Received: 28 February 2022 | Revised: 31 March 2023 | Accepted: 23 October 2023

DOI: 10.1111/jors.12673

RESEARCH ARTICLE

JOURNAL OF REGIONAL SCIENCE WILEY

Do municipal unions improve cost efficiency for the social function? A quasi-experimental endogenous stochastic frontier approach

Francesco Vidoli[1] | Fabio Quintiliani[2] | Giorgio Ivaldi[2] | Giorgia Marinuzzi[3] | Francesco Porcelli[4] | Walter Tortorella[3]

[1]Department of Economics, Society and Politics, University of Urbino Carlo Bo, Urbino, Italy

[2]Banca d'Italia, Rome, Italy

[3]IFEL—Fondazione Anci, Rome, Italy

[4]Department of Law, University of Bari, Bari, Italy

Correspondence
Francesco Porcelli, Department of Law, University of Bari, Bari, Italy.
Email: francesco.porcelli1@uniba.it

Abstract

Homogeneous national policies can generate heterogeneous effects on the territory. This paper aims to verify the impact of the entry of single Italian municipalities into an inter-municipal association on the unitary costs of the social services supplied to the local communities. Panel cost stochastic frontier model in the presence of endogeneity has been introduced based on two pillars: a counterfactual setting to verify the aggregation effects concerning a set of similar municipalities and the use of a cost efficiency estimation methodology which considers the endogeneity of output with respect to cost. The results of our analysis show that the effects of adopting associated forms of service delivery can be very multifaceted and diversified according to the typology and the degree of implementation of the Municipal Union itself. The analysis sheds light on the effects of municipalities' organisational choices, addressing the consolidation of small municipalities not in binary terms but also—and above all—suggesting that it is

The views expressed in the articles are those of the authors and do not involve the responsibilities of the employing institutions. Any mistake or inaccuracy is the sole responsibility of the authors.

the type of union that impacts the possibility of convergence towards greater cost efficiency.

KEYWORDS
cost efficiency, difference-in-difference, endogeneity, inter-municipal cooperation, local authorities, social function, stochastic frontier analysis

1 | INTRODUCTION AND RESEARCH QUESTIONS

Italy is often referred to as the "*land of a thousand bell towers*." This saying probably draws from the long history of the communal heritage, dating back to at least the Middle Ages. However, it also highlights the fact that the number of small administrative units nationwide is still very large. At the end of 2020, the total number of Italy's municipalities was 7903 (down from 8092 10 years before and from 8100 20 years before). Small municipalities (i.e., towns with less than 5000 inhabitants) represented about 69% and hosted less than 10 million residents, amounting to 17% of the Italian population. The average size of small towns is just below 1800 inhabitants (as opposed to 7500 at the national level).

Over the past two decades, local populations' needs have evolved significantly in Italy, like in all parts of the EU. The citizens' expectation of the quality level of the services that municipalities provide has increased. At the same time, it has become increasingly difficult for small administrative units to supply these services efficiently or to reach acceptable qualitative levels in providing for them. On this basis and given the relevant number of small municipalities nationwide, several laws have tried to reform how small municipalities provide services to their citizens, trying to induce or "compel" them to do it through mergers with other neighbouring small municipalities or through forms of inter-municipal cooperation.

The first legislative text ever to refer Municipal Unions (henceforth, MUs) was law 142/1990, which at article 26 stated that "*with a view to their subsequent merger, two or more neighbouring municipalities lying within the same province [i.e. county] and having each no more than 5,000 inhabitants, can constitute a Union to carry out a plurality of services*".[1] Law 122/2010 (and Law 135/2012) listed the services and functions that small municipalities were asked to provide jointly, specifying deadlines by which they should have started some form of cooperation. Law 56/2014, on the other hand, specified that inter-municipal cooperation should take place through either a formal covenant (convenzione) signed by the involved municipalities' representatives or a "Municipal Union." MUs constitute supra-municipal bodies with budgetary autonomy and second-degree political representation. The main task of an MU is to provide local public services on behalf of the single municipalities that are part of it. In contrast, the municipalities involved do not have to devolve functions to a different entity through a covenant.

Over the years, the speed of the reform process that could give MUs more effective powers has been significantly reduced by a series of factors, such as annual extensions of the original deadlines by which small municipalities should have started to supply local services cooperatively; lack of effective sanctions for deviant municipalities and lack of indication as to how "efficient cooperation" was meant to be achieved.

[1]Over the years a copious amount of laws has been piling up to regulate both the MUs and the communal mergers. The three most notable acts were Law 42/2009, Law 122/2010 and Law 56/2014. Law 42/2009, in particular, regulated many aspects regarding the MUs and delegated the Italian Regional authorities the definition of the MUs' optimal territory borders within which they could supply their services. On the other hand, Law 122/2010 listed the"fundamental functions" that municipalities must manage and supply to their citizens. Finally, Law 56/2014 (the Delrio Act), which also defined the tasks provincial administrations should carry out, dictated that small municipalities should supply their services through MUs or formal covenants. See Ivaldi et al. (2017) and Manestra et al. (2018), among others, for a more detailed survey of the main laws regulating the MUs.

In this paper, we focus on Italy's MUs, and we try to assess whether their rather widespread adoption has increased cost efficiency in the provision of local social services, especially for small municipalities, or not. Our paper is based on a recent methodology for endogenous cost efficiency estimation proposed by Karakaplan and Kutlu (2019) within a counterfactual setting that thus allows a comparison between municipalities that have chosen to join forms of union and similar municipalities. Focusing on a specific function, social services, allowed us to specify the cost function, the determinants of supply and demand, and the output measure.

Our paper tries to face three critical points in this literature. (i) First, by focusing on how aggregations of municipalities are formed: to this end, it should be pointed out that, for a reform to be effective, it is not so much a matter of whether or not to adopt a form of union or overall optimum sizing, but above all of the way how the union is designed, an element that the legislator has never considered. (ii) Second, the measure of comparison between municipalities and MUs must not only take into account the expenditure side but also—and most importantly—the output that is produced to avoid higher performance being obtained only from the financial side: it is, therefore, necessary to evaluate the cost efficiency and this must be done taking into due consideration the endogeneity between expenditure and out- put. Moreover, in our empirical approach, we are using a"direct" measure of output, avoiding a typical caveat of previous studies where the cost function specification is based only on a proxy of the services produced. (iii) Finally, the aforementioned comparison must be carried out within a counterfactual framework so as not to obtain biased or otherwise unreliable results.

The paper is organized as follow: Section 2 introduces our contribution to the literature on MUs and efficiency estimation methods; the Italian institutional framework is described in Section 3 and the empirical approach in Section 4. The application of the proposed methodology to the Italian case is reported in Section 5, highlighting the strong heterogeneity of the effects of MUs; Section 6 concludes.

2 | LITERATURE

A large body of literature deals with the problem of assessing if and why small administrative units should join forces to supply services to their local communities. The debate, whose start could be traced as far back as the Fifties of the 20th century, can be for simplicity's sake polarized under two conflicting headings. On the one hand, one view supports the idea that a large number of local governments is efficiency-enhancing because polycentric governance units generate competition, which in turn fosters efficiency (Ostrom et al., 1961; Tiebout, 1956). Consumers of local public services choose to locate in territories where the municipalities can best satisfy their needs. In other terms, "*citizens can shop with their feet and locate in a community that best maximizes their preferences*" (Goodman & Leland, 2013).

On the other hand, the opposite view maintains that the growing complexity of the local population's needs can only be effectively and efficiently satisfied by large-sized municipalities: small administrative units should, therefore, join forces either through covenants or common bodies or possibly mergers.[2] The effective provision of local public services often requires a variety of specialized personnel and capital goods, which in turn entails that a minimum scale of operation has to be reached (Bel et al., 2010; Hulst et al., 2009).

The debate was revived in the aftermath of the 2007–2008 financial crisis, as the austerity measures adopted in many countries worldwide resulted in both significant cuts in national transfers to municipalities and in constraints being applied to their expenditures (Bel & Warner, 2015). With the aim of both relaxing the budget constraints and supporting the efforts of local administrations to provide local services efficiently, inter-municipal

[2]While the quest for efficiency is probably the most relevant reason why small administrative units may want to cooperate, other reasons may include: (i) relaxation of possible expenditures or fiscal constraints defined by the central government; (ii) wealth and income level of local communities, which may be inversely correlated with the incentives to join forces; (iii) social homogeneity in local communities, whereby similarly structured communities may more easily cooperate; (iv) improvements in the quality services provided locally (Aulich et al., 2014; Bel & Warner, 2016; Hepburn et al., 2004).

cooperation and/or amalgamation were advocated even by many international institutions, such as UNDP-CoE (2010)[3] and OECD (OECD 2015a, 2015b). As a result, inter-municipal cooperation in local services provision has been adopted in many countries, such as in Spain with the "comarcas" or in France with the "communautès de communes" or in Germany with the "Gemeindeverbande" (see e.g., Manestra et al., 2018 for a comprehensive review).

At the same time the empirical literature increased significantly too, as a large number of studies were published employing a rich variety of parametric and non-parametric approaches. Their common aim was to assess whether the very numerous experiences of inter-municipal cooperation around the world were in fact proving efficiency-enhancing (an extensive and comprehensive review of the existing literature on local governments' efficiency can be found in Narbon-Perpina & De Witte, 2018).

The economies of scale resulting from being able to produce at a larger scale remain the key factor in many analyses: Bönisch and Haug (2018) found that a decentralized institutional setting improves the efficiency of municipal production, while Reingewertz (2012), through a difference-in-differences (DID) methodology calculates that on average there is a decrease of about 9% in municipal expenditures finding no evidence of a decrease in the level of services provided. Pérez-Lopez et al. (2018) also indicate that joint forms of management achieve greater efficiency of scale in the long run.

More in general Tavares (2018), in his literature review, highlight the effects of mergers in three areas, economic efficiency—in which it is shown that cost savings are mainly limited to general administration function—on managerial implications—highlighting limited change in the quality of services provided—and in democratic outcomes showing a slight degradation in local representation. Gendźwiłł et al. (2021) also points out that in many applications based on quasi-experimental research designs the result clearly confirms savings in administrative spending, while for other areas no economies of scale are demonstrated.

