International Center for Public Policy
Working Paper 21-16
November 2021

Evidence on Economies of Scale in Local Public Service Provision: A Meta-Analysis

Juan Luis Gómez-Reino
Santiago Lago-Peñas
Jorge Martinez-Vazquez

Georgia State University
Andrew Young School of Policy Studies

International Center for Public Policy
Working Paper 21-16

Evidence on Economies of Scale in Local Public Service Provision: A Meta-Analysis

Juan Luis Gómez-Reino
Santiago Lago-Peñas
Jorge Martinez-Vazquez
November
2021

International Center for Public Policy
Andrew Young School of Policy Studies
Georgia State University
Atlanta, Georgia 30303
United States of America

Phone: (404) 413-0235
Fax: (404) 651-4449
Email: paulbenson@gsu.edu
Website: http://icepp.gsu.edu/

International Center for Public Policy
Andrew Young School of Policy Studies

The Andrew Young School of Policy Studies was established at Georgia State University with the objective of promoting excellence in the design, implementation, and evaluation of public policy. In addition to two academic departments (economics and public administration), the Andrew Young School houses seven leading research centers and policy programs, including the International Center for Public Policy.

The mission of the International Center for Public Policy (ICePP) at the Andrew Young School of Policy Studies is to provide academic and professional training, applied research, and technical assistance in support of sound public policy and sustainable economic growth in developing and transitional economies.

ICePP is recognized worldwide for its efforts in support of economic and public policy reforms through technical assistance and training around the world. This reputation has been built serving a diverse client base, including the World Bank, the U.S. Agency for International Development (USAID), the United Nations Development Programme (UNDP), finance ministries, government organizations, legislative bodies, and private sector institutions.

The success of ICePP reflects the breadth and depth of its in-house technical expertise. The Andrew Young School's faculty are leading experts in economics and public policy and have authored books, published in major academic and technical journals, and have extensive experience in designing and implementing technical assistance and training programs. Andrew Young School faculty have been active in policy reform in over 40 countries around the world. Our technical assistance strategy is not merely to provide technical prescriptions for policy reform, but to engage in a collaborative effort with host governments and donor agencies to identify and analyze the issues at hand, arrive at policy solutions, and implement reforms.

ICePP specializes in four broad policy areas:

- Fiscal policy (e.g., tax reforms, public expenditure reviews)
- Fiscal decentralization (e.g., reform, intergovernmental transfer systems, urban finance)
- Budgeting and fiscal management (e.g., local, performance-based, capital, and multiyear budgeting)
- Economic analysis and revenue forecasting (e.g., micro-simulation, time series forecasting)

For more information about our technical assistance activities and training programs, please visit our website at http://icepp.gsu.edu or contact us by email at paulbenson@gsu.edu.

Evidence on Economies of Scale in Local Public Service Provision: A Meta-Analysis*

Juan Luis Gómez-Reino[a], Santiago Lago-Peñas[b] and Jorge Martinez-Vazquez[c]

November 2021

Abstract

The standard theory of optimal jurisdictional size hinges on the existence of economies of scale in the provision of local public goods and services. However, despite its relevance for forced local amalgamation programs and related policies, the empirical evidence on the existence of such economies of scale remains elusive. The main goal of this paper is to produce an updated and comprehensive quantitative review of the existence of economies of scale in the provision of local public goods using a meta-analysis approach to systematize the wide range of empirical approaches and modeling frameworks found in the previous literature. Our analysis confirms the presence of moderately increasing to constant returns to scale in the provision of local services across traditional local service sectors such as education, water and sanitation, and garbage collection. We identify best practices for future empirical research in this area, which should rely on physical output as the metric of activity, production cost data as the measure of input expense, and a translog specification function for the modeling of cost functions. Finally, we find evidence that the determinants of output cost elasticity include bidirectional publication bias and population density but do *not* include the presence or absence of modern "lean" production technologies or the (perceived) capital intensity of the sector, contrary to conventional wisdom. These findings have significant policy implications for countries considering jurisdictional consolidation programs.

Keywords: Economies of scale, local public service provision; meta-analysis

JEL Classification: H11; H40; H73

* We gratefully acknowledge superb research assistance by Alejandro Dominguez, very helpful comments from Chris Thayer, and financial support by the Spanish Ministry of Science, Innovation and Universities [grant number CSO2017-85024-C2-2-P (AEI/FEDER)].

[a] Corresponding Author. Inter-American Development Bank; E-mail: jgomezreino@IADB.org
Website: https://www.researchgate.net/profile/Juan_Gomez150

[b] Governance and Economics research Network (GEN). University of Vigo (Spain); E-mail: slagop@uvigo.es Website: http://slagop.webs.uvigo.es

[c] International Center for Public Policy (Georgia State University, USA) & (GEN); E-mail: jorge.martinez@gsu.edu Website: https://aysps.gsu.edu/profile/jorge-martinez-vazquez

1. Introduction

The standard theory of optimal jurisdictional size developed by Oates (1972) and extended later on by other authors (e.g., Alesina & Spolaore, 2003) hinges on the existence of economies of scale in the provision of local public goods and services. However, the empirical evidence on the existence of such economies of scale remains elusive. Obtaining sound evidence on this issue is as relevant as ever for efficient decentralization policy design. Many countries have embarked over the years on forced jurisdictional consolidation or amalgamation programs based on the supposedly insufficient economies of scale in the delivery of local public services among their existing, and allegedly small, local governments. However, the evidence for such moves is far from conclusive, as Gendzwill et al. (2020) show after reviewing 31 studies for 14 countries implementing territorial reforms in recent times. Hence, this issue calls for a systematic and in-depth quantitative analysis that summarizes and evaluates the evidence available. The main goal of this paper is to produce an updated and comprehensive quantitative review of this important issue using a meta-analysis approach.

The rest of the paper is organized as follows: in section two, we review the definitions and different interpretations in the literature concerning economies of scale in the delivery of public services. Section three provides a summary of the stylized facts on economies of scale in the public economics literature. Section four offers a systematic quantitative review of the literature as the initial step to conducting the meta-analysis. Section five shows the results from the meta-analysis' regressions. Section six concludes.

2. On Alternative Definitions of Economies of Scale

In its classical definition, a production process is characterized by economies of scale if: "when all inputs are increased by a certain factor λ, output increases by a factor larger than that λ" (Panzar & Willig, 1977). Alternatively, economies of scale exist when we can increase the production of a good or service without increasing productions costs in the same proportion. The sources of such economies of scale are varied. They could be derived from the specialization of the production process (which may only be viable for larger levels of output); they may originate from increased bargaining power with suppliers once production increases (leading to lower or discounted prices for inputs); or they may be related to the spread of fixed costs across larger production levels (thus reducing average prices).

The most commonly used mathematical formulation of economies of scale is owed to Baumol et al. (1988): $S = \frac{C(q)}{q\frac{\partial C}{\partial q}} = \frac{1}{\frac{\partial\, ln(C)}{\partial\, ln(q)}} = \frac{1}{\varepsilon_y}$. Economies of scale (S) or increasing returns to scale (used interchangeably here forth) exist when S >1; that is, when the marginal cost of production is below the average cost. Constant or decreasing returns to scale exist when S is equal or less than unity, respectively. In elasticity terms, economies of scale exist when the cost elasticity of output is smaller than unity ($\varepsilon_y < 1$).[1]

The literature has predominantly used this definition of economies of scale, although other contributions have also merited attention. In particular, Caves et al. (1984), in their analysis of scale economies of local service airlines' costs, include a measure of network length (or points served) for the calculation of economies of scale. In their interpretation, short-term economies of scale are defined as $RTS = \frac{1}{\varepsilon_y + \varepsilon_N}$ where ε_N is the cost elasticity of the network length. As in Baumol et al, returns to scale exist when RTS >1. In addition, Caves et al. (1984) argue that the estimation of long-term economies of scale needs to take into account the quasi-fixed production

[1] The definition implies that the cost elasticity of output cannot be zero, as that would lead to infinite economies of scale.

inputs (Z) and thus $RTS = \frac{1-\varepsilon_Z}{\varepsilon_y+\varepsilon_N}$, with ε_Z as the cost elasticity of quasi-fixed production inputs.[2]

The empirical literature on the existence of economies of scale in the production of public services has concentrated on the estimation of the cost elasticity of output, using a variety of modeling frameworks. However, important contributions to the literature have adopted the interpretation of economies of scale proposed by Caves et al. (1984), such as in Mizutani and Urakami (2001), Aubert and Reynaud (2005), or Filippini and Prioni (2003).

