

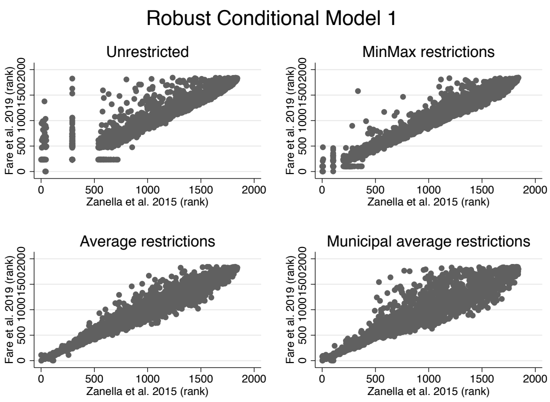

Figure E.6: Zanella et al. (2015) vs Fare et al. (2019).

We look also at the expenditure budget shares obtained from the shadow prices in both model specifications. For the sake of brevity, we replicate Table C.3 presented in Appendix C: we reported the budget shares obtained for the Unrestricted unconditional and for the Robust conditional estimates for Model 1 including the economic variables - Fiscal income, Financial debt and Unemployment. As for the scatter plot findings, we observe that even using a different model formulation to deal with undesirable output, by imposing the weight restriction we get increasingly similar optimal expenditure shares. Especially in the case of the Average weight restrictions, the optimal budget shares are very similar, both estimating the composite indicator by using the Zanella et al. (2015b) model and the one by Färe et al. (2019).

Table E.2: Descriptive statistics of the budget shares across the municipal functions (Unrestricted unconditional and Robust conditional estimates)

Unrestricted unconditional estimates

	Zanella et al. (2015)				Fare et al. (2019)			
	Mean	sd	Min	Max	Mean	sd	Min	Max
Administration	7%	4%	0%	18%	8%	6%	0%	23%
Culture	3%	3%	0%	19%	9%	5%	0%	19%
Care services	6%	13%	0%	69%	42%	25%	0%	84%
Education	7%	7%	0%	46%	16%	20%	0%	72%
Housing	3%	5%	0%	27%	18%	10%	0%	34%
Local mobility	8%	9%	0%	76%	1%	4%	0%	43%
Security	9%	11%	0%	61%	5%	10%	0%	57%
Environment	56%	21%	0%	89%	1%	5%	0%	59%

Robust conditional - MinMax restrictions

	Zanella et al. (2015)				Fare et al. (2019)			
	Mean	sd	Min	Max	Mean	sd	Min	Max
Administration	29%	11%	9%	49%	34%	11%	8%	49%
Culture	7%	6%	2%	25%	6%	4%	2%	24%
Care services	8%	5%	1%	23%	9%	6%	1%	23%
Education	16%	10%	0%	38%	17%	10%	0%	38%
Housing	7%	5%	1%	16%	8%	5%	1%	16%
Local mobility	11%	7%	3%	30%	8%	5%	3%	30%
Security	11%	4%	4%	22%	10%	4%	4%	23%
Environment	11%	5%	3%	22%	9%	4%	3%	20%

Robust conditional - Average restrictions

	Zanella et al. (2015)				Fare et al. (2019)			
	Mean	sd	Min	Max	Mean	sd	Min	Max
Administration	23%	4%	11%	27%	25%	3%	14%	27%
Culture	10%	3%	7%	19%	10%	3%	7%	20%
Care services	13%	3%	6%	17%	15%	2%	8%	17%
Education	11%	3%	5%	14%	12%	2%	6%	14%
Housing	6%	2%	3%	8%	7%	1%	3%	8%
Local mobility	14%	4%	7%	22%	12%	3%	7%	21%
Security	11%	3%	5%	16%	10%	3%	5%	16%
Environment	12%	2%	5%	15%	9%	2%	5%	14%

Robust conditional - Municipal average restrictions

	Zanella et al. (2015)				Fare et al. (2019)			
	Mean	sd	Min	Max	Mean	sd	Min	Max
Administration	23%	7%	6%	60%	25%	8%	5%	65%
Culture	10%	5%	2%	29%	11%	5%	2%	36%
Care services	13%	5%	1%	28%	15%	5%	1%	33%
Education	10%	8%	0%	44%	12%	9%	0%	46%
Housing	6%	4%	1%	22%	7%	4%	1%	23%
Local mobility	15%	7%	3%	37%	12%	6%	2%	33%
Security	11%	4%	5%	29%	10%	3%	3%	30%
Environment	12%	5%	2%	27%	9%	4%	2%	23%

A variant of the Zanella et al. (2015b) model

Second, we consider a variant of the model proposed by Zanella et al. (2015b). Specifically, we replace the equality constraint $\sum_{j=1}^{n} \lambda_j = 1$ in the CI model (7, p.523) with an inequality constraint$\sum_{j=1}^{n} \lambda_j \leq 1$. As a result, v_{j_0} is not free anymore and the multiplier formulation changes as follows.