Waste collection service is one of the functions with the highest intensity of improvement due to cooperation between municipalities: Bel and Mur (2009) found that small towns that cooperate incur in lower costs, an evidence confirmed especially for municipalities with a sub-optimal size (Bel et al., 2014). Savings are dependent not only on the scale and density economies and externalities, but also to the structure of local government and the governance framework at the local/regional level where cooperation may vary from informal to formal (Bel & Warner, 2015). It is evident how union is not the only form of cooperation between Local Authorities: Pérez-Lopez et al. (2023) compare the relative cost of waste collection services under inter-municipal cooperation and autonomous production by local governments in Spain from 2009 to 2015 showing how significant cost savings have been produced by representing a viable policy option for municipal functions that benefit from a multi-production service approach.

Few analyses are available at the level of Italian municipalities: Ivaldi et al. (2017) used a non-parametric free disposal hull approach developed by Deprins et al. (1984) and found mixed evidence as to the supposed higher expenditure efficiency of MUs in providing local services as opposed to single municipalities. In particular, differences across Italian regions emerged, whereby, only in some regions, municipalities in MUs reached higher efficiency scores comparing with those not included in MUs. Ferraresi et al. (2018), adopting quasi-experimental methodologies, found that MU reduces the total per capita current expenditures by around 5% without affecting the level of local public service.[4] Luca and Modrego (2021) fail to find any positive effect, in terms of efficiency, of municipal cooperation among Italian municipalities using robust data envelopment analysis (Banker et al., 1984) and fuzzy regression discontinuity design around a threshold of 5000 inhabitants; compared to the exercise proposed in this paper, with a similar database, the authors use only 1 year and employ non-parametric techniques that do not

[3]UNDP-CoE went so far as to publish a handbook for inter-municipal cooperation to the benefit of small administrations aiming to join forces and provide services (especially large-scale ones, like waste management, water supply or public transport).

[4]However, measured indirectly only through exogenous demand factors. Instead, our approach uses a "direct" measure of output in terms of total social care sector users restricting the cost function's specification only to social care.

allow them to instrument the output and solve the problem of heterogeneity in a one-stage framework. More recently, Di Ielsi et al. (2022), using our same source of data, show that in Italy, at least two different models of MUs emerge: the "Piemontese" model and the "Emilia-Romagna" model with different structures and different levels of performance. In particular, this last evidence strongly motivates our analysis of the performance heterogeneity that different types of MUs may generate.

All the studies mentioned so far, however, and also others related to the estimation of the efficiency of municipalities (see e.g., Blochliger & Charbit, 2008; Dafflon & Mischler, 2007; Reschovsky, 2007), do not consider the intimate link that exists between the standard level of local expenditure and the standard level of local services, since attention is focused mainly on the financial side of the problem.

The originality of the proposed exercise, therefore, lies in enhancing the information on the level of the service provided, which is usually not available, allowing as in Porcelli et al. (2016), Porcelli and Vidoli (2020) and Vidoli and Fusco (2018) to stress the absolute need to consider the cost function (and the relative endogeneity issues) in public setting; in other terms, without taking the directly produced output into account there is a risk of large overestimate the benefits of consolidation. In addition, the counterfactual setting and the endogenous model allow us to strengthen the estimates by controlling for the self-selection of Municipalities that have chosen Unions and bypassing simultaneous estimation problems with respect to endogenous variables, respectively.

Finally, the heterogeneity of the different forms of collaboration is a key issue: in other terms, the way in which change is achieved is by no means independent of the final results. Allers and van Ommeren (2016), for example, stress the difference between amalgamation and inter-municipal cooperation showing how the latter consistently pays higher interest rates on loans due to higher inefficiency; Blesse and Baskaran (2016) similarly but with a different angle, focus on whether mergers are compulsory or not, finding significant reductions in administrative costs after compulsory mergers and no effect at all for voluntary mergers.

3 | INSTITUTIONAL FRAMEWORK

There are presently just over 7900 municipalities in Italy: about 70% of municipalities are small (i.e., featuring less than 5000 inhabitants each) mainly located along the Alps and the Apennines. All in all about 10 million people (i.e., one sixth of total population) live in these towns. In about a quarter of municipalities the population lies between 5000 and 20,000 inhabitants, while in 5% of cases it lies between 20,000 and 60,000 citizens. Municipalities with more than 60,000 inhabitants represent a mere 1% of the overall number.

Small municipalities are the target territories for the reform process that stretched over the last two decades with a view to increasing the number of cases of associated management of functions, thereby enhancing efficiency in public spending and in the provision of local public services. Law no. 142/1990 introduced the Municipal Unions,[5] as an inter-municipal cooperation tool for the joint management of functions and supply of services. Their introduction aimed at helping small municipalities to increase efficiency in the provision of services, without altering their identities nor the individual administrations and councils. In 2010 (Law no. 122, art. 14) this tool (and formal covenant) became mandatory[6] for small municipalities, but no specific organisational and functional model has been required; because of this, there are different forms of municipal unions, and each type is more or less widespread across Italy's regions.

How many MUs are there in Italy? Unlike in the case of single municipalities, the answer to this question is not as simple as one might expect, since there exists no official list of MUs: ISTAT (the Italian statistical office) publishes

[5]Members of the MUs' governing councils are chosen from (and elected by) the members of the city councils of the single municipalities that form the union.

[6]They are mandatory but no sanctions are applied if municipalities do not oblige. Furthermore, the obligation has been postponed many times and is still suspended today.

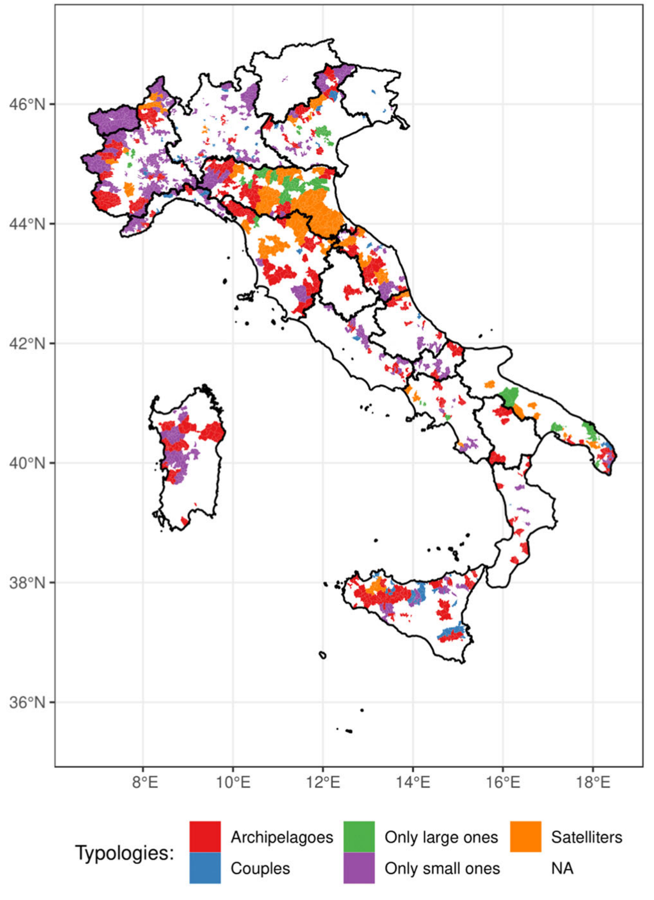

FIGURE 1 Municipal unions by typology, year 2015. *Source*: "Cittalia-Fondazione ANCI Ricerche" (2010). [Color figure can be viewed at wileyonlinelibrary.com]

annually a list of all Italian administrative bodies which includes MUs, but some of them may provide no effective services and, therefore, may exist only on paper. For this reason, we followed a more pragmatical approach, based on cross-referencing different databases (ANCI, Ministry of Interior, ISTAT, MEF, and SIOPE)[7] and identifying 524 MUs[8] across Italy in 2015 (see Figure 1). This year has been chosen because the output measure we use in our econometric analysis is available only for some years between 2010 and 2016.

Most of the MUs are located in Piedmont (98), Lombardy (77), Sicily (52), and Emilia-Romagna (45) involving 2969 municipalities, 37% of the 8047 municipalities existing in 2015. This share, however, hides the stark heterogeneity that exists across Italian regions, ranging from 98.6% to 86.8% of all municipalities respectively in Aosta's Valley and in Emilia- Romagna to 4.2% and 2.5% in the case of Friuli-Venezia Giulia and Trentino-South Tyrol.

[7]In particular as of December 2021 databases are accessible as follows: ANCI (https://www.anci.it/unioni-di-comuni/); Ministry of Interior (https://dait.interno.gov.it/territorio-e-autonomie-locali/sut/elenco_unioni_comuni_comp.php); ISTAT (https://www.istat.it/it/archivio/190748); MEF (https://www.rgs.mef.gov.it/VERSIONE-/e_government/amministrazioni_pubbliche/siope/elenco_delle_amministrazioni/); SIOPE (https://www.siope.it/Siope/).

[8]We classified as active the MUs for which we have evidence of at least one payment having been made according to the SIOPE database. Then we used all the other dataset available to define the member municipalities for each active MU.

Such regional heterogeneity depends both on different regional legislation, on the diverse presence of small municipalities across territories, and on the different forms of unions created and is clearly a key factor in performance evaluation. For these reasons, to account for the relevant heterogeneity, the taxonomy of MUs proposed in "Cittalia-Fondazione ANCI Ricerche" (2010) has been adopted and MUs have been classified according to it as follows:

- "Couples," referring to MUs featuring only two municipalities (13%);
- "Only small ones," if formed only by municipalities (at least 3) with less than 5000 inhabitants: they represent 45% of all Italian MUs;
- "Archipelagoes," if they include municipalities (at least 3) of different sizes whose total MUs' population sums up to 15,000 residents (28% of total MUs);
- "Satelliters," when MUs are formed by at least 3 municipalities both small and one or two very large municipalities, the latter with over 15,000 inhabitants each (9% of total MUs);
- "Only large ones," if MUs are formed only by at least 3 large municipalities (5%).