3. Stylized Facts in the Literature

An initial review of the literature unveils a series of stylized facts which help shape our quantitative analysis below.[3]

3.1 Capital vs. labor-intensive services

From production theory, it would be reasonable to assume that economies of scale are more likely to be found in capital-intensive goods or services, where the investment in capital goods (i.e., fixed costs) can be spread across more units of output (Dollery & Fleming, 2006; Bel, 2013). As we see below, this conjecture is only partially fulfilled.

In Chile, Albala-Bertrand and Mamatzakis (2004) find economies of scale in the provision of transport, sewerage, and power grid services. Bel (2005) and Alvarez et al. (2003) show (for Spain) that solid waste collection and processing offers important savings in production costs derived from larger client populations. This is a finding shared both by Callan and Thomas (2001) in their study of 110 municipalities in the Massachusetts area and by McDavid (2000), who studies cost patterns for 327 local governments of less than 1000 citizens

[2] By quasi-fixed production inputs, the authors refer to the fact that, although in the long-run all inputs are traditionally assumed to be variable, some of them, including capital and labor for instance, can be partially adjustable.

[3] This section benefits from earlier reviews including Boyne (1995) and Andrews et al. (2002) in the area of education, Byrnes and Dollery (2002) for Australian local governments, and Bel (2009) for selected sectors.

in Canada. Conversely, Bel and Mur (2009) do not find scale economies on solid waste services in small rural municipalities in the region of Aragon (Spain), insofar as most of them rely upon intermunicipal cooperation or outsourcing. This conclusion is reinforced by the empirical analysis of Hortas-Rico and Salinas (2014). Supra-municipal aggregation of services would fade scale economies.

Concerning the water sector, Cunha Marques and De Witte (2011) found significant economies of scale, with an optimal scale of the utilities located between 160,000 and 180,000 inhabitants, well over the average Portuguese municipality (their study population). These figures are close to those estimated by Turley et al. (2018) for Ireland, where economies of scale are found to exist up to 140,000 inhabitants. For Spain, Prieto et al. (2015) also find significant economies of scale for water supply, sewerage, and water cleansing; this effect would be reinforced by population density.

In the area of urban transport, seemingly contradictory results are found depending on the sample used for the analysis. For example, Berechman (1983) finds economies of scale in the operation of buses in Israel but constant returns to scale are found by Matas and Raymond (1998) for Spain and by Filippini and Prioni (2003) for Switzerland. However, moderate increasing returns to scale are found by Farsi et al. (2007) for Swiss urban transport. More recently, Avenali et al. (2016), using data for Italy, show the existence of weak economies of scale and only for small size services. There may be additional factors specific to transportation modes and geographic conditions that influence transportation scale economies and that do not have the same, or as strong an, effect on other sectors' economies of scale.

More conclusive is the evidence obtained in the area of garbage collection, where solid evidence of economies of scale is generally found (Bel, 2009). However, in his survey of the literature, Bel (2013) also concludes that this effect is diluted when jurisdiction population is over a threshold of between 20,000 and 50,000 inhabitants. Hence, increasing returns would be stronger in countries where the average size of municipalities is smaller.

A corollary of the above proposition, based on production and cost theory, is that labor-intensive local services should offer less potential for economies of scale. A pioneering reference from this perspective is the work of Hirsch (1959) for police services in U.S. municipalities, who found no evidence of economies of scale. Analogously, Ahlbrandt (1973) examined 44 cities and districts of Seattle's metropolitan area and found no evidence of economies of scale in the provision of firefighting services. Similar conclusions are reached by Alt (1971) and Boaden (1971) and Danzinger (1978) in the cases of England and Wales. Furthermore, Ostrom and Parks (1973), Dilorenzo (1981), and Gyimah-Brempong (1987) find evidence of higher production costs for firefighting and police services with the greater population size of the jurisdiction in the United States. However, this early literature was not totally void of positive evidence on economies of scale. For example, Bodkin and Conklin (1971) report declining average costs of production with higher population size for police and fire services in local governments in the United States. The study, however, is substantially old and there has been considerable technological change in the provision of policing that could substantially change the scale economies estimate.

In the case of the provision of public schooling, evidence of economies of scale is found by Chambers (1978), Butler and Monk (1985), Callan and Santerre (1990), Duncombe et al. (1995), Jimenez (1986), Reschovsky and Imazeki (1997, 1999), and Andrews (2013). On the other hand, Gyimah-Brempong and Gyapong (1991) find decreasing returns to scale in the production of education in Michigan school districts. A general conclusion of practically all these studies on education services has been that scale economies vanish at relatively low levels of student enrollment. In this same vein, Duncombe et al. (1995) show that the consolidation of school districts in the State of New York may have offered savings in education costs, although the gains were limited to the consolidation of districts with fewer than 500 students. A similar study for Iowa by Edelman & Knudsen (1990) concluded that the gains in terms of economies of scale were found for student populations between 800 and 900. For the state of Maine, Deller and Rudnicki (1992) estimated the optimal size of the education district to be at around 2,000 students. These studies, however, predate widespread use of internet and mobile technologies, which could impact estimates.

Concerning other labor-intensive public services, Hortas-Rico and Salinas (2014) do not find economies of scale in social services in Spanish municipalities. In the case of security (municipal police), scant savings in cost are shown and only up until 500 inhabitants. Moreover, they only find evidence of significant economies of scale in the case of general administration for up to 20,000 inhabitants.

3.2 Measurement, measurement, measurement

The mixed evidence on economies of scale gathered from the empirical literature may well be due, at least partially, to critical differences in the measures of output and production costs used in the different analyses. In their review of previous works on the existence of economies of scale in Australian local governments, Byrnes and Dollery (2002) conclude that, even when homogeneous goods are analyzed, the evidence as to whether economies of scale exist in their production is inconclusive. They argue that inaccurate measures of output/production and costs are partly to blame for the variety of results found in the literature.

This general critique to the body of empirical contributions on the existence of economies of scale in public service delivery is in fact an old one. Tiebout (1960) criticized Hirsch's (1959) seminal contributions to the literature for his use of population as a proxy for public service output. Tiebout (1960; p. 444) argued that "there is no necessary relationship between population and either the output or quality of the good." In fact, larger population may lead, Tiebout argued, to larger per capita expenditures, implying no economies of scale. Studies using population as a proxy for output levels are rare nowadays, although they represented a substantial share of early works in this empirical area.

The use of expenditure data as a substitute or instead of cost data for the estimation of cost functions has also been criticized for obvious reasons. Changes in per capita expenditures in a public service may be due to reasons other than production costs; including administrative inefficiencies (Tiebout, 1960; Breton, 1965; Duncombe & Yinger, 2007). As cost data have been made increasingly available, empirical works have favored their use and the number of academic contributions using per capita expenditure as a proxy for average cost has declined over time.

There is also substantial evidence in the literature that the size of economies of scale is largely affected by the measure of output selected, even when the measures refer to the same service. In their study of cost of bus services provision in Switzerland, Filippini and Prioni (2003) find larger economies of scale when the output measure is the number of bus stops as opposed to bus-kilometers. This finding corroborates the results from Berechman and Giuliano (1984), who document diseconomies of scale in bus operation in the United States if bus-miles are used as the output measure, and instead economies of scale were found if revenue per passenger is used. Unfortunately, in most cases, there may not be a clear superior choice of output measure. Therefore, we need to be aware that evidence on the existence of economies of scale in the delivery of public services may be dependent on the output measure selected for the analysis.