Zanella et al. (2015, mod.7-8 p.523)

Primal formulation

$$\max \beta$$

$$\text{s.t.} \sum_{j=1}^{n} b_{kj}\lambda_j \leq b_{kj_0} - \beta g_b \quad \text{for } k = 1, \ldots, l$$

$$\sum_{j=1}^{n} y_{rj}\lambda_j \geq y_{rj_0} + \beta g_y \quad \text{for } r = 1, \ldots, s$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda_j \geq 0 \quad \text{for } j = 1, \ldots, j_0, \ldots, n$$

Multiplier formulation

$$\beta_{j_0} = \min - \sum_{r=1}^{s} y_{rj_0}u_{rj_0} + \sum_{k=1}^{l} b_{kj_0}p_{kj_0} + v_{j_0}$$

$$\text{s.t.} \quad \sum_{r=1}^{s} g_y u_{rj_0} + \sum_{k=1}^{l} g_b p_{kj_0} = 1$$

$$-\sum_{r=1}^{s} y_{rj}u_{rj_0} + \sum_{k=1}^{l} b_{kj}p_{kj_0} + v_{j_0} \geq 0 \quad \text{for } j = 1, \ldots, j_0, \ldots, n$$

$$u_{rj_0} \geq 0 \quad \text{for } r = 1, \ldots, s$$

$$p_{kj_0} \geq 0 \quad \text{for } k = 1, \ldots, l$$

$$v_{j_0} \in \Re$$

$$\phi_r \leq \frac{u_{rj_0}\bar{y}_r}{\sum_{r=1}^{s} u_{rj_0}\bar{y}_r + \sum_{k=1}^{l} p_{kj_0}\bar{b}_k} \leq \psi_r \quad \text{for } r = 1, \ldots, s$$

$$\phi_k \leq \frac{p_{kj_0}\bar{b}_k}{\sum_{r=1}^{s} u_{rj_0}\bar{y}_r + \sum_{k=1}^{l} p_{kj_0}\bar{b}_k} \leq \psi_k \quad \text{for } k = 1, \ldots, l$$

$$\text{where} \quad CI_{j_0} = 1/(1 + \beta_{j_0})$$

Variant Zanella et al. (2015)

Primal formulation

$$\max \beta$$

$$\text{s.t.} \sum_{j=1}^{n} b_{kj}\lambda_j \leq b_{kj_0} - \beta g_b \quad \text{for } k = 1, \ldots, l$$

$$\sum_{j=1}^{n} y_{rj}\lambda_j \geq y_{rj_0} + \beta g_y \quad \text{for } r = 1, \ldots, s$$

$$\sum_{j=1}^{n} \lambda_j \leq 1$$

$$\lambda_j \geq 0 \quad \text{for } j = 1, \ldots, j_0, \ldots, n$$

Multiplier formulation

$$\beta_{j_0} = \min - \sum_{r=1}^{s} y_{rj_0}u_{rj_0} + \sum_{k=1}^{l} b_{kj_0}p_{kj_0} + v_{j_0}$$

$$\text{s.t.} \quad \sum_{r=1}^{s} g_y u_{rj_0} + \sum_{k=1}^{l} g_b p_{kj_0} = 1$$

$$-\sum_{r=1}^{s} y_{rj}u_{rj_0} + \sum_{k=1}^{l} b_{kj}p_{kj_0} + v_{j_0} \geq 0 \quad \text{for } j = 1, \ldots, j_0, \ldots, n$$

$$u_{rj_0} \geq 0 \quad \text{for } r = 1, \ldots, s$$

$$p_{kj_0} \geq 0 \quad \text{for } k = 1, \ldots, l$$

$$v_{j_0} \geq 0$$

$$\phi_r \leq \frac{u_{rj_0}\bar{y}_r}{\sum_{r=1}^{s} u_{rj_0}\bar{y}_r + \sum_{k=1}^{l} p_{kj_0}\bar{b}_k} \leq \psi_r \quad \text{for } r = 1, \ldots, s$$

$$\phi_k \leq \frac{p_{kj_0}\bar{b}_k}{\sum_{r=1}^{s} u_{rj_0}\bar{y}_r + \sum_{k=1}^{l} p_{kj_0}\bar{b}_k} \leq \psi_k \quad \text{for } k = 1, \ldots, l$$