In line with the regional heterogeneity reported in Figure 1, Di Ielsi et al. (2022) provide a clear analysis of two different regional models: the "Piemontese" model where unions are mainly formed by only small municipalities and archipelagos; and the "Emilia-Romagna" model in which unions are formed around a municipality medium-sized leader (Satellites) or involve only large municipalities. These two models are a clear results of two forces: the different structure of municipalities; and the different regional incentives and legislation. Di Ielsi et al. (2022) propose a descriptive analysis that shows the higher performance of Emilia-Romagna model compared to the Piemontese model using the same source of data adopted in our research (OpenCivitas portal). This evidence motivates our in depth analysis of the performance heterogeneity of different types of municipal unions.

4 | EMPIRICAL APPROACH

The proposed methodological and empirical framework are intended to be an up-to-date and rigorous approach to three different branches of research: the approach to cost function estimation in the public domain (subsection 4.1) and the endogenous parametric estimation methods of stochastic frontiers (subsection 4.2) within a counterfactual setting (subsection 4.3). Here we discuss each branch. Nevertheless, this discussion cannot be exhaustive, and we refer to the literature for further specific details.

4.1 | Estimating cost function for local governments

The "pure" Regression base Cost Approach (RCA) can be estimated using a structural model of demand and supply of local public goods, in which the method of instrumental variables is used to solve the endogeneity issues related to the historical level of output. This strategy is very difficult to implement and the results' accuracy, based on several assumptions, is not easy to verify. The risk, therefore, is to obtain biased estimates.

More specifically, following an RCA approach the variables selection is based on a theoretical model in which the demand for public services expressed by the citizens interacts with the supply of public services provided for by local governments. In this framework the efficient cost for each service depends on three basic dimensions: the optimal quantity of service offered, the prices for the inputs used in the production process (primarily labour costs) and the contextual variables related to the supply (i.e., the external factors that, all things being equal, can favour or hinder the supply of local public goods—e.g., the morphological characteristics of the territory or the surface area and so on).

Formally in this framework the unitary local public service supply can be expressed in terms of unitary cost as:

$$y = s(g, \quad p, \quad A), \tag{1}$$

where g are endogenous outputs, p the input price vector and A the supply variables that represent morphological and socio-economic constraints changing the unitary service costs.

The unitary demand for local public service (g) can be expressed as:

$$g = d(Q, \quad R, \quad y), \tag{2}$$

where Q are demographic and socio-economic contextual variables, R is the average income and y represents the unitary cost of local public service.

Please also note that the presence of the output variables within the estimation model presents three disadvantages: (i) outputs are not always measurable and there may be a very serious lack of information. (ii) Even if the public service outputs were observable with extreme accuracy, they might be external to the initial theoretical model (e.g., when the central authority establishes specific service levels to which individual local authorities must conform) (iii) outputs might be endogenous, since their optimal quantity is determined simultaneously in terms of expenditures; in this case, to obtain unbiased estimations of the output's weight (price) it is necessary to use two-stage estimation techniques by finding the correct instrumental variables (from a practical point of view this statistical approach is very difficult, if not, sometimes, impractical).

In this setting, the cost function (3) is obtained by the simultaneous estimation of the Equations (2) and (1):

$$\begin{cases} y = s(g, p, A) \\ g = d(Q, R, y) \end{cases}. \tag{3}$$

4.2 | Stochastic frontier literature

The public service activity—aimed at the provision of public utilities established by law—requires that local administrations act in the most appropriate and convenient way possible to ensure the effectiveness, efficiency, speed, and economy of administrative action itself.

To translate these general principles into concrete objectives, efficiency measurement techniques have been introduced in the public sector since the early 1980s using a multitude of parametric and non-parametric techniques. Parametric stochastic production frontier (SFA) models, introduced by Aigner et al. (1977) and Meeusen and van den Broeck (1977), is the most widely used technique for the evaluation of public and private enterprises (see Greene, 2008; Kumbhakar & Lovell, 2000 and Narbon-Perpina & De Witte, 2018 for overviews of developments in this area).

More specifically, SFA specify a single output in terms of a response function and a composite error term, where the compound error consists of a two-sided error representing random effects and a one-sided term representing technical inefficiency. In general terms, it can be defined as:

$$y_i = f(\mathbf{x}_i, \boldsymbol{\beta}) + v_i - u_i, i = 1, \ldots, n, \tag{4}$$

where $y_i \in R^+$ is the output of Municipality unit i, $\mathbf{x}_i \in R^{p+}$ is the vector of inputs, $f(\cdot)$ defines a production (frontier) relationship between inputs X and the output Y, v_i is the symmetric two-sided error representing random effects and $u_i > 0$ the one-sided error term which represents technical inefficiency. Given these premises, technical efficiency is defined as $TE_i = \exp(u_i)$ (typically y_i is log-output).

In applications, the two-sided error term is usually[9] assumed to be normally distributed: $v \sim N(0, \sigma^2)$ while u is distributed half-normally on the non-negative part of the real number line: $u \sim N^{+}(0, \sigma^2)$; moreover, following common practice, it is assumed that v and u are each identically independently distributed (*i.i.d.*).

Starting from the model reported in Equation (4)—if time series are available—a multitude of panel models (see for instance, the Battese and Coelli (1988) time-invariant model, the Battese and Coelli (1992) time-varying model and the Greene (2005) "true" fixed-effects (TFE) and random-effects (TRE) models have been presented in the literature to deal with the underlying longitudinal correlation structure more appropriately.

But, while the more specific panel aspects have attracted the attention of many authors, less attention has been paid to the simultaneity aspects especially in cost functions, both for average cost and especially for minimum cost functions.

As stated by Brinkman (1985), in fact, for OLS, "one of the standard assumptions, expressed intuitively, is that causality runs in one direction only in the regression equation, from the independent variable(s) to the dependent ones"; simultaneous-equation bias occurs, therefore, when mutual influence between cost and its determinants is in place. More in particular—within stochastic frontier models—"endogeneity problems can arise due to a couple of major reasons: first, the determinants of the cost frontier and the two-sided error term can be correlated [and] second, the inefficiency term and two-sided error term can be correlated, or in particular, the determinants of the inefficiency can cause this correlation" (Karakaplan & Kutlu, 2019). If the determinants of the frontier or the inefficiency term are correlated with the two-sided error term of the model, then the outcomes of the standard estimators would be contaminated by endogeneity. In other terms, source of endogeneity can be identified in the presence of omitted variables and simultaneity (between X and the two-sided error term) or because of the correlation between the two-sided error term and v.

Given this background, the topic of endogeneity in cross-sectional frontiers has been debated in a limited, but very interesting literature that includes mainly the contributions of Greene (2008); Amsler et al. (2016); Griffiths (2016); Karakaplan and Kutlu (2017) and Amsler et al. (2017).

Only a few years ago, finally, Karakaplan and Kutlu (2019) proposed a panel version of their Karakaplan and Kutlu (2017) endogenous model allowing the estimation of a panel cost frontier model in the presence of endogeneity; the Karakaplan and Kutlu (2019) endogenous SFA panel model can be expressed as follows:

$$\begin{cases} y_{it} = x'_{yit}\beta + v_{it} - su_{it} \\ x_{it} = Z_{it}\delta + \varepsilon_{it} \\ with \\ u_{it} = h\left(x'_{uit}\varphi_u\right)u_i^* \\ \begin{bmatrix} \tilde{\varepsilon}_{it} \\ v_{it} \end{bmatrix} = \begin{bmatrix} \Omega^{-1/2}\varepsilon_{it} \\ v_{it} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I & \sigma_v\rho \\ \sigma_v\rho & \sigma_v^2 \end{bmatrix}\right), \end{cases} \tag{5}$$

where $y_{it} \in R^{+}$ is the output of unit i at time t, $x_{it} \in R^{p+}$ is the vector of inputs, v_{it} is the symmetric two-sided error representing random effects, s is equal to 1 for production functions and −1 for cost functions and $u_{it} > 0$ is the one-sided error term which represents technical inefficiency.[10] The covariates are split into three groups: x_{yit}, the exogenous and endogenous variables explaining y; x_{it}, the endogenous variables; and x_{uit}, the exogenous variables explaining u. The endogeneity of x_{it} is corrected using Z_{it}, the vector of all exogenous instrumental variables. In this framework, $u_{it} > 0$ is, therefore, a one-sided error term capturing the inefficiency depending by a vector of exogenous and endogenous variables and by a producer-specific random component u_i^* independent from v_{it} and ε_{it}. Finally, Ω is the variance–covariance matrix of ε.

[9]Many distributions can be assumed for the one-sided error term; e.g., half-normal, truncated normal, gamma, exponential, etc.

[10]The two-sided residual term is usually assumed to be normally distributed: $v \sim N(0, \sigma_v^2)$ while u is distributed as a half-normal and is always positive: $u \sim N^{+}(0, \sigma_u^2)$. The classical model also assumes that v and u are each identically independently distributed (i.i.d.) and the covariates in the model.

4.3 | Quasi-experimental design

Another source of bias may arise from comparing different groups of municipalities in terms of cost efficiency: the self-selection bias, that arises in any situation in which individuals select themselves into a group. To overcome this bias, the hypothesis of being in presence of purely experimental data must be assumed or, in other terms, the exogeneity in the choice to pool among municipalities.

It is trivial to note, conversely, how, in the last 10 years in Italy, the choice to enter a MU has been left to the municipality itself: key variables such as the municipality size or the spending efficiency can, therefore, predetermine whether or not municipalities have chosen the associated forms, arising to biased results.

A DID framework can provide a solution to this shortcoming, and it is often used in empirical research in economics to evaluate the effects of public interventions and other treatments of interest. The basic idea (for a more exhaustive review, please see Angrist & Pischke, 2009; Donald & Lang, 2007; Imbens & Wooldridge, 2008) is to use a quasi-experimental design by reconstructing a matching set of untreated municipalities which, before the policy, had similar trends with respect to the treated ones and then assessing, after the policy, the differences (in our case, the difference between cost efficiencies) between the treated and untreated municipalities.