Finally, Tran et al. (2018) outline the relevance of controlling for population density when analyzing the extent of scale economies attached to involved population size. Given the strong correlation between both variables, omitting the first one upwardly biases the effect of the second. Using data for 68 Australian municipalities, they show that local government expenditure is characterized by significant economies of scale. However, once municipalities are stratified by population density, the evidence for scale economies largely disappears. In contrast, evidence for Brazilian municipalities reported by Bernardelli et al. (2020) shows that economies of scale remain when the analysis controls for the effect of population density.

3.3 A U-shaped average cost function

A third salient finding in the literature is the concentration of economies of scale in smaller (population-wise) jurisdictions in studies using jurisdictions (as opposed to production units) as the focus of their analysis. This would appear to signal the expected U-shape of the long-term average cost curve for public service production. The seminal Hirsch (1959) study reports evidence of economies of scale in firefighting services in municipalities of less than 100,000 people and increasing average costs over that population size. In the same vein, Bodkin & Conklin (1971) find evidence of declining average costs in police and firefighting services for localities of between 5,000 and 10,000 people. Additionally, Gyimah-Brempong (1987), in his analysis of the Florida case, estimated that diseconomies of scale in the provision of police services start at population levels of around 50,000 residents. Sole-Olle & Bosch (2005) find substantial economies of scale for provision of local government services in Spanish municipalities with a population below 5,000 citizens but growing unit costs of provision until the population is over or about 50,000. Using more sophisticated spatial econometric tools, the recent contribution by Hortas-Rico and Rios (2019) corroborate the existence of a U-shape curve for Spanish municipalities over the period 2003-2011. Total current spending decreases with population and increases with the square of population, suggesting optimal size would be slightly over 10,000 inhabitants. Again, for Spain, the recent contribution by Piñero et al. (2021) shows a U-shape with the minimum attained for population sizes in the bracket 5,000-10,000. However, this result is not based on econometric estimates, but on careful descriptive analysis of data.

Using a sample of Catalonian (Spanish) municipalities, Bel & Fagueda (2009) show evidence of substantial economies of scale in solid waste collection (attained by inter-municipal cooperation in the provision of this service) for municipalities below 20,000 citizens, but no gains in unit costs for municipalities over that population. In Sweden, Nelson (1992) shows that savings in the production of local services derived from municipal consolidation seem limited to municipalities of very small population size (below 2,000 citizens after the consolidation). In light of these and other contributions, Bish (2001; p.14) concludes that approximately 80 percent of local government activities do not possess economies of scale beyond relatively small municipalities with populations of 10,000 to 20,000.

The review of the previous empirical literature also shows different within-country results in terms of whether economies of scale are present, depending on the sample of jurisdictions and databases used. This result is more apparent in a third group of studies which analyze overall expenditure patterns before and after processes of jurisdictional consolidation or inter-municipal cooperation. This may be a result of aggregating all local expenditures on public goods and services. The evidence on economies of scale would become even more inconclusive after aggregating both capital- and labor-intensive services, with different potential in the reduction of average production costs due to larger volume. For example, in a series of studies in the early 1970s, Davies et al. (1971) and Davies and McMillan (1972) report increasing costs of provision (measured as total local public expenditure, excluding social services) with higher population size in the U.K., a result also partially confirmed by Mehay (1981). Byrnes and Dollery (2002) review a good number of studies with similar but also contradictory findings following the local government consolidation process that took place in Australia during the 1990s. More recently, Miyazaki (2018) analyzes the relationship between municipal consolidation in Japan, discovering that per capita current expenditure does increase after consolidation, but subsequently they gradually fall. The author suggest that this result could be explained by factors such as internalization of spillovers in local public good provision, increased heterogeneity of preferences in consolidated municipalities, and/or less competitiveness among municipalities. Blom-Hansen et al. (2016) also conclude with a pessimistic view on the net gains of amalgamations. Using a differences-in-differences approach on mergers in Denmark, they conclude that cost savings in some areas are offset by deterioration in others, while for most public services, jurisdiction size did not matter at all. Moreover, the analysis by Blesse and Baskaran (2016) on Germany add two suggestive results. First, reductions in administrative expenditures are limited to compulsory (not voluntary) mergers. Second, reductions in costs are more relevant in the case of mergers with a larger number of participants and when there is a dominant partner in the merger (annexations).

3.4 Modeling frameworks for the cost function

The empirical literature on economies of scale in local public service delivery has seen three somewhat overlapping but otherwise well-differentiated stages in the modeling of the cost functions of the production process. Early works in this area used a linear function, quadratic in the measure of output, to establish the existence of U-shaped cost curves (Hirsch 1959, 1965; Bodkin & Conklin, 1971; Knapp, 1982; Kumar, 1983, among others). The reporting standards of the early contributions are arguably weaker than those of more recent works. For example, many of the articles reviewed from this early stage do not provide descriptive statistics of the variables used, naturally making the calculation of the elasticities difficult. The sample of observations for the meta-analysis below suffers therefore from a bias towards more recent articles.

A second wave of contributions to the analysis of economies of scale in local service delivery incorporates logarithmic cost functions that allow the direct estimation of the cost elasticity of production. This sub-sample of works assumes a Cobb-Douglas production function, a modeling framework that still incorporates important limitations in the analysis of economies of scale, such as the assumed constant elasticity of substitution of factors of production and returns to scale and the homotheticity of the production function (Gyimah-Brempong, 1987). The largest number of available works has used this modeling framework, including the seminal contributions from Stevens (1978) in the sector of refuse collection, Duncombe et al. (1995) in education, and Christoffersen et al. (2007) in the cleaning of schools in Denmark.

The third and more recent wave of empirical works in this area have heavily favored the use of the translogarithmic cost functions as a modeling framework. In contrast to the Cobb-Douglas function, the multi-product translog cost function places fewer restrictions on the parameters (e.g., does not assume constant elasticity of substitution of factors of production), and allows for the analysis of multi-product production processes that are common in certain sectors (e.g., primary and secondary education for instance). An early contribution in the area of bus transport that uses the translog modeling framework is that of Berechman (1983), but the framework has been applied to every possible service including Drake and Simper (2002) in the area of police, Prieto et al. (2015) in sewerage, paving and lighting, Fabbri and Fraquelli (2000) and Prieto et al. (2015) in water supply, and Jimenez (1986) in education. The more sophisticated modeling framework offered by the translog function signals the potentially more accurate and solid estimates of economies of scale, so any quantitative analysis of the literature must control for this important development. However, more recent contributions explore the possibility of relying upon even more flexible functional forms. For instance, Bikker and Van der Linde (2016) compare results for administrative costs using the translog cost function with up to three alternatives to analyze local public administration expenditure in Netherlands.

In conclusion, evidence of economies of scale for local public good provision that would justify Oates' (1972) theoretical argument for larger governmental units can be found but needs to be adequately contextualized. To the significant empirical limitations related to the measurement of output and cost of service production, we must add the complications generated by different (and coexisting) technologies and the specificities of geographical areas. Our initial review of the literature leads us to conclude that economies of scale, when found, are sector specific, population bound, and perhaps even temporary in their range and size, depending on the available technologies of production of the particular time period. Economies of scale arising from larger size jurisdictions may only exist for a small number of locally provided services. We next turn to our attempt to produce more precise information on these issues.

4. Meta-analysis: A Systematic Quantitative Literature Review

Following Stanley (2001) our meta-analysis is structured in four stages.

4.1 Identifying all relevant studies and choosing a common metric

The first stage is to identify as complete a sample of studies in the area of interest as possible. We reviewed 115 empirical studies in total from a variety of sources and journals on economies of scale in local public service delivery, including several PhD dissertations and unpublished papers.

The scope of our meta-analysis includes all published or unpublished empirical papers in the area, published in English or Spanish. Our search included two standard databases (EconLit and Dissertation abstracts) and the Google Scholar, Google, and Bing standard search engines, without any time-period restriction. We also contacted several authors for unpublished papers, dissertations, and government reports, but only with mixed success. In addition, the bibliographies of all papers reviewed were scanned for additional studies, applying a "snowball" approach to study identification.