$$\text{where} \quad CI_{j_0} = 1/(1 + \beta_{j_0})$$

where y_{rj_0} and b_{kj_0} respectively refer to the observed r desirable and k undesirable indicator of the evaluated municipality j_0; u_{rj_0} and p_{kj_0} respectively represent the most favourable BoD-weights for the r desirable and k undesirable indicator for the evaluated municipality j_0; y_{rj} and

b_{kj} respectively refer to the r desirable and k undesirable indicator of every municipality j in the dataset; n is the number of municipalities under analysis (n=307); s and l respectively signify the number of municipal functions linked to desirable and undesirable indicators considered in this application (s=5 and l =3).

Looking both at the descriptive statistics in Table E.3 and at scatter plot of ranks, we find that the equality constraint might play a role merely in the unrestricted model. In all the other scenarios, the findings are absolutely the same, confirming that, in our application, disposability on the unitary input are not a matter of concern in the proposed flexible directional distance composite indicator with weight restrictions to measure the service level provision in municipalities. As for the budget shares, the same reasoning applies straightforwardly.

Table E.3: Descriptive statistics of the service provision composite indicator scores estimated for 307 municipalities over 2006-2011

	Zanella et al. (2015)				Variant Zanella et al. (2015)			
	Mean	sd	Min	Max	Mean	sd	Min	Max
Unrestricted Unconditional	0.8388	0.0643	0.6357	1.0000	0.8198	0.0648	0.6357	1.0000
Unconditional								
MinMax restrictions	0.7832	0.0670	0.6302	1.0000	0.7827	0.0669	0.6302	1.0000
Average restrictions	0.7178	0.0649	0.5883	1.0000	0.7178	0.0649	0.5883	1.0000
Municipal-specific restrictions	0.7067	0.0723	0.5121	1.0000	0.7067	0.0723	0.5121	1.0000
Robust Unconditional								
MinMax restrictions	0.9618	0.1098	0.8008	2.0990	0.9531	0.0827	0.8023	1.3977
Average restrictions	0.8752	0.0936	0.7002	1.5770	0.8729	0.0857	0.7003	1.3010
Municipal-specific restrictions	0.8662	0.1041	0.6652	1.8767	0.8632	0.0918	0.6648	1.3650
Robust Conditional Model 1								
MinMax restrictions	0.9753	0.0315	0.8007	1.0013	0.9752	0.0316	0.8009	1.0006
Average restrictions	0.9215	0.0611	0.7034	1.0004	0.9214	0.0612	0.7034	1.0005
Municipal-specific restrictions	0.9158	0.0672	0.6571	1.0003	0.9157	0.0672	0.6600	1.0004

Note: Unrestricted indicates the absence of weight restrictions. *MinMax* restrictions refer to the minimum and maximum share of expenditure in each municipal area across all the municipalities. *Average* restrictions consider the average spending share (lower and upper bound equal to its ±50%). *Municipal-specific* restrictions are based on the municipal-specific current spending allocation (lower and upper bound equal to the ±50% of each municipal spending share). *Model 1* includes the economic and financial characteristics (Fiscal income, Financial debt and Unemployment) and a year dummy.

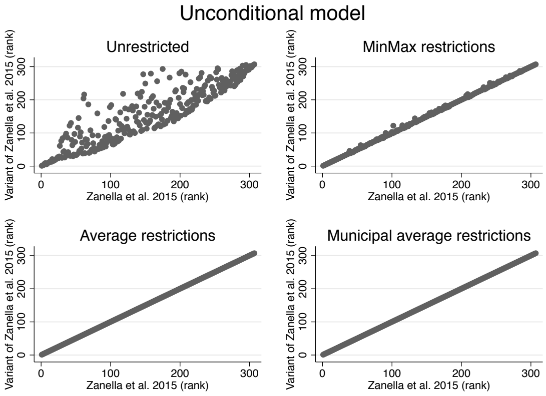

Figure E.7: Zanella et al. (2015) vs Variant of Zanella et al. (2015). Service provision CI scores aggregated at municipal level

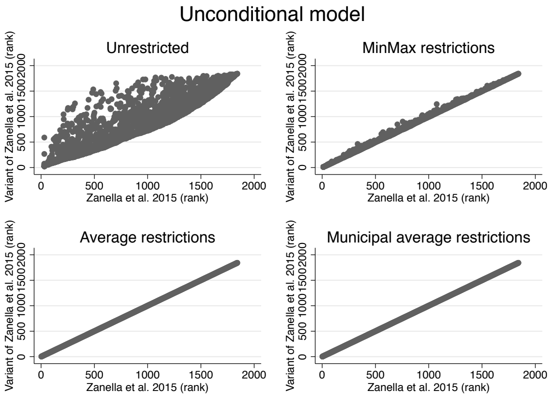

Figure E.8: Zanella et al. (2015) vs Variant of Zanella et al. (2015).