To ensure the estimation accuracy, the standard approach requires that the cost variable to be compared before and after treatment should be the same; unfortunately, this cannot be the case for this exercise due to the lack of information about the output for the years before 2010 (see Figure 2 for clarity and Section 5.1 for more details about the structure and source of data). Therefore, a non-standard analysis procedure has been chosen.

In other terms, given the absence of information on outputs before 2010, the matching phase (see Figure 2) has been carried out by evaluating the municipalities in terms of expenditure function (i.e., expenditure and supply and demand factors), while the evaluation of the effect of the union has been expressed in terms of cost (i.e., also taking into account the output produced). Obviously, the underlying hypothesis—which cannot be circumvented given the lack of data—is that before joining the union, similar municipalities produced similar outputs that were perfectly correlated with local supply and demand factors.

After the matching phase, therefore, it is necessary to evaluate a unitary cost function starting with the definition of Equation (3) and including the information about the membership of a union (dummy d) and the year (t) in which the municipality became part of a union:

$$\begin{cases} y = s(g, p. A, d, t) \\ g = d(Q, R, y) \end{cases}, \tag{6}$$

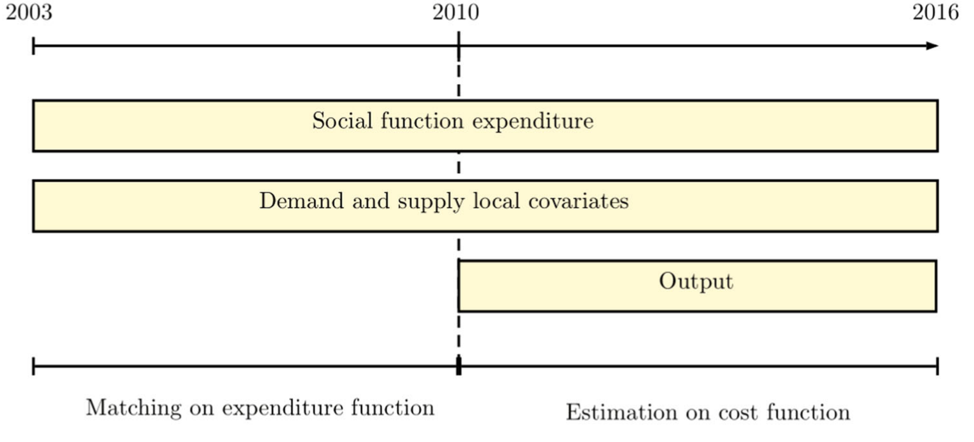

FIGURE 2 Overview of data availability by year.

and, taking advantage of Karakaplan and Kutlu (2019) endogenous SFA panel model given in the Equation (5), finally, the empirical cost model ($s = -1$ in the inefficiency term) can be written as:

$$\begin{cases} y_{it} = \beta_1 A_{it} + \beta_2 p_{it} + \beta_3 g_{it} + v_{it} + u_{it} \\ g_{it} = Z_{it}\delta + \varepsilon_{it} \\ with \\ u_{it} = \gamma x_{uit} + d + t + dt + \varepsilon'_{it} \end{cases}, \tag{7}$$

where $dt = d \cdot t$ is the key treatment DID effect parameter.

Finally, to identify treated and control observations in the matching step, the panel matching proposed by Imai et al. (2021) has been chosen[11]; this matching method is particularly helpful when time-series cross-section data are present creating a matched set for each treated observation and refining the matched set via matching/weighting methods thanks to additional adjustments for past outcomes and confounders.

5 | APPLICATION TO THE ITALIAN MUNICIPALITIES

Our goal is to identify the impact that the joint management of services can generate in terms of efficiency, i.e. producing more output at the same cost or reducing the unit cost per output. Restricting the analysis to just social care services diminishes the possibility of listing economies of scale as a potential source of efficiency enhancement, as usually highlighted in the literature (Bönisch & Haug, 2018; Pérez-Lopez et al., 2018; Reingewertz, 2012).

As reported in the official methodological notes of municipal standard expenditure needs released by the Italian Ministry of Finance (see Porcelli, 2015 and Porcelli et al., 2016), social care is a labour-intensive sector. Therefore, expenditure per user (a proxy of the average cost) does not decrease with the increase of the municipal size in terms of the resident population, ruling out the possibility of increasing economies of scale.

To mitigate the risk of confounding congestion phenomena, typical of the social care sector, with inefficiency, we have introduced the second-order polynomial of the resident population in all specifications. At the same time, following Di Ielsi et al. (2016), each municipality enters the regression sample with its size in terms of the resident population since the efficiency gains resulting from municipal associations could be distorted if associations are compared with single municipalities with the same features (especially in terms of population size). Instead, these efficiency gains would be correctly measured only if the municipalities participating in an association (and not the association itself) are directly compared with similar municipalities not involved in any association forms, constructing a synthetic counterfactual through matching techniques.

5.1 | Data and estimation strategy

The empirical framework proposed in Section 4 has been applied to the Italian Municipalities data to check whether, and for which forms of union, a cost differential was significant.

The paper takes advantage of a recent and innovative dataset constructed within the Standard Expenditure Needs project started with Legislative Decree No. 216 of November 26, 2010 and subsequently published online on OpenCivitas (www.opencivitas.it) by the Italian Ministry of Finance and SOSE S.p.A., a government-controlled company that periodically collects data from local governments through dedicated surveys validated by a Technical Commission instituted by the Ministry of Economy. It is important to underline here that the fulfilment of these

[11]Thanks to PanelMatch R package.

surveys is mandatory only for the 6565 Italian municipalities located in the 15 Ordinary Regions, as data are used for calculation of the expenditure needs of each municipality and the allocation of State funds across local governments located only in ordinary regions. The survey collects information about expenditures, divided by functions: general administration, local police, education, social services, road maintenance and public transports, territory management and environment. In addition to data about expenditures, the surveys require information about the volume of operations and activities (see Agasisti & Porcelli, 2022 for more details). The information provides a unique example of statistical and administrative effort to collect specific direct indicators of services' outputs on a wide array of areas, instead of solely indirect proxies for the demand of services. The dataset covers all the municipalities in the Ordinary Regions for four different consistent time frames: 2010, 2013, 2015, 2016.

To take into account, the evident heterogeneity among the different fundamental missions carried out by municipalities, it has been decided to evaluate the cost, the output and the related determinants only for the social function.[12] In other terms, different functions mean different economies of scale—optimal sizing or the ability or not to provide for services in an aggregated form.

Social care function in Italy is also interesting—in this exercise—for another reason: municipalities do not receive funds only from the central state, but also, they are mediated by the regions to which they belong. For this reason, different organisational models are present in each region, making it possible to verify cost differentials for different forms of union. The analysed dataset (see Table 1) consists of a panel of 4 years after 2010 (treatment year) with reliable[13] and non-abnormal information for approximately 850 municipalities in unions[14] and 650 matched municipalities not in union.[15]

In the endogenous stochastic cost frontier model—proposed in Equation (7)—the per-capita cost is related to the demand factors, the input prices and the social services output instrumented with the supply and demand determinants of the output itself; the treatment variable (post-2010 union membership) and the scale variables describe the cost inefficiency (please see Table 2 for main descriptive statistics).

The variables reported in Table 2 are all taken from the OpenCivitas database and cover the municipalities belonging to ordinary statute over a time span of 4 years (2010, 2013, 2015, and 2016).[16]

In particular the main variables included in the cost function are: the per capita cost that corresponds to the total expenditure allocated for the provision of social serviced divided by the total resident population, the social services output is used to comprehensively identify the amount of output actually delivered by municipalities and is measured using the OpenCivitas official composite indicator that considers several variables reflecting the total number of people assisted for various social problems, such as immigrants, elderly, abandoned minors, etc. divided by the total resident population. It is important to emphasize that in this way, we can employ a "direct" measure of output in the cost function specification, avoiding the adoption of proxies of the services produced. Population is designed to consider the size of the municipalities and to identify the likely presence of economies of scale in the provision of public goods (the quadratic term to identify possible non-linearities was also considered). Real estate prices per square meter aim to represent a proxy for the price of the capital factor, while the average salary per worker in the private sector is a proxy for the labour costs. Finally, the main contextual variables used to measure the external factors that may affect the service provision are the surface area of the municipality, and expenditure allocated to the provision of administrative services in per capita terms.

[12]The social function is related to family (young and elders), poverty, disabilities, mental health, immigration, addiction.

[13]More precisely, municipalities with missing data and identified as anomalous by the procedure proposed by Filzmoser et al. (2008) on the variables of interest in the analysis have been deleted.

[14]Please note that, although in theory municipalities can withdraw from the union, in the case under review this—at least in the years analysed—never happened.

[15]Output information was only available for the years 2010, 2013, 2015, and 2016. For this reason we had to omit the years 2013 and 2015.

[16]In the empirical analysis we focus only on the "ordinary statute regions" because the output variable is only available for the municipalities that are administratively included in them. We therefore exclude from the analysis the "special statute regions" (namely Aosta's Valley, Trentino-S¨ud Tyrol, Friuli-Venezia Giulia, Sicily and Sardinia).

TABLE 1 Panel of municipalities analysed by union (yes/no) and year of presence of information.

Year	2011	2012	2014	2016
Treated				
No	624	659	629	559
Yes	824	893	891	728

TABLE 2 Main descriptive statistics and their relative role in analysis.

Statistic	Role	Mean	Standard deviation	Min	Pctl(25)	Pctl(75)	Max
Social services per-capita total cost (Euro)	Dependent variable	71.14	69.70	4.77	37.68	82.11	1395.49
Population (*N*)	Returns to scale	4395	8710	36	822	4372	123,130
Social services per-capita total output	Output	0.864	0.5	0	0.786	0.906	3.82
Real estate prices (Euro per sq. meter)	Capital price	4.315	1.686	1.265	3.102	5.165	11.473
Average salary per worker (Euro)	Labour price	30,258	1959	24,392	28,854	31,603	34,544
Surface area in km^2	Supply	0.023	0.040	0.0002	0.005	0.026	0.658
Expend. administrat. (per capita)	Supply	303.829	195.039	64.647	184.301	356.634	1903.874
Social deprivation index	Demand (IV)	13.746	30.056	100.000	34.198	2.306	275.132
% Populat. 0–2 years	Demand (IV)	2.259	0.719	0.000	1.825	2.701	6.977
Incidence index elderly pop.	Demand (IV)	0.249	0.056	0.081	0.213	0.279	0.608
Social target	Supply (IV)	3.633	1.942	0.000	1.000	5.000	6.000
Social facilities (dummy)	Supply (IV)	0.669	0.471	0.000	0.000	1.000	1.000
Average income per capita	Income	18,623.880	3235.454	6178.329	16,604.300	20,574.210	40,216.540

Source: OpenCivitas database.