Having obtained 115 papers overall on the topic, the selection of pieces to be included in the meta-analysis used the following criteria: First, we selected only papers that referred to local government-provided services. There exist substantial contributions on the economies of scale of some public services that are not local in nature, such as power generation or regional transportation services. These non-local studies were excluded from the analysis. Second, we eliminated from our pool of papers those which did not use regression analysis as the main estimation methodology for economies of scale. This may have biased the sample towards more recent contributions, which are more prone to use regressions as opposed to other methodologies, such as simple correlation coefficients.

Third, the review of the literature unveiled a series of studies using the production function approach to the analysis of local public service provision. None of these studies can be incorporated into the dataset as they do not truly test for economies of scale but rather for the impact of financing levels on critical performance indicators. In these studies, the left-hand side variable is traditionally a measure of service quality (i.e., average value of standardized tests in education, etc.), not the production costs. Thus, they offer complementary evidence which cannot be incorporated into our quantitative review.

Fourth, the selection of our variable of interest introduced additional limitations to the papers that could be used in the analysis. The proposition we want to test in this meta-analysis is that *economies of scale exist in the production of local public services*. Note that we bundle all local services. We do so because, with very few exceptions, local governments are expected to take on the delivery of a standard of all local services and not just a selection of them. Asymmetric assignments of functional responsibilities across local governments are rare in the international experience. Thus, for example, in most countries jurisdictional consolidation is sought for and requires the delivery of *all* services to a larger population, as opposed to a small or selective set of public services.[4]

Given the theoretical framework discussed and the proposition we want to test, the natural variable of interest for our meta-analysis is the cost elasticity of output. This statistic best summarizes the empirical results in the previous literature.[5] The use of the cost elasticity of output as our measure of economies of scale immediately imposes additional restrictions on our sample of studies. Papers where the statistic is not reported or from which it cannot be calculated were discarded. As discussed, several of the early contributions in this area use a linear (and quadratic) cost functional form. When descriptive statistics are provided, we can calculate the attached elasticity and thus the paper is added to the dataset. In many cases, however, such information was not available and the paper was discarded.

The selection process outlined above left us eventually with 56 studies that reported 76 values of the cost elasticity of output for different services. Of those 76 observations, 55 reported their attached standard errors and 21 did not. The availability of standard errors is crucial as the meta-analysis essentially weights the observations by their variance. In their absence, other

[4] If special districts exist for the delivery of separate services, as is the case in the U.S. with school districts, etc., then it can make sense to conduct the study for unbundled services. In this case, the number of observations available to conduct reliable statistical analysis can become the challenge.
[5] Earlier contributions in the literature used the total cost of production (in monetary terms) as the dependent variable. We rejected this option as it did not offer a realistic or theoretically sound alternative.

measures of study size can be used, such as the inverse of the degrees of freedom of the study, but generally they are less satisfactory. Our empirical work will include estimates of the average "true" value of economies of scale for those observations for which standard errors are reported and for the whole sample of observations (76) using a different measure of study size.

The data set includes studies from 1978 to 2020. As discussed already, several studies offer more than one observation, and we include all of them in the dataset. Observations for a total of 8 services and 22 countries are included. As is traditionally the case with cross sections, the data set of studies suffers from unobserved heterogeneity, as not all relevant moderators or control variables may have been coded.

The inclusion of studies that used translog cost functions also required calculating the individual statistic where the estimate of economies of scale, as opposed to the cost elasticity of output, was reported. When the study estimated both the Cobb-Douglas and translog functional forms of the cost function, the translog estimate of cost elasticity of output was selected for consistency, unless the Cobb-Douglas estimate was the preferred estimate of the author.

The alternative use of the Caves et al. (1984) measure of economies of scale and the Baumol et al. (1988) measure introduces a certain level of heterogeneity in the value of the dependent variable. The former, as we discussed earlier, includes in the estimation of economies of scale a measure of network length. As it was not possible to completely homogenize the statistics from the different studies, our meta-regression accordingly controls for this fact with a variable (*baumol*) which takes a value of one for the studies using the Baumol et al. (1988) measure of economies of scale and a value of zero otherwise.

Sample dependency in meta-analysis is a common risk that can manifest itself in a variety of forms. First, it may be that many observations are obtained from the same study (and thus the same sample). This may include observations obtained from different estimation methods over the same sample or the use of the same data sample by many different researchers. Fortunately, all of the studies included in our dataset use different samples. This limits sample dependency to the greatest possible extent, making our observations virtually fully independent. In our dataset, we also include the author's preferred model specification or estimation in the cases where more than one regression is run on the same sample. However, it is the case that several studies present estimations over different samples. In some cases, the samples are independent from each other, and their inclusion as separate observations does not present further problems. In other cases, several estimations are obtained from different sub-sets of the same sample. We also include them as separate observations, but control in our meta-regression with a dummy variable for studies from which we obtain more than one data point.

A different type of dependency is that caused by errors in the specifications of the econometric model that are reproduced in other studies. In our sample, this would include, for instance, the need to control for the possibility of joint public and private provision of the service being analyzed, or the possibility of multi-product functions. We define moderators in our right-hand side of the equation to control for such occurrences.

4.2 Meta-regression

As discussed, the studies analyzed differ in many critical dimensions, including the functional form of the cost equation, the estimation method, and even the measure of production. The presence of this significant "between-study heterogeneity" requires the application of meta-regression techniques that allow for random-effects estimation rather than the fixed-effect meta-analysis commonly employed for highly homogenous sets of studies. With the meta-regression we can investigate "the extent to which statistical heterogeneity between results of multiple studies can be related to one or more characteristics of the studies" (Hardbord & Higgings, 2008,

p. 493).

In a random-effects meta-regression, the individual study estimates of the variable of interest y are assumed to be distributed normally around a mean effect θ and with a between-study variance τ^2 and a standard error of each study denoted as σ_i. More specifically, we assume, following Harbord and Higgings (2008) that: $y_i|\theta_i \sim N(\theta_i, \sigma_i^2)$where $\theta_i \sim N\left(\theta, \tau^2\right)$: so $y_i \sim N(\theta, \sigma_i^2 + \tau^2)$. Or equivalently: $y_i = \theta + \mu_i + \varepsilon_i$; where $\mu_i \sim N(0, \tau^2)$ and $\varepsilon_i \sim N(0, \sigma_i^2)$

Our dependent variable (y) for the meta-analysis is the cost elasticity of output. Regressors include the moderators identified during the process of coding and discussed above. We perform variance-Weighted Least Squares (WLS) regression analysis to attach more weight to estimations with lower standard error and thus believed to be more accurate.[6] We aim to estimate the average (true) cost elasticity of output for local public service delivery. The algorithm used for the estimation calculates first the between study variance (τ^2) and later the β-coefficients using as weights $\frac{1}{\sigma^2 + \tau^2}$.

The unconditional (average) cost elasticity of output in our sample of 76 observations is 0.597: if service output increases by 1 percent, the cost of provision increases by 0.6 percent, signaling some substantial economies of scale. The results of the meta-regression are presented in the next section below.

An additional issue we need to address in our analysis is that of potential publication bias in our data sample. As indicated above, publication bias may be present because of the higher likelihood that a study is published if it reports statistically significant results. Thus, following Bom and Ligthart (2009) we assume that: $\hat{\theta}_i = \theta_i + g\left(se\left(\hat{\theta}_i\right)\right) + \mu_i$; where $\hat{\theta}_i$ represents the

[6] Specifically, we use the *metareg* command of Stata.

observed estimates, $\theta_i$is the population parameter and μ is the sampling error.

So, if there is publication bias, the insertion of the standard errors in the meta-regression should yield statistically significant coefficients for that variable. Previous studies have assumed publication bias is linear (Card & Krueger, 1995), whereas others (Stanley & Doucouliagos, 2007) argue the relationship between the estimate and its standard error is more likely to be non-linear and propose a quadratic approximation. In their study of the output elasticity of public capital, Bom and Ligthart (2009) present a comprehensive analysis of publication bias and we follow their methodology in our analysis. Our analysis will allow for identifying the direction of the bias and select the appropriate control for our meta-regression including all relevant moderators.