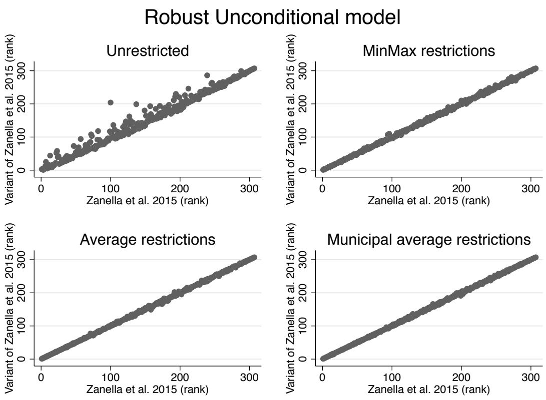

Figure E.9: Zanella et al. (2015) vs Variant of Zanella et al. (2015). Service provision CI scores aggregated at municipal level

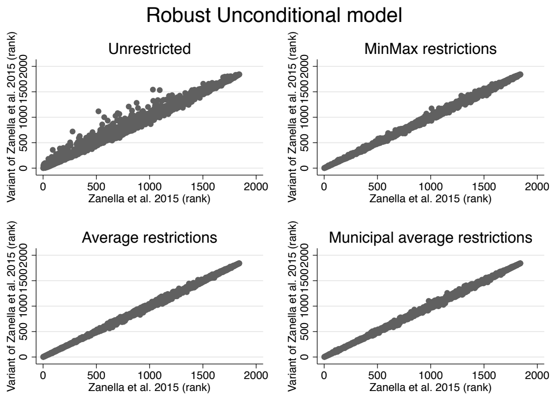

Figure E.10: Zanella et al. (2015) vs Variant of Zanella et al. (2015).

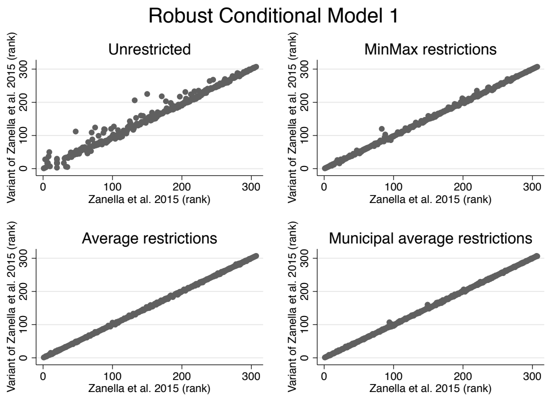

Figure E.11: Zanella et al. (2015) vs Variant of Zanella et al. (2015). Service provision CI scores aggregated at municipal level

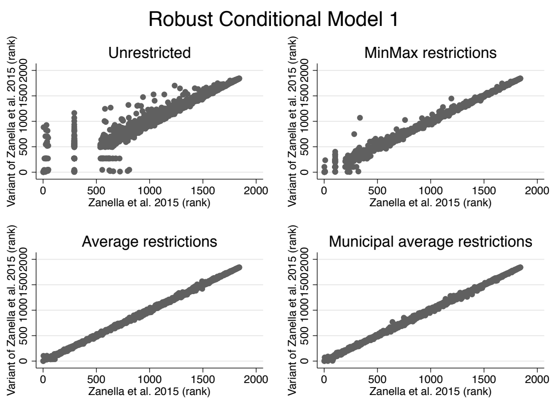

Figure E.12: Zanella et al. (2015) vs Variant of Zanella et al. (2015).

Appendix E.2. Alternative sample specification to control for municipal size effect

In the following, the main robustness check results are presented. Specifically, this further analysis has been performed excluding from the sample the 13 largest Flemish cities (so-called "Centrumsteden"), to make sure that the main findings are not influenced by the municipal size.