As for the instrumental demand variables they are related to the Social deprivation index calculated as a weighted average of five indices related to low school attendance, the unemployment index, the index of rented houses compared to the total number of houses used for housing, the index of family density and an index of economic hardship; to the Percentage of children (0–2 years) of daycare service age; to the Incidence of the elderly population as a percentage of the total; to Social target which is a politomic variable (values from 1 to 6) that aims to identify both the presence of the service and, in part, its intensity with a maximum of 6 if the municipality offers a wider range of services; and to the presence or absence (Social facilities) of direct service delivery points within the municipality; finally, the Average income per capita declared for income tax purposes is a proxy of the local well-being.

5.2 | Empirical strategy

In this section, we discuss the identification of the matched set based on 2004–2009 municipal expenditure and supply/demand data, to reconstruct a counterfactual set for the municipalities that have given rise to Unions since 2010.

The Imai et al. (2021) matching method to time-series cross-sectional data constitutes a crucial advancement in standard techniques because it allows to "*link each treated observation from a given unit in a particular time period with control observations from other units in the same time period that have a similar treatment and covariate history.*" The matching is, therefore, not only drawn on two different time periods (before and after MUs), but also by comparing the trends of several significant covariates.

This method, based on covariate balanced propensity score matching, allows to measure the similarities/distances among units including lagged versions of any variable as well; in our application, therefore, demand variables such as the number of residents aged 0–14, 15–64, and over 65 (lagged 2004–2009); accounting variables such as municipal expenditure per capita for social function (lagged 2004–2009) and price variables such as average income per capita (lagged 2004–2009); and fixed effects supply variables such as the altitude of the municipality, housing density, mountain and coastal dummies, have been used.

To better reflect the specificity and structure of the municipalities treated in the period after 2009 and those never treated, parallel trends both in absolute terms (Figure 3 left side), to grasp the dimension, and in relative terms (Figure 3 right side), are reported showing approximately optimal parallel trends after matching.

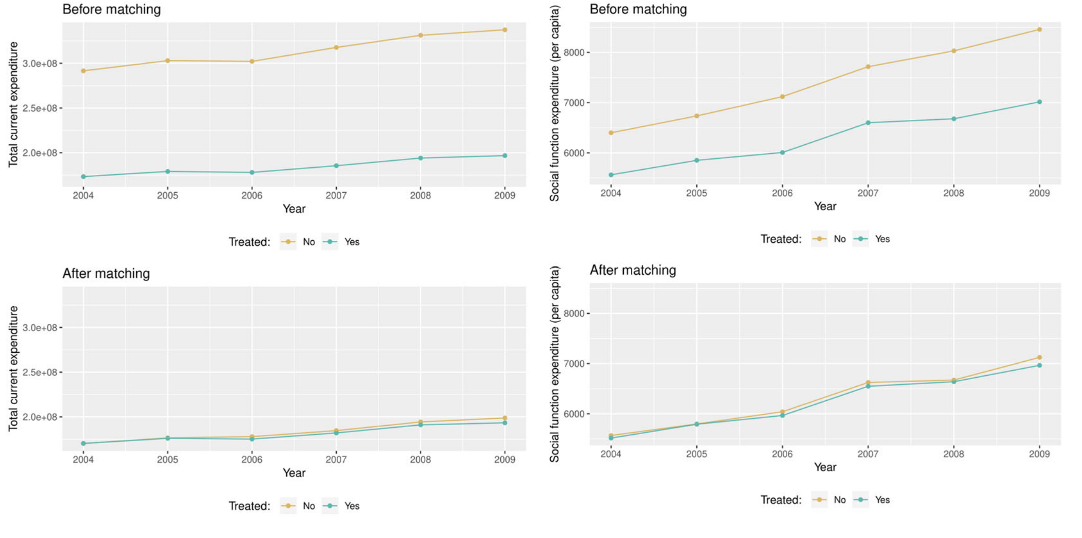

FIGURE 3 Parallel trends before and after matching. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 3 *t*-Test between treated and not treated between and after the matching phase.

Matching	Statistic	*p*-value	Mean (euro)	Mean (euro)
Social expenditure (per capita)			Not treated	Treated
Before	6.274	0.000	89.95	76.68
After	−1.713	0.087	73.74	76.39
Total inhabitants				
Before	30.255	0.000	7336.42	3448.29
After	0.769	0.442	3497.73	3407.08
Average income				
Before	−2.856	0.004	18,738.73	18,832.79
After	−2.462	0.014	18,716.69	18,837.92
Total current expenditure				
Before	24.587	0.000	5,714,572.84	2,633,834.33
After	0.899	0.368	2,679,755.13	2,592,634.78
Total current staff expenditure				
Before	22.890	0.000	1,856,852.00	846,962.54
After	0.552	0.581	852,080.64	834,753.15

Table 3 confirms how in terms of the average distribution of the two groups for major structural variables—between and after the matching phase—the matching procedure has been effective in allowing a comparison between similar municipalities.

5.3 | Baseline model

After reconstructing the two groups of treated and untreated municipalities, a linear baseline model linking per-capita cost and the main supply and demand factors has been investigated.

Table 4 shows the random-effects GLS regression estimates: it should be noted the U-shape of the population, the (positive) significance of the composite indicator of social service output and input prices and, finally, the positive estimate of the dummy variable treatment, which shows that, on a very first analysis, municipalities that choose associated forms present higher costs.

These initial and incomplete insights are also confirmed by the panel stochastic frontier model reported in Table 5 where the parameter σ_v represents the variance of the random error and σ_u of inefficiency; the coefficient of the γ statistic equal to 0.649 shows that about 65% of the distance of the units from the minimum cost frontier is attributable to inefficiency. Membership of the treated group always shows a positive and significant estimate.

The two introductory models, which made it possible to evaluate the single contribution of covariates in the variation of per-capita costs, present at the same time two main limitations that must be overcome: (i) the first one concerning the presence among the covariates of an endogenous variable such as the “social services composite output” on the frontier equation [first and second lines of the Equation (7)] allowing this variable to be instrumented through appropriate demand variables and (ii) the second one concerning the possibility of identifying, again in one-stage, the determinants of inefficiency [third line of the Equation (7)].

TABLE 4 Baseline random-effects GLS regression model (years 2011–2016).

	Coef.	*p* > *z*
Population	−0.0694	0.485
Population (squared)	0.0169	0.010
Social services composite output	0.0849	0.000
Real estate prices	0.111	0.001
Average salary	0.823	0.000
Treated (0/1)	0.137	0.000
Constant	−5.227	0.002
Observations	5772	

Note: Dependent variable = *Social services per-capita total cost*.

TABLE 5 Panel stochastic frontier model.

	Coef.	Standard error	*z*	*p* > *z*	[95% Confidence interval]	
Population	−0.020	0.084	−0.240	0.808	−0.186	0.145
Population (squared)	0.013	0.006	2.410	0.016	0.003	0.024
Social services composite output	0.062	0.010	6.370	0.000	0.043	0.081
Real estate prices	0.088	0.026	3.360	0.001	0.036	0.139
Average salary	0.729	0.121	6.040	0.000	0.492	0.965
Treated (0/1)	0.108	0.025	4.250	0.000	0.058	0.158
Constant	−5.515	1.278	−4.320	0.000	−8.019	−3.011
σ^2	0.385	0.014			0.359	0.414
γ	0.649	0.014			0.621	0.676
σ_u^2	0.250	0.014			0.223	0.277
σ_v^2	0.135	0.003			0.129	0.142

Note: Dependent variable = *Social services per-capita total cost*.

The endogenous panel stochastic frontier models proposed by Karakaplan and Kutlu (2019) makes it possible to describe the frontier in terms of exogenous covariates, to instrument the endogenous variables (social services composite output, in our case) and to identify the factors that explain a greater or lower inefficiency, all in a single step.

Table 6 confirms the main findings of the two baseline models with several important differences: (i) in this model the social composite output indicator is instrumented by a plurality of significant variables such as deprivation indices and indices related to the elderly and newborns; (ii) the DID term is positive and significant; please note that, in this specification, the counter-factual term has been more correctly, compared to the approach in Table 5, introduced to explain the cost inefficiency term [u_{it} in the Equation (7)] and not so much the deterministic part (the frontier) of the cost. This from an interpretation point of view is a definite advantage since it allows us to directly associate treatment and inefficiency; (iii) other scale covariates (per-capita surface area of the municipality) and per-capita administrative expenditure are used to better control the inefficiency part.

TABLE 6 Endogenous panel stochastic cost frontier model.

	Coef.	Standard error	z	p > z	95% Confidence interval	
Frontier estimation						
Population	0.065	0.076	0.850	0.396	−0.085	0.214
Population (squared)	0.010	0.005	1.990	0.047	0.000	0.019
Social services output	0.517	0.053	9.800	0.000	0.414	0.620
Real estate prices	0.069	0.025	2.780	0.005	0.020	0.118
Average salary	1.408	0.114	12.310	0.000	1.184	1.633
Constant	−12.138	1.239	−9.790	0.000	−14.566	−9.709
Instrumental variable estimation for output						
Social deprivation index	−0.001	0.000	−2.480	0.013	−0.001	0.000
% Populat. 0–2 years	0.044	0.014	3.120	0.002	0.016	0.071
Incidence index elderly pop.	0.723	0.197	3.670	0.000	0.337	1.110
Social target	0.041	0.004	9.170	0.000	0.032	0.050
Social facilities	0.182	0.018	9.910	0.000	0.146	0.217
Average income	0.000	0.000	6.570	0.000	0.000	0.000
Constant	−1.187	0.095	−12.470	0.000	−1.374	−1.001
Inefficiency term						
Treated (0/1)	0.245	0.073	3.360	0.001	0.102	0.388
Surface area in km^2	4.397	1.099	4.000	0.000	2.242	6.551
Expend. administrat. (percapita)	0.001	0.000	3.910	0.000	0.000	0.001
Constant	−0.939	0.081	−11.540	0.000	−1.098	−0.779

Note: Dependent variable = *Social services per-capita total cost*.