Again following Bom and Ligthart (2009), we estimate the equation:

$$\hat{\theta}_i = \theta_0 + \sum_{j=1}^{N} \beta_j x_{ij} + \alpha_p se(\hat{\theta}_i)^h D_{pi} + \alpha_n se(\hat{\theta}_i)^h D_{ni} + \mu_i \qquad (1)$$

where, as discussed, $\hat{\theta}_i$ represents the observed cost elasticity estimates, θ_i is the population parameter and μ is the sampling error. The term $\sum_{j=1}^{N} \beta_j x_{ij}$ represents the moderator variables coded in our meta-analysis, whereas D_{pi} (D_{ni}) are dummy variables coded 1 if $\hat{\theta}_i$ is positive (negative) and zero otherwise. They are interacted with the standard errors of $\hat{\theta}_i$. The structure of the equation allows us to test for different versions of publication bias.

If the term $\sum_{j=1}^{N} \beta_j x_{ij}$ is eliminated, we can test for publication bias in the (assumed) absence of heterogeneity between studies. If both terms: $\alpha_p se(\hat{\theta}_i)^h D_{pi}$ and $\alpha_n se(\hat{\theta}_i)^h D_{ni}$ are included in the regression as moderators, we are able to test for bidirectional publication bias. Lastly, if we include solely the standard error as one term in the regression (that is, $D_p = D_n = h = 1$), we are able to test for unidirectional publication bias. The superscript h allows us to introduce the non-linear publication bias test. Thus, if h=1, we test for linear publication bias, but if h=2, we assume a quadratic, non-linear relation between the estimates and their standard errors.

4.3 Identifying moderator variables: The coding process and related hypotheses

During the process of coding the empirical papers identified, we uncovered a number of features that may have influenced the overall results found in the previous empirical literature. Accordingly, dummy variables were created to control for them. A complete list of variables is provided in Table 1. Not all of these potential moderators were used in the regression analysis. Those that lacked statistical significance were eventually dropped in the analysis. Here we discuss the main issues addressed.

Table 1: Moderator Variables Coded from the Literature Review

Dimension	**Variable**	**Definition**	**Obs.**
Year of Survey	1970s	Value 1 if survey year from that decade, 0 otherwise.	18
	1980s	Value 1 if survey year from that decade, 0 otherwise.	6
	1990s	Value 1 if survey year from that decade, 0 otherwise.	21
	2000s	Value 1 if survey year from that decade, 0 otherwise.	18
	2010	Value 1 if survey year from that decade, 0 otherwise.	13
Data Years	YearsData	Value 1 if more than one year, 0 otherwise.	29
Sector	Education	Value 1 if it focuses on Education, 0 otherwise.	15
	Water and Sanitation	Value 1 if it focuses on Water and Sanitation, 0 otherwise.	18
	Garbage Collection	Value 1 if it focuses on Garbage collection, 0 otherwise.	20
	Urban Transportation	Value 1 if it focuses on Urban transportation, 0 otherwise.	10
	Other Services	Value 1 if it focuses on a sector not specified before, 0 otherwise.	13
Country	USA	Value 1 if the data is from USA, 0 otherwise	28
	Spain	Value 1 if the data is from Spain, 0 otherwise	16
	UK	Value 1 if the data is from UK, 0 otherwise	4
	Other countries	Value 1 if the data is from a country not listed before, 0 otherwise.	28
Unit of Analysis	Jurisdictions	Value 1 if a jurisdictional unit is focus of analysis, 0 if a production unit (municipal firm, etc.)	45
Estimation Methodology	OLS	Value 1 if it uses Ordinary Least Squares, 0 otherwise.	35
	SUR	Value 1 if it uses Seemingly Unrelated Regression, 0 otherwise.	22
	MLE	Value 1 if it uses Maximum Likelihood Estimation, 0 otherwise.	6
	2SLS	Value 1 if it uses Two-Stages Least Squares, 0 otherwise.	4
	FE	Value 1 if it uses Fixed Effects, 0 otherwise.	4
	GLS	Value 1 if it uses Generalized Least Squares, 0 otherwise.	3
	GMM	Value 1 if it uses Generalized Method of Moments, 0 otherwise.	1
	BYS	Value 1 if it uses Spatial Bayesian, 0 otherwise.	1
Dataset Structure	Cross Section	Value 1 if the structure is a Cross Section, 0 otherwise.	48
	Panel	Value 1 if the structure is a Panel, 0 otherwise.	22
	Pooled	Value 1 if the structure is a Pooled dataset, 0 otherwise.	4

	Time series	Value 1 if the structure is a Time Series, 0 otherwise.	2
Cost function form	Log linear	Value 1 if the cost function is Log linear, 0 otherwise.	37
	Linear	Value 1 if the cost function is Linear, 0 otherwise.	6
	Translog	Value 1 if the cost function is a Translog, 0 otherwise.	31
	Quadratic	Value 1 if the cost function is Quadratic, 0 otherwise.	2
Expend. Data	Expenditure	Value 1 if expenditure data used for dependent variable, 0 if cost data	27
Output Data	Physical Output	Value 1 if a measure of physical output is used, 0 if population used as a proxy for output.	42
Dummies for Elasticity	Positive Elasticity	Value 1 is the cost elasticity observed is > 0	60
	Negative Elasticity	Value 1 if the cost elasticity observed is < 0	16
Definition of EOS	Baumol	Value 1 if Baumol et al (1988) definition of EOS, 0 if Caves et al (1984)	69
Multiple Observations	Multiple Observation	Value 1 if it has multiple observations, 0 otherwise	33
Population Density	Population Density	Value 1 if Population Density is used as control variable, 0 otherwise.	31

First, our coding process included the creation of a variable for the country in which the study took place. Almost half of the 76 observations eventually considered for the meta-regression were from U.S.-based studies (28), but the sample also included several European, Asian, and Latin American countries. The dataset also creates a moderator variable representing whether the unit of analysis for the study was a production unit (e.g., a bus public company) or a jurisdiction (including studies with municipal, district, or city focus). A certain overlap is observed, however, between this variable and the one that denotes whether cost or expenditure data were used as the dependent variable of the analysis, perhaps increasing the risk of multi-collinearity if both variables are included at the same time.

In terms of the characteristics of the dataset used in studies reviewed, we created dummies denoting cross-section, panel, or time series data sets. As Berechman and Giuliano (1984) point out, cross sectional data renders biased estimates as it assumes homogeneity of the observations (i.e. jurisdictions, public companies, etc.). However, the direction of this bias is not clear and the answer must be left to the empirical analysis. Equally, we coded the estimation method of the cost function, a variable also closely linked to the dataset structure.

Regarding the modeling framework of the cost function, we created dummy variables for studies using linear, log-linear or translog functional forms of the production cost function. As discussed, the translog modeling framework, which reduces the assumptions imposed on the behavior of the dependent variable, is expected to provide more solid estimations results; but here, again, the impact on the estimated coefficient is an empirical issue.

Another dummy variable was created to denote whether expenditure (as opposed to cost) data were (1) or were not (0) used to create the dependent variable in the analysis; the presence of this moderator is expected to bias downwards the estimates of economies of scale. In addition, we created a dummy variable denoting whether the output variable used population as a proxy (0) or used a physical measure of output (1), such as gallons of water or tons of garbage.

Other moderator variables created during the coding process included a variable denoting whether the Baumol et al. (1988) or the Caves et al. (1984) definitions of economies of scale were used in the study. It is to be expected—as already discussed above—that the estimated cost elasticity of output would be lower when the latter measure is used in the analysis. In addition, we created dummies for whether the study controlled for service production alternatives (i.e., private, public, volunteer services), for cases where the analysis was disaggregated by population groups, for cases where more than one observation was obtained from the same study, and for whether the study controlled for population density. Finally, our model included the elasticity's standard errors, the degrees of freedom for each study, the number of years for which data were available in each study, and the total number of observations as size variables and controls for the robustness of results.

The model specification also included moderators that should allow us to control for important econometric considerations. We coded papers by the year of their publication and the year of the survey. These are potentially important moderators that may absorb variations in the values of our dependent variable due to changes in productive technology. In some sectors, technological advances have been offering greater flexibility of production (i.e., possibilities for diversification with relatively lower levels of production) with lower relative fixed capital investment requirements, which may have somewhat reduced the potential for economies of scale if total costs of production are considered. From that point of view, earlier analysis may show greater potential for economies of scale than later ones.