Table E.4: Descriptive statistics of the service provision composite indicator scores estimated for 294 municipalities over 2006–2011 (Analysis without "Centrumsteden")

	Mean	St. Dev.	Min	Max
Unrestricted Unconditional	0.8324	0.0639	0.6261	1.0000
Unconditional				
MinMax restrictions	0.7704	0.0674	0.6219	1.0000
Average restrictions	0.7003	0.0632	0.5790	1.0000
Municipal-specific restrictions	0.6895	0.0713	0.4933	1.0000
Robust Unconditional				
MinMax restrictions	0.9501	0.1171	0.7768	2.1155
Average restrictions	0.8596	0.0973	0.6815	1.5953
Municipal-specific restrictions	0.8519	0.1104	0.6593	1.9019
Robust Conditional Model 1				
MinMax restrictions	0.9718	0.0355	0.7820	1.0009
Average restrictions	0.9142	0.0642	0.6893	1.0005
Municipal-specific restrictions	0.9084	0.0703	0.6558	1.0002
Robust Conditional Model 2				
MinMax restrictions	0.9967	0.0098	0.8622	1.0000
Average restrictions	0.9836	0.0283	0.7703	1.0000
Municipal-specific restrictions	0.9821	0.0319	0.7142	1.0000
Robust Conditional Model 3				
MinMax restrictions	0.9981	0.0063	0.9059	1.0000
Average restrictions	0.9900	0.0207	0.7771	1.0000
Municipal-specific restrictions	0.9887	0.0241	0.7194	1.0000

Note: Unrestricted indicates the absence of weight restrictions. *MinMax* restrictions refer to the minimum and maximum share of expenditure in each municipal area across all the municipalities. *Average* restrictions consider the average spending share (lower and upper bound equal to its ±50%). *Municipal-specific* restrictions are based on the municipal-specific current spending allocation (lower and upper bound equal to the ±50% of each municipal spending share).

Model 1 includes the economic and financial characteristics (Fiscal income, Financial debt and Unemployment). *Model 2* adds to the economic and financial characteristics the socio-demographic structure (Share of elderly people, Share of foreigners and Population growth). *Model 3* adds to the economic-financial and socio-demographic variables the political component (Ideological Complexion of the local Government). In every conditional model specification a year dummy is also included.

Table E.5: Influence of background conditions on municipal service composite indicator (Analysis without “Centrumsteden")

MinMax weight restrictions

	Model 1			Model 2			Model 3		
	Influence	p-value		Influence	p-value		Influence	p-value	
Economic-financial									
Fiscal income	Unfavourable	0.000	***	Unfavourable	0.000	***	Unfavourable	0.000	***
Financial debt	Unfavourable	0.000	***	Favourable	0.000	***	Favourable	0.000	***
Unemployment	Unfavourable	0.000	***	Unfavourable	0.075	*	Unfavourable	0.155	
Socio-demographic									
Residents over 65				Favourable	0.995		Favourable	0.005	***
Foreigners				Favourable	0.885		Favourable	0.355	
Population growth				Unfavourable	0.000	***	Unfavourable	0.000	***
Political									
ICG							Unfavourable	0.000	***

Average weight restrictions

	Model 1			Model 2			Model 3		
	Influence	p-value		Influence	p-value		Influence	p-value	
Economic-financial									
Fiscal income	Unfavourable	0.000	***	Unfavourable	0.000	***	Unfavourable	0.000	***
Financial debt	Unfavourable	0.000	***	Favourable	0.000	***	Favourable	0.000	***
Unemployment	Unfavourable	0.000	***	Unfavourable	0.000	***	Unfavourable	0.000	***
Socio-demographic									
Residents over 65				Favourable	0.000	***	Favourable	0.000	***
Foreigners				Favourable	0.345		Favourable	0.570	
Population growth				Unfavourable	0.005	***	Unfavourable	0.000	***
Political									
ICG							Unfavourable	0.000	***

Municipal-specific weight restrictions

	Model 1			Model 2			Model 3		
	Influence	p-value		Influence	p-value		Influence	p-value	
Economic-financial									
Fiscal income	Unfavourable	0.000	***	Favourable	0.000	***	Favourable	0.000	***
Financial debt	Unfavourable	0.000	***	Unfavourable	0.000	***	Unfavourable	0.000	***
Unemployment	Unfavourable	0.000	***	Unfavourable	0.010	**	Unfavourable	0.000	***
Socio-demographic									
Residents over 65				Favourable	0.025	**	Favourable	0.000	***
Foreigners				Favourable	0.005	***	Favourable	0.005	***
Population growth				Unfavourable	0.000	***	Unfavourable	0.000	***
Political									
ICG							Unfavourable	0.000	***

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table E.4 shows the descriptive statistics of the service provision composite indicator scores estimated for 294 municipalities over 2006–2011. Table E.5 presents the statistical influence results for each weight restriction and each conditional model specification.