But what is the real contribution of having used a more complex model that deals with the endogeneity of output and also allows the inefficiency part to be instrumented?[17] There are two main findings that allow us to do this.

The first one is an endogeneity test proposed by Karakaplan and Kutlu (2019) that test if correction for endogeneity is not necessary (H_0) or if there is endogeneity in the model and correction is needed (Ha); in our case, test $\chi^2(1) = 73.37$ (Prob > $\chi^2 = 0$) leads us to reject the null hypothesis.

The second one is linked with the comparison between the exogenous and the endogenous model (reported in Table 7) in which the "exogenous model" is the model that ignores endogeneity, while "endogenous model" is the model that handles it. Please note a substantial stability in all the coefficients between the two models except for the instrumented variable which highlights a substantial difference between the two approaches and confirms what Karakaplan and Kutlu (2019) claimed: "*if not handled, the endogenous covariates may have a larger bias effect on the*

[17]Two types of test have been evaluated to verify the IV variables: (i) the tests of underidentification (Anderson canon. corr. LM statistic) to verify if instruments are irrelevant ($p < 0.00001$ allows us to reject this hypothesis) and (ii) the weak identification test (Stock & Yogo (2005) method) to evaluate if instrument are weak; also in this case, we can reject the null hypothesis that our instruments are weak ($F >$ J10) because in our case $F = 22.19$ is greater than J10 = 11.12; finally, with regard to the exogeneity of the instrumental variables chosen, it should be noted that they are all not dependent on the endogenous choices of the municipalities themselves.

TABLE 7 Comparison between exogenous and endogenous frontier specifications.

	Exogenous model		Endogenous model	
	Coef.	*p* > *z*	Coef.	*p* > *z*
Population	0.046	0.076	0.065	0.076
Population (squared)	0.012	0.005	0.010	0.005
Social services output	0.075	0.010	0.517	0.053
Real estate prices	0.092	0.025	0.069	0.025
Average salary	1.334	0.115	1.408	0.114
Constant	−11.511	1.250	−12.138	1.239
Dep.var:	Inefficiency			
Treated (0/1)	0.266	0.073	0.245	0.073
Surface area in km^2	3.627	1.042	4.397	1.099
Expend. administrat. (per capita)	0.001	0.000	0.001	0
Constant	−0.914	0.082	−0.939	0.081
Observations		5589		5589
Log Likelihood		4004.84	−9279.44	

Note: Endogeneity test $\chi^2 = 73.37$, $p = 0.000$.

frontier parameters and the endogenous variables in the one-sided error term may have a larger bias effect on the efficiency parameters."

Table 7, finally, shows the positive effect on cost inefficiency of being part of the treated group[18]: indeed, it seems that, on average, belonging to a MU leads, at least in the short term, to an increase in per-capita costs. Can we take this as a definitive result, seemingly inconsistent with general findings in the literature, or are there still some critical keys missing? A very interesting key to interpretation can be found in the strong heterogeneity of Italian municipalities as well as differences in the MUs' typologies.

In the model specification shown in Table 8 the DID parameter has been multiplied by the different types discussed in Section 3. Important differences emerge in terms of explaining cost inefficiency: more in detail, the unions of "only large" municipalities seem to obtain a positive effect from adherence to associated forms, while those of the "satellite" group yield strongly negative effects, at least in the short term.

It is, therefore, necessary, in our opinion, to deal with the issue of consolidation of small municipalities not in binary terms, but instead through the analysis of the type of association form. In other terms, it is not only a question of whether or not to adopt a form of union or optimal overall sizing, but more importantly of designing in what way and for what types of municipality it is useful to do so.

This interpretation of the effects of the association between municipalities is certainly more interesting and richer than only the average effect leaving aside the very different forms of association. But another and final step allows us to evaluate the dynamics of this effect, too. By making the treatment parameter interact with the type of

[18]Please note that, working with Time-Series Cross-Sectional data, we might be in the case of staggered adoption design (SAD), a special case of the general difference-in-differences (DID) set up where units can switch back and forth between being exposed or not to the treatment (Athey & Imbens, 2022); this could potentially be true, but in our present case (i.e., in the short term, 4 years after treatment) no Municipality has left the relevant Union. Regarding the random assignment of the adoption of the treatment, however, a caveat must be added: the choice to join the relevant union may in some cases not be fully exogenous, i.e. favoured only by regional authorities or national policies, but may also be decided by the municipality itself. It is therefore a weak point in our identification that cannot be bypassed.

TABLE 8 Endogenous panel stochastic cost frontier model, treated × typology; in parentheses the numbers of municipalities treated for the 4 years analysed.

	Coef.	Standard error	z	p > z	95% Confidence interval	
Frontier estimation						
Population	0.0449	0.0889	0.5000	0.6140	−0.1294	0.2192
Population (squared)	0.0119	0.0057	2.0900	0.0370	0.0007	0.0231
Social services output	0.4398	0.0602	7.3000	0.0000	0.3218	0.5579
Real estate prices	0.0314	0.0288	1.0900	0.2760	−0.0251	0.0878
Average salary	1.0934	0.1419	7.7000	0.0000	0.8152	1.3716
Constant	−8.8383	1.5188	−5.8200	0.0000	−11.8152	−5.8615
Instrumental variable estimation for output						
Social deprivation index	−0.0013	0.0004	−3.2800	0.0010	−0.0020	−0.0005
% Populat. 0–2 years	0.0489	0.0172	2.8400	0.0040	0.0152	0.0826
Incidence index elderly pop.	0.5911	0.2468	2.3900	0.0170	0.1073	1.0749
Social target	0.0496	0.0058	8.5300	0.0000	0.0382	0.0611
Social facilities	0.1647	0.0229	7.1800	0.0000	0.1198	0.2096
Average income	0.0000	0.0000	2.7100	0.0070	0.0000	0.0000
Constant	−1.0196	0.1203	−8.4800	0.0000	−1.2553	−0.7838
Inefficiency term (Treated × Typology)						
Treated × Archipelagoes (1255)	0.1775	0.1421	1.2500	0.2120	−0.1010	0.4559
Treated × Couples (57)	0.5648	0.4030	1.4000	0.1610	−0.2251	1.3548
Treated × Only large ones (135)	−1.0324	0.3073	−3.3600	0.0010	−1.6347	−0.4300
Treated × Only small ones (1463)	0.1099	0.1403	0.7800	0.4330	−0.1651	0.3849
Treated × Satelliters (507)	0.6154	0.1706	3.6100	0.0000	0.2810	0.9498
Surface area in km^2	2.8495	1.0540	2.7000	0.0070	0.7836	4.9153
Expend. administrat. (per capita)	0.0005	0.0001	3.4000	0.0010	0.0002	0.0008
Constant	−0.8592	0.1372	−6.2600	0.0000	−1.1282	−0.5903

Note: Dependent variable = *Social services per-capita total cost*.

association form, and with the posttreatment year (see Table 9), it is possible to study the average dynamics of treated unions by type, verifying whether the positive effect (on cost inefficiency) is temporary or persistent.

Thanks to this very detailed estimates, which takes due account of the dynamics of the different typologies (Figure 4) within an endogenous cost framework, it is possible to capture the differences in a more detailed form.[19] In particular, the estimated DID coefficients reported in Table 9 are shown as points, separately by type and year; the smooth regression curves are also shown to better appreciate the overall

[19]Please note that from a strictly statistical point of view, the relatively small numerosity and the few years of posttreatment lead us to evaluate the dynamics and its significance from an economic point of view and perspective more carefully than the strict statistical significance. It is evident that the clearest results are for large unions, but the shift from higher to lower efficiency is also evident for 'pairs' and'satelliters'.

TABLE 9 Endogenous panel stochastic cost frontier model, treated × typology × year.

	Coef.	Standard error	z	p > z	95% Confidence interval	
Frontier estimation						
Population	0.077	0.089	0.870	0.385	−0.097	0.252
Population (squared)	0.010	0.006	1.740	0.081	−0.001	0.021
Social services output	0.422	0.055	7.640	0.000	0.314	0.531
Real estate prices	−0.002	0.029	−0.060	0.949	−0.059	0.055
Average salary	1.516	0.156	9.690	0.000	1.209	1.822
Constant	−13.281	1.660	−8.000	0.000	−16.535	−10.027
Instrumental variable estimation for output						
Social deprivation index	−0.001	0.000	−2.840	0.005	−0.002	0.000
% Populat. 0–2 years	0.023	0.017	1.380	0.169	−0.010	0.057
Incidence index elderly pop.	0.406	0.243	1.670	0.095	−0.071	0.882
Social target	0.061	0.006	10.430	0.000	0.049	0.072
Social facilities	0.125	0.023	5.480	0.000	0.080	0.170
Average income	0.000	0.000	1.550	0.122	0.000	0.000
Constant	−0.903	0.123	−7.370	0.000	−1.143	−0.663
Inefficiency term (Treated × Typology × Year since foundation)						
Treated × Archipelagoes (year 0)	0.331	0.149	2.220	0.027	0.039	0.623
Treated × Archipelagoes (year 1)	0.192	0.157	1.230	0.220	−0.115	0.499
Treated × Archipelagoes (year 2)	0.185	0.213	0.870	0.386	−0.233	0.603
Treated × Archipelagoes (year 3)	0.056	0.159	0.350	0.726	−0.256	0.367
Treated × Archipelagoes (year 4)	0.312	0.197	1.580	0.114	−0.074	0.698
Treated × Archipelagoes (year 5)	0.186	0.302	0.620	0.537	−0.406	0.778
Treated × Couples (year 0)	0.478	0.392	1.220	0.222	−0.290	1.245
Treated × Couples (year 1)	0.368	0.531	0.690	0.488	−0.672	1.408
Treated × Couples (year 2)	−0.124	0.695	−0.180	0.859	−1.486	1.239
Treated × Couples (year 3)	−0.092	0.740	−0.120	0.901	−1.543	1.359
Treated × Couples (year 4)	−0.717	1.498	−0.480	0.632	−3.652	2.219
Treated × Couples (year 5)	(empty)					
Treated × Only large ones (year 0)	−0.569	0.527	−1.080	0.280	−1.601	0.463
Treated × Only large ones (year 1)	−0.778	0.573	−1.360	0.174	−1.901	0.345
Treated × Only large ones (year 2)	−0.579	0.467	−1.240	0.215	−1.493	0.335
Treated × Only large ones (year 3)	−1.606	0.807	−1.990	0.047	−3.188	−0.024
Treated × Only large ones (year 4)	−1.156	0.498	−2.320	0.020	−2.133	−0.179
Treated × Only large ones (year 5)	−1.101	0.652	−1.690	0.091	−2.379	0.178