Another important control variable that was coded is the type of public service that was the focus of the study. As discussed in our theoretical framework, we would expect to find greater economies of scale in more capital-intensive services such as urban transportation or water supply and sanitation, where spreading fixed costs among larger clienteles could lead to lower average costs.

Finally, our sample of studies used a great diversity of data sets, which minimizes the presence of data dependency or sample overlap.[7]

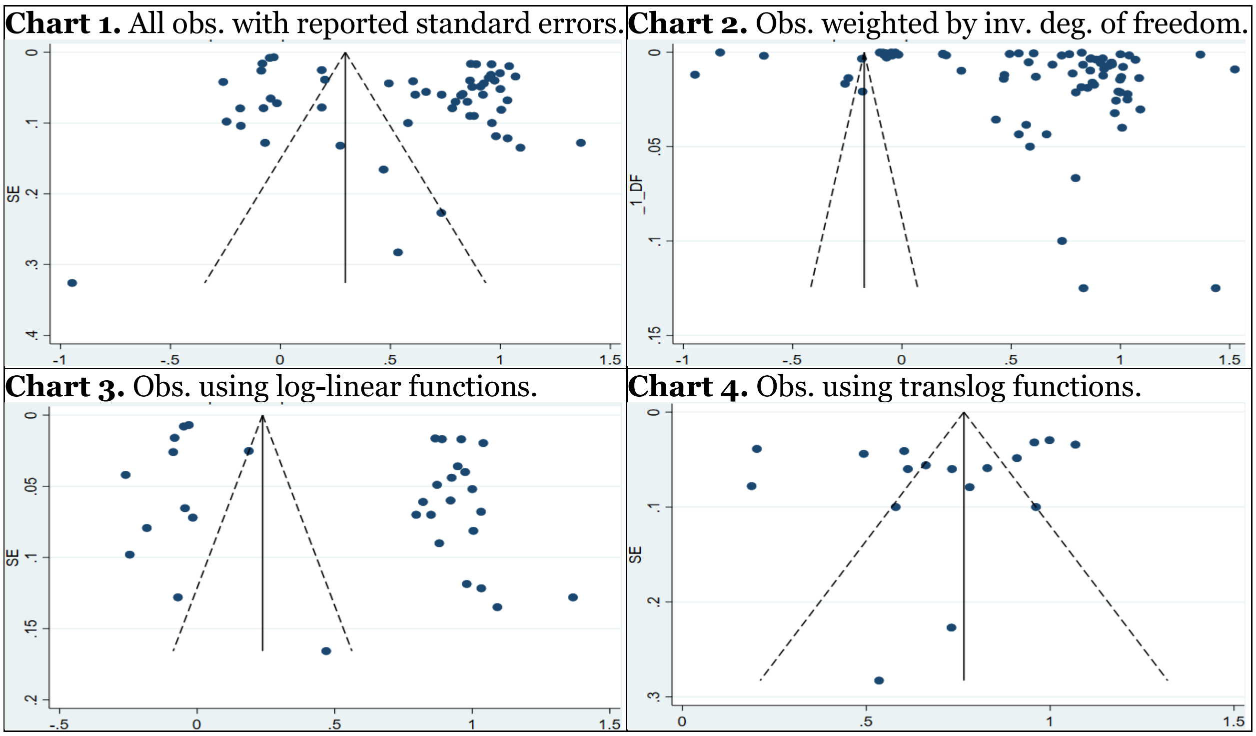
Funnel plots are a visual tool for dealing with publication and other bias in meta-analysis (Sterne & Harbord, 2004; p. 127). Publication bias exists when the probability of a study being published is higher if it reports statistically significant results (Bom & Ligthart, 2009). The term funnel plot is drawn from the "inverted funnel" shape that the scatter plot of the variable of interest and the measure of study size take in the absence of publication bias.

In our meta-analysis, each point in the funnel plots presented below depicts a particular study's value of the cost elasticity of production in the horizontal axis, and its standard error (or inverted degrees of freedom when so stated) as the measure of study size in the vertical axis. If

[7] For example, an area where there is potential for dataset dependency is that of comparative fiscal decentralization studies, commonly using the International Monetary Fund's Global Financial Statistics.

the sample of 76 observations (from 56 studies) considered in this meta-analysis were not to display publication bias, we should expect the data points representing the studies of smaller size (larger standard error) to scatter widely at the bottom of the funnel, whereas those studies with the smaller standard error (or lower value of the inverse degrees of freedom) would concentrate at the top around the “true” effect value.

Figure 1. Cost Elasticity Funnel Plot Analysis, Pseudo 95% Confidence Limits



As shown in Table 2, the sample unconditional mean is 0.597, which indicates the existence of (some) economies of scale, with a minimum value of the cost elasticity of output of -0.947 and a maximum of 1.524. Thus, the observations in the data set range from showing large economies of scale to sizable decreasing returns to scale.

Table 2: Descriptive Statistics

	Observations	Mean	Std. Dev.	Min	Max
All	76	0.597	0.515	-0.947	1.524
Education	15	0.311	0.558	-0.632	1.524
Garbage Collection	20	0.933	0.198	0.272	1.366
Water and Sanitation	18	0.716	0.314	-0.245	1.086

Chart 1 in Figure 1 displays the first funnel plot, where all 55 observations for which we have standard errors reported are plotted. It includes observations for most of the public services considered and for all types of functional forms of the cost function being reviewed. The plug-in routine for Stata-generated funnel plots calculates the fixed effects meta-estimate, which determines the position of the solid vertical line of the chart. The dependent variable is the cost elasticity of output and the independent variable the standard error of the coefficient. This is a weighted average where the weights represent the inverse variance of the estimate. The discontinuous lines that form the inverted "funnel" represent the 95% confidence limits around the summary treatment effects. It is important to note that the fixed effects estimate obtained from the funnel plots does not include any of the moderator variables that will be used later on in the meta-regression.

The funnel plot on Chart 1 presents a twin-peak structure that is relatively uncommon in meta-analysis. A first group of studies with relatively low standard errors concentrate around a value of 1 for the cost elasticity of output, signaling from limited returns to scale or slight diseconomies of scale. A second peak is found around the value 0 of cost elasticity of output, signaling relatively large economies of scale with similarly small standard errors. This latter group of studies is considerably smaller in number though. Only 19 observations out of the 55 for which standard errors are available in our sample reported a cost elasticity of output below 0.5. In this group, 8 of those observations corresponded to studies in the education sector, and 12 of them used a log-linear function to model the cost function.

The remaining 36 observations in the sample with standard errors reported included 18 observations on the garbage services sector, 8 in the water supply services and sanitation services and 4 in urban transport services. The studies' unit of analysis is mostly jurisdictional units (23), 24 of them use cross sectional data, 14 assume a translog cost functional form, and 13 of them took place in the USA.

The inclusion in the funnel plot analysis of those observations that do not report standard errors does not significantly change the results (see Chart 2). In this chart, the measure of size used is the inverse of the degrees of freedom, a common alternative to the individual standard errors. The pattern is somewhat less clear, although the two-peak structure is also identifiable around values of the cost elasticity coefficient of 1 and 0. Neither funnel plot (Charts 1 and 2) presents the symmetrical distribution that would signal absence of publication bias. The large heterogeneity among the studies and services analyzed prevents this. The first two funnel plots, in addition, do not allow us to establish the direction of the publication bias, and thus additional quantitative analysis will be undertaken to test for it in the next section.

We can, however, look more closely at the drivers of the "twin peak" distribution obtained from the general funnel plots. As discussed, it would seem to be partially determined by the distribution of studies using a log-linear function as the modeling framework for the estimation of cost elasticities. Chart 3 shows the funnel plot obtained from the representation of just such studies. The two peaks around the 1 and 0 values of the cost elasticity of output dependent variable are clearly identifiable.

This compares with a completely different distribution of studies which use the translog function as the modeling framework for the estimation of production costs. In Chart 4 we can observe that those studies report, in general, very low standard errors and although the funnel shape that would indicate absence of publication bias is also absent, values of the cost elasticity coefficient bunch in the interval 0.5 to 1.

In terms of the sectoral distribution of observations, results from studies on education services are leading the overall distribution of observations towards the two-peaked structure observed. Figure 2 below depicts the funnel plots for the four sectors that contain the largest numbers of observations, namely education, garbage collection, water supply and sanitation, and urban transportation services. These four sectors account for 59 out of the 76 observations in the sample. As we mentioned, a total of six observations from studies on education reported cost elasticity coefficients lower than 0.5 (Chart 5), while only two observations for water supply and one for garbage collection report those low elasticities.

Figure 2. Cost Elasticity Funnel Plot Analysis, Pseudo 95% Confidence Limits, Sectoral Distribution

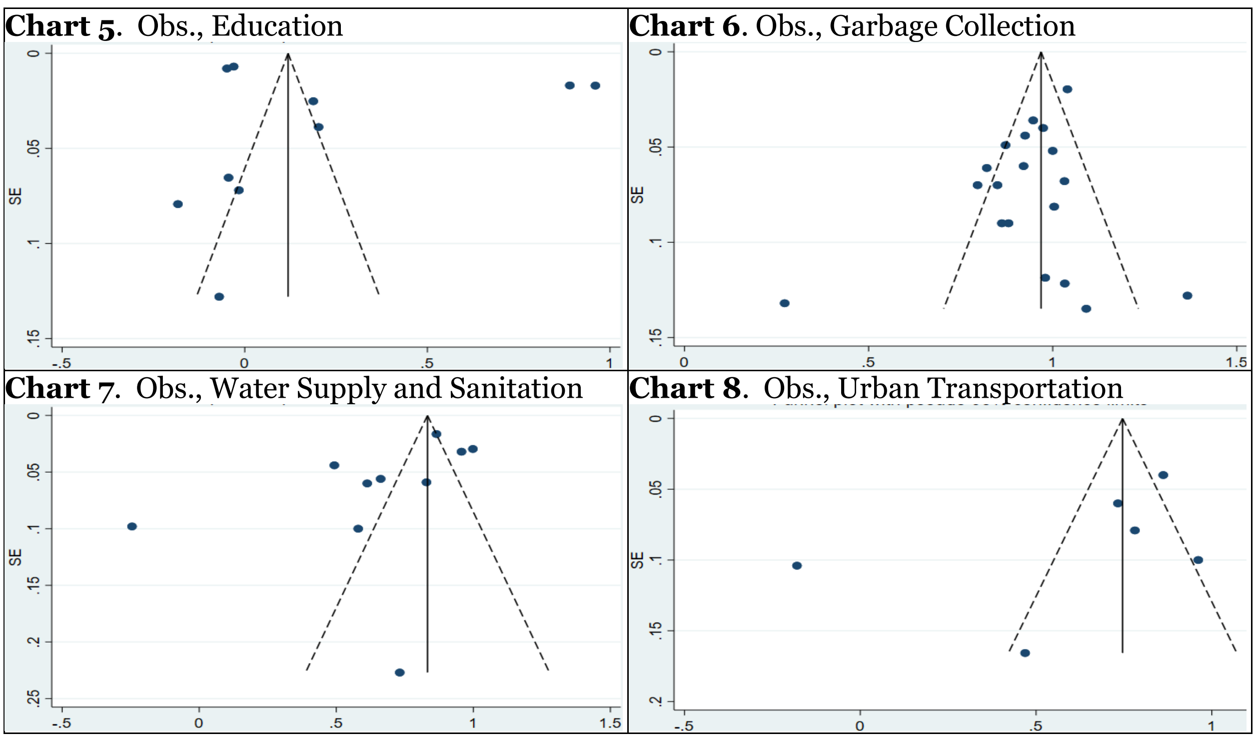


Chart 6 in Figure 2, depicting the funnel plot for observations on the garbage collection sector, offers the single-peaked, well-behaved funnel plot distribution that presumes absence of publication bias. Most of the observations are within the 95% confidence interval defined by the funnel, and the "true" value of the cost elasticity of output in this sector seems to be defined at around 0.9. The plots for the water supply and urban transportation sector show also a relatively standard distribution (meaning single peaked, with most of the observations contained within the 95% confidence interval of the inverted funnel) of observations.

Thus, we can conclude that the double peaked plot obtained in Chart 1 is driven by the observations from studies in the education sector which yield values of the cost elasticity of output that are close to 0, signaling sizable economies of scale. For those observations, a log-linear form for the estimation of the cost function was primarily used. We anticipate this may have important implications for our meta-regression analysis, to which we turn next.

5. Meta-regression Analysis

5.1 Testing for publication bias

We test first for publication bias under the assumption of homogeneity among studies in specification (1). Thus, we do not include any of the moderator variables identified during the coding process in our meta-regression. If the only differences between studies are due to sampling error ("within-study" heterogeneity), then the fixed effects estimation would be the adequate estimation methodology. If, however, large "between-study" heterogeneity is expected, then we should consider the use of random effects estimation to account for both sources of heterogeneity. Main results are reported in Tables 3 and 4.

Table 3: Publication Bias Analysis, Fixed Effects

	Output Cost Elast.	α	α_p	α_n	R2	N	Q-test	I^2
No PB	0.295** (0.141)				0.293	55	14699.26***	99.60%
Linear PB. Unidirectional	0.270* (0.146)	31.843* (18.11)			0.316	55		
Linear PB. Bidirectional	0.263* (0.143)		64.068** (24.400)	-599.918	0.369	55		
Non-linear PB. Unidirectional	0.293** (0.143)	34.203 (44.030)			0.294	55		
Non-linear PB. Bidirectional	0.294** (0.144)		244.196 (164.459)	-130.756** (30.362)	0.295	55		

Notes: Where I^2 is variation due to "between-study" heterogeneity and the Q-test is of Heterogeneity with degrees of freedom of 54. Standard errors in parenthesis. ***, **, * = statistical significance at 1%, 5% and 10%.

Table 3 shows the fixed effects estimates. Our estimate of cost elasticity of output under this assumption is somewhat smaller than the simple average for all estimates. The literature may have favored the publication of studies that reported negative cost elasticity of output, signaling large economies of scale in public good provision. As we have seen, most of these studies belong to the area of education. The fixed effects estimates are however compromised by the large amount of between-study heterogeneity as indicated by the Q-test.[8] In addition, the I^2 test shows that 99.6% of the heterogeneity found in the sample is due to "between-study" differences.

Table 4. Publication Bias Analysis, Random Effects

	Output Cost Elast.	α	α_p	α_n	N	I^2 Residual
No PB	0.581*** (0.064)				55	99.49%
Linear PB. Unidirectional	0.681*** (0.105)	-1.391 (1.149)			55	99.56%
Linear PB. Bidirectional	0.591*** (0.075)		1.946** (0.948)	-6.724*** (1.131)	55	99.25%
Non-linear PB. Unidirectional	0.640*** (0.071)	-28.238691			55	99.63%
Non-linear PB. Bidirectional	0.600*** (0.064)		8.542 (20.218)	-49.473*** (15.855)	55	99.64%

Notes: Standard errors in parenthesis. ***, **, * = statistical significance at 1%, 5% and 10%.

Due to the large heterogeneity among studies (in sectors, modeling frameworks, etc.), random effects estimation is recommended. These results are shown in Table 4. These estimates confirm the nature and the direction of the publication bias, with studies reporting negative values of cost elasticity of output driving the "true" average value in our sample.

However, the estimations presented in Tables 3 and 4 explain a small amount of the between study variation, an average of 30 percent. We therefore turn to analyzing the case where observed heterogeneity between studies with the insertion of moderator variables in the meta-regression is allowed. In line with Harbord and Higgings (2008; p. 497) and our earlier results,

[8] The Q-test is a common measure of heterogeneity used in the literature. It is the sum of the squared deviations of each study's effect estimate from the overall effect estimate (Huedo-Medina et al., 2006).

we do not estimate the fixed effects meta-regression, as such estimation assumes that “all heterogeneity can be explained by the covariates,” leading to excessive type I errors in cases (as ours) of unobserved heterogeneity. In the random effects model, the weights used in the weighted least squares estimation include not only the standard errors of each individual observation, but the between study variance.