TABLE 9 (Continued)

	Coef.	Standard error	*z*	*p* > *z*	95% Confidence interval	
Treated × Only small ones (year 0)	0.090	0.146	0.610	0.540	−0.197	0.376
Treated × Only small ones (year 1)	0.053	0.156	0.340	0.735	−0.253	0.359
Treated × Only small ones (year 2)	0.287	0.169	1.700	0.089	−0.044	0.618
Treated × Only small ones (year 3)	0.034	0.166	0.210	0.836	−0.291	0.360
Treated × Only small ones (year 4)	0.341	0.174	1.960	0.050	−0.001	0.683
Treated × Only small ones (year 5)	0.331	0.319	1.040	0.299	−0.294	0.956
Treated × Satelliters (year 0)	0.756	0.182	4.150	0.000	0.399	1.113
Treated × Satelliters (year 1)	0.605	0.196	3.080	0.002	0.220	0.990
Treated × Satelliters (year 2)	0.729	0.201	3.630	0.000	0.335	1.123
Treated × Satelliters (year 3)	0.341	0.218	1.560	0.118	−0.086	0.769
Treated × Satelliters (year 4)	0.461	0.219	2.100	0.036	0.031	0.890
Treated × Satelliters (year 5)	−0.037	0.515	−0.070	0.943	−1.046	0.972
Surface area in km^2	3.059	1.071	2.860	0.004	0.960	5.157
Expend. administrat. (per capita)	0.000	0.000	3.280	0.001	0.000	0.001
Constant	−0.880	0.136	−6.450	0.000	−1.147	−0.612

Note: Dependent variable = *Social services per-capita total cost*.

trend, again by type of union; a positive value indicates that inefficiency is increasing, while a value less than zero indicates that inefficiency is decreasing and thus that the aggregate form provides cost savings.

It is, therefore, possible to understand: (i) how the effect of the union may imporve efficiency for the “couples” and “satellite” typologies, but that cost savings only occur after 2 or 3 years; (ii) how the unions of large municipalities are immediately more efficient thanks to a greater administrative structure; and (iii) finally how the “small” and ”archipelago” unions are not able, at least in the short and medium term (4 years), to counterbalance a greater administrative structure with lower management costs (these results are broadly in line with the descriptive analysis provided by Di Ielsi et al., 2022).

In our opinion, it is crucial to understand whether and under what conditions Unions of Municipalities allow for cost savings addressing two primary sources of heterogeneity. Our results show that the first source of heterogeneity can be identified in the structure of the association determined either by morphological features or by regional incentives and legislation, elements that are very diversified across Italy; both these forces give rise to two models: one characterised by larger municipalities or by small municipalities associated around one large city (this model can be found mainly in Emilia-Romagna); a second model where only small municipalities form the association (this model is spread all over the country and is prominent in Piemonte). Clearly, our results show that the former model is more efficient than the latter. The dynamics of the performance capture the second source of heterogeneity over time that allows us to trace what trajectory and how many years must pass on average before the reorganisation of the service pays off. Again, our results show that only municipal associations that include at least one large municipality improve the overall efficiency over time.

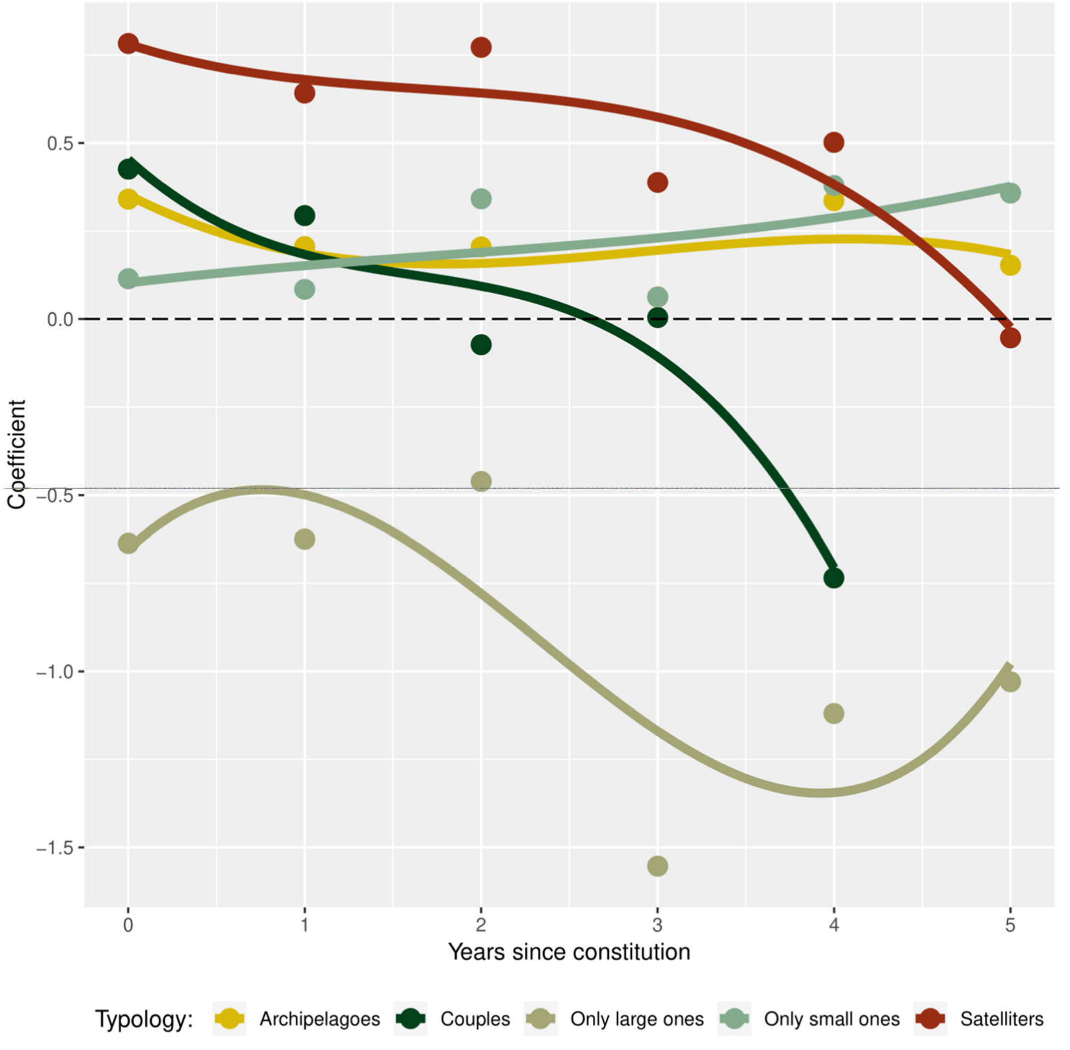

FIGURE 4 Dynamics of the inefficiency terms by associative typology—smooth over years since constitution. [Color figure can be viewed at wileyonlinelibrary.com]

6 | FINAL REMARKS

In this paper, the question was asked whether or not joining an associated form by individual municipalities would lead to a decrease in unitary costs; we answered this question by adopting a counterfactual framework—years 2004–2009 for matching and 2010–2016 for impact analysis—that would allow us to verify the evolution of endogenous post-treatment cost efficiency over time.

The analysis has been carried out on a novel dataset for the social function of Italian Municipalities that allowed us to have the usual local supply and demand variables at our disposal and, above all, a measure of the output provided. This measure allowed us to ensure that the cost variations did not occur by simply cutting the services provided to the population. Therefore, standard expenditure analysis that does not consider the corresponding supply may lead to sub-optimal policy implications and recommendations. The results of our analysis show that the effects of adopting associated forms of service delivery can be very multifaceted and diversified according to the typology and the degree of implementation of the MU itself. In other terms, heterogeneity among aggregation forms in Italy is the salient feature in estimating cost efficiency; a one-size policy does not fit all in this case.

A different type of association also corresponds to a different response following the decision to join a MU. While larger municipalities achieve cost savings immediately, smaller ones take 2–3 years to offset higher coordination costs and rationalise the service supply network, but only when at least one large municipality is in the association. These savings are not evident for associations composed only of small municipalities, at least in the post-intervention time window analysed.

The results obtained in this paper, although regarding only the social service function, may help to explain why (especially for the Italian case) the existing empirical literature shows contrasting evidence on the benefit derived from the municipal association, providing at the same time valuable tips for the policymakers. First of all, is the association constructed around at least one large municipality that emerges as the most efficient model (this model is widespread in Emilia-Romagna and should be encouraged, when possible, also in other regions), instead the association model that includes only tiny municipalities may lack the required expertise to improve efficiency. In conclusion, our results show that achieving higher efficiency after the association involves time, so the analysis of the possible benefits should be measured over the long run. Hence the association among municipalities should be seen as a structural process.

It is clear, however, that further and more in-depth analyses can indeed be carried out for the other main functions of municipalities and in the presence of many post-intervention years data.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Francesco Porcelli https://orcid.org/0000-0002-6803-4368

REFERENCES

Agasisti, T., & Porcelli, F. 2022. Local governments' efficiency and its heterogeneity – empirical evidence from a stochastic frontier analysis of Italian municipalities 2010-2018. *Applied economics*, Forthcoming. https://doi.org/10.1080/00036846.2022.2107988

Aigner, D., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, *6*, 21–37.