Table 5 presents the random effects estimation of the model with observed heterogeneity. The first column reports results without testing for publication bias, while the second and third columns of Table 5 include the publication bias test in its linear and non-linear form, respectively. This first set of results offer interesting insights as to the determinants of the estimations found in the empirical literature of the cost elasticity of output. The first relevant result is the value of the conditional mean of the cost elasticity of output. In principle, due to the larger amount of “between-study heterogeneity” explained under the linear publication bias test, it would seem that such estimation presents the better fit for the model. The estimated coefficient for the constant term, the conditional mean of the sample, is 0.529. This signals the presence of some economies of scale for works published in the 2010s, in sectors other than education and garbage collection that used predominantly the Cobb-Douglas form approach to the cost production function, and also used cost as opposed to expenditure data and population as a proxy for output. The non-linear publication bias test yields a conditional mean of the cost elasticity of output of 0.297, signaling economies of scale. The results, using studies published in the last decade as the reference group, also show large variations in the value of the conditional mean depending on the inclusion of the publication bias test. Once we control for bidirectional publication bias, the estimates of the cost elasticity of output increase substantially, lowering the extent of economies of scale.

Our time dummies show that, having the 2010s as reference group, estimates of economies of scale seem to have been larger in the 1980s; this is a period in which the sophistication of the analyses increases considerably with the generalization of log linear function estimation specifications and the first contributions using translog cost functions. However, this effect falls considerably later, with the estimates of the cost elasticity of output being similar in size between the 1990s and the 2010s. Despite the consistency in the sign of this relationship, the results are not statistically significant. We thus find no strong support for the hypothesis that modern production methods, incorporating "leaner" technologies and lower requirements in terms of capital investment, offer lower potential for economies of scale.

Table 5. Meta-Regression Results, Random Effects Estimation

	No PB Test	PB Linear Bidirectional Test	PB Non-Linear Bidirectional Test
Constant	0.224 (0.248)	0.529** (0.228)	0.297 (0.244)
1970s	0.662*** (0.158)	0.573*** (0.137)	0.637*** (0.153)
1980s	0.149 (0.220)	0.205 (0.188)	0.120 (0.215)
1990s	0.329* (0.175)	0.335** (0.150)	0.293* (0.170)
2000s	0.375** (0.151)	0.370*** (0.131)	0.373** (0.147)
Education	-0.453** (0.180)	-0.508*** (0.155)	-0.504*** (0.176)
Garbage Collection	0.732*** (0.156)	0.528*** (0.146)	0.708*** (0.152)
Translog Cost Function	0.539*** (0.132)	0.318** (0.127)	0.507*** (0.128)
Expenditure Data	0.235* (0.132)	0.202* (0.113)	0.272** (0.129)
Physical Output	-0.276* (0.135)	-0.289** (0.117)	-0.279 (0.131)
Baumol	-0.170 (0.130)	-0.195 (0.151)	-0.185 (0.170)
Multiple Observations	0.172 (0.130)	0.152 (0.111)	0.146 (0.126)
Population Density	-0.309*** (0.096)	-0.280*** (0.082)	-0.303*** (0.093)
α_p		-0.681 (0.835)	-7.821 (14.961)
α_n		-4.465*** (0.991)	-31.968** (12.276)
Observations	76	55	55
Residual Variation due to Heterogeneity	96.23%	97.00%	97.56%
Proportion of Between-Study Variance Explained	70.02%	79.01%	72.09%
τ^2 (Between-study Variance)	0.065	0.045	0.060
Prob > F	0.0000	0.0000	0.0000

Notes: Standard errors in parenthesis. ***, **, * = statistical significance at 1%, 5% and 10%.

Our initial hypotheses regarding the sectoral distribution of economies of scale and capital input intensity are now confronted with opposing empirical evidence. Among the critical sectors analyzed, education consistently displayed a negative and highly significant coefficient in different model specifications. Economies of scale seem to be potentially greater in the education sector despite the presumption that this service displays a more labor-intensive production method and even after other moderator variables are included in the analysis. Studies analyzing economies of scale in the garbage collection sector, an assumed capital-intensive sector, displayed higher cost elasticities of output, meaning lower economies of scale—although the results were not statistically significant. The results for the water and sanitation sector were not found statistically significant either in the alternative model specifications.

From the results in Table 5, we can also see that greater sophistication in the modeling of production costs leads to smaller estimates of economies of scale. The use of the translog function has, all other things equal, led to higher estimates of cost elasticity of output in the literature. Acknowledging that the translog functions offers substantial advantages in the study of economies of scale, we can conclude that the use of log linear (Cobb-Douglas based) functional forms for the estimation of economies of scale may have led to the overestimation of economies of scale across the board. Lastly, the empirical results show the relevance of controlling for population density in the evaluation of economies of scale, as Tran et al. (2018) have observed. Studies using population density as a moderator generally report lower estimates of cost elasticity of output, increasing the extent of economies of scale.

The bias introduced in the analysis of economies of scale using inadequate measures of output is also made obvious from the results in Table 5. The use of expenditure data, as opposed to cost data, substantially increases the estimates of cost elasticity of output, thus reducing the perceived potential of economies of scale. Also, as previously discussed, the use of expenditure data as a proxy for production costs introduces distortions in the analysis, as expenditure data includes administrative items not necessarily related to the production of services. In addition, the use of physical output instead of population as a proxy for production proves to be critically important for the results obtained. As expected, more accurate (physical) measures of output led to larger estimates of economies of scale in the literature.

Lastly, the meta-regression results in Table 5 show no impact from the use of different definitions of economies of scale (i.e., Baumol or Caves). We interpret this as a positive sign for the consistency of our sample; more specifically, the negative result offers some relief regarding the possible distortion introduced by the heterogeneity in the measurement of our dependent variable. The results in Table 5 also show that studies with multiple observations may tend to report greater estimates of cost elasticity of output, that is, smaller economies of scale, although the significance of the coefficients was not robust to different model specifications.

Several other control variables were included in earlier model specifications but were found not significant. Country dummies were consistently found to be not significant. Their introduction as moderators in some cases was even pernicious as they could display high correlation with other moderators (for instance, 10 of the 16 observations from Spain come from studies in the garbage collection sector, creating multicollinearity between the country and the sectoral dummy). Individual significance tests also recommended their elimination from the sample with no loss of explanatory value. The variables identifying the structure of the dataset used in the study (e.g., cross-sections, panel, etc.), and the estimation method (e.g., OLS, SUR, etc.) proved to be equally non-significant. As expected, high correlation was found between the variables measuring the dataset structure, the estimation methodologies, and the form of the cost function, so most had to be discarded from the final specification. We also note that the results in Table 5 were robust to the inclusion or not of these other control variables.

As we conjectured above, our general results are greatly driven by the studies in education. We saw in Table 1 that the mean cost elasticity of output in the sector of education is the lowest among the three main sectoral groups of observations. Also, in Chart 5 in Figure 2, we observed that most of the observations reporting negative or low output cost elasticity are obtained from studies in the education sector. Thus, in order to test the robustness of our results to sectoral composition, we estimate additional model specifications which alternatively exclude each of the three sectors for which the largest number of observations is obtained (i.e., education, garbage collection, and water and sanitation). The results are shown in Table 6. For brevity, only selected variables are reported.

Table 6. Sectoral Disaggregation. Random Effects Estimation

	(1) Without Education		**(2) Without Water and Sanitation**		**(3) Without Garbage**	
Constant	0.267	(0.256)	-0.380	(0.394)	-0.312	(0.349)
(1) Garbage Collection (2) Education (3) Water and Sanit.	0.793***	(0.158)	-0.859***	(0.190)	0.075	(0.173)
1970s	0.489**	(0.201)	0.989***	(0.179)	0.549***	(0.180)
1980s	0.349	(0.201)	0.709***	(0.209)	-0.212	(0.309)
1990s	0.272	(0.197)	0.827***	(0.190)	0.154	(0.224)
2000s	0.363**	(0.172)	0.857***	(0.180)	-0.083	(0.267)
Translog Cost