Allers, M. A., & van Ommeren, B. (2016). Intermunicipal cooperation, municipal amalgamation and the price of credit. *Local Government Studies*, *42*(5), 717–738.

Amsler, C., Prokhorov, A., & Schmidt, P. (2016). Endogeneity in stochastic frontier models. *Journal of Econometrics*, *190*(2), 280–288.

Amsler, C., Prokhorov, A., & Schmidt, P. (2017). Endogenous environmental variables in stochastic frontier models. *Journal of Econometrics*, *199*(2), 131–140.

Angrist, J., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion* (1 edn.). Princeton University Press.

Athey, S., & Imbens, G. W. (2022). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*, *226*(1), 62–79.

Aulich, C., Sansom, G., & McKinlay, P. (2014). A fresh look at municipal consolidation in Australia. *Local Government Studies*, *40*(1), 1–20.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, *30*, 1078–1092.

Battese, G. E., & Coelli, T. J. (1988). Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics*, *38*(3), 387–399.

Battese, G. E., & Coelli, T. J. (1992). Frontier production functions, technical efficiency and panel data: With application to paddy farmers in India. *Journal of Productivity Analysis*, *3*, 153–169.

Bel, G., Fageda, X., & Mur, M. (2014). Does cooperation reduce service delivery costs? Evidence from residential solid waste services. *Journal of Public Administration Research and Theory*, *24*(1), 85–107.

Bel, G., Fageda, X., & Warner, M. E. (2010). Is private production of public services cheaper than public production? a meta-regression analysis of solid waste and water services. *Journal of Policy Analysis and Management*, *29*(3), 553–577.

Bel, G., & Mur, M. (2009). Intermunicipal cooperation, privatization and waste management costs: Evidence from rural municipalities. *Waste Management*, *29*(10), 2772–2778.

Bel, G., & Warner, M. E. (2015). Inter-municipal cooperation and costs: Expectations and evidence. *Public Administration*, *93*(1), 52–67.

Bel, G., & Warner, M. E. (2016). Factors explaining inter-municipal cooperation in service delivery: A meta-regression analysis. *Journal of Economic Policy Reform*, *19*(2), 91–115.

Blesse, S., & Baskaran, T. (2016). Do municipal mergers reduce costs? Evidence from a German federal state. *Regional Science and Urban Economics*, *59*, 54–74.

Blochliger, H., & Charbit, C. (2008). Fiscal equalisation. *Oecd Journal: Economic Studies*, *1*, 1–22.

Bönisch, P., & Haug, P. (2018). The efficiency of local public-service production: The effect of political institutions. *Finanz-archiv*, *74*(2), 260–291.

Brinkman, P. T. (1985). Simultaneous-equation bias in higher-education cost functions. *Research in Higher Education*, *23*(2), 201–218.

Cittalia-Fondazione ANCI Ricerche. (2010). *Lo stato delle unioni. rapporto 2010 sulle unioni di comuni*. Anci-Cittalia.

Dafflon, B., & Mischler, P. (2007). Expenditure needs equalisation at the local level: Methods and practice. In J. Kim & J. Lotz (Eds.), *Measuring local government expenditure needs*. Copenhagen Workshop.

Deprins, D., Simar, L. L., & Tulkens, H. (1984). *The performance of public enterprizes: Concepts and measurements*. North-Holland.

Donald, S. G., & Lang, K. (2007). Inference with difference-in-differences and other panel data. *Review of Economics and Statistics*, *89*(2), 221–233.

Ferraresi, M., Migali, G., & Rizzo, L. (2018). Does intermunicipal cooperation promote efficiency gains? Evidence from Italian municipal unions. *Journal of Regional Science*, *58*(5), 1017–1044.

Filzmoser, P., Maronna, R., & Werner, M. (2008). Outlier identification in high dimensions. *Computational Statistics & Data Analysis*, *52*(3), 1694–1711.

Gendźwiłł, A., Kurniewicz, A., & Swianiewicz, P. (2021). The impact of municipal territorial reforms on the economic performance of local governments. A systematic review of quasi-experimental studies. *Space and Polity*, *25*(1), 37–56. https://doi.org/10.1080/13562576.2020.1747420

Goodman, C., & Leland, S. (2013). Cost shocks and their relationship to the creation, consolidation and dissolution of us local governments. *Public Finance and Management*, *13*(2). https://pfm.spaef.org/article/1435/Cost-shocks-and-their-relationship-to-the-creation-consolidation-and-dissolution-of-US-local-governments

Greene, W. (2005). Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis*, *23*, 7–32.

Greene, W., H. O. Fried, C. A. K. Lovell & S. S. Schmidt, eds. (2008). The econometric approach to efficiency analysis, *In*: *The measurement of productive efficiency and productivity change*. Oxford University Press.

Griffiths, W. E. & Hajargasht, G. (2016). Some models for stochastic frontiers with endogeneity. *Journal of Econometrics*, *190*(2), 341–348. https://doi.org/10.1016/j.jeconom.2015.06.012

Hepburn, N., La Sage, E., & McMillan, M. (2004). *Shared service arrangements: Determinants of success*. Western Centre for Economic Research.

Hulst, R., van Montfort, A., Haveri, A., Airaksinen, J., & Kelly, J. (2009). Institutional shifts in inter-municipal service delivery. *Public Organization Review*, *9*(09), 263–285.

Di Ielsi, G., Fiorillo, F., & Porcelli, F. (2022). Le unioni di comuni in Italia: Modelli di gestione associata a confronto. *Moneta e credito*, *75*(295), 11–40.

Di Ielsi, G., Porcelli, F., & Zanardi, A. (2016). La valutazione dell'efficienza nelle forme associate dei comuni italiani: la lezione dei fabbisogni standard. *Economia Pubblica*, *1*, 37–58.

Imai, K., Kim, I. S., & Wang, E. H. (2021). Matching methods for causal inference with time-series cross-sectional data. *American Journal Of Political Sciencen*/a(n/a).

Imbens, G., & Wooldridge, J. (2008). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(09), 5–86.

Ivaldi, G., Marinuzzi, G., Quintiliani, F., & Tortorella, W. (2017). Quali confini? Territori tra identit'a e integrazione internazionale. *Franco Angeli*.

Karakaplan, M., & Kutlu, L. (2017). Handling endogeneity in stochastic frontier analysis. *Economics Bulletin*, *37*(2), 889–901.

Karakaplan, M. U., & Kutlu, L. (2019). School district consolidation policies: Endogenous cost inefficiency and saving reversals. *Empirical Economics*, *56*(5), 1729–1768.

Kumbhakar, S. C., & Lovell, C. A. K. (2000). *Stochastic frontier analysis*. Cambridge University Press.

Luca, D., & Modrego, F. (2021). Stronger together? Assessing the causal effect of inter- municipal cooperation on the efficiency of small Italian municipalities. *Journal of Regional Science*, *61*, 261–293.

Manestra, S., Messina, G., & Peta, A. (2018). *L'unione (non) fa la forza? alcune evidenze preliminari sull'associazionismo comunale in Italia*. Questioni di economia e finanza. Bank of Italy.

Meeusen, W., & van den Broeck, J. (1977). Efficiency estimation from cobb-douglas production functions with composed error. *International Economic Review*, *18*, 435–444.

Narbón-Perpiñá, I., & De Witte, K. (2018). Local governments' efficiency: A systematic literature review—Part I. *International Transactions in Operational Research*, *25*(2), 431–468. https://doi.org/10.1111/itor.12364

OECD. (2015a). *Governing the city*. OECD Publishing.

OECD. (2015b). *The metropolitan century*. OECD Publishing.

Ostrom, V., Tiebout, C. M., & Warren, R. (1961). The organization of government in metropolitan areas: a theoretical inquiry. *American Political Science Review*, *55*(4), 831–842.

Pérez-López, G., Prior, D., & Zafra-Gómez, J. L. (2018). Temporal scale efficiency in dea panel data estimations. an application to the solid waste disposal service in Spain. *Omega*, *76*, 18–27.

Pérez-L'opez, G., Tran, C.-T. T. D., & Dollery, B. (2023). Is council co-operation cost efficient? An empirical analysis of waste collection in Spanish local government. *Public Money & Management*, *43*(2), 105–115.

Porcelli, F. (2015). The evaluation of standard expenditure needs of municipalities: The case of social care services in Italy. *The Italian journal of public economics*, *3*, 123–157.

Porcelli, F., & Vidoli, F. (2020). A comprehensive model for the evaluation of standard expenditure needs and standard level of local services. *Local government studies*, *46*(5), 734–762.

Porcelli, F., Vidoli, F., & Dispotico, R. (2016). Public sector economics and the need for reforms. MIT Press, CESifo.

Reingewertz, Y. (2012). Do municipal amalgamations work? Evidence from municipalities in Israel. *Journal of Urban Economics*, *72*(2–3), 240–251.

Reschovsky, A. (2007). Intergovernmental fiscal transfers: Principles and practice. R. Boadway and A. Shah.

Stock, J., & Yogo, M. (2005). Testing for weak instruments in linear iv regression. *Pages 80–108 of*: In: Andrews, Donald W.K. (Ed.), *Identification and inference for econometric models*. New York: Cambridge University Press.

Tavares, A. F. (2018). Municipal amalgamations and their effects: A literature review. *Miscellanea geographica*, *22*(1), 5–15.

Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of Political Economy*, *64*(5), 416–424.

UNDP-CoE. (2010). Toolkit manual: Inter-municipal cooperation. UNDP-CoE.

Vidoli, F., & Fusco, E. (2018). Level of services, spatial dependence and allocative efficiency in local governments. *Local Government Studies*, *44*(6), 848–873.

How to cite this article: Vidoli, F., Quintiliani, F., Ivaldi, G., Marinuzzi, G., Porcelli, F., & Tortorella, W. (2023). Do municipal unions improve cost efficiency for the social function? A quasi-experimental endogenous stochastic frontier approach. *Journal of Regional Science*, 1–25. https://doi.org/10.1111/jors.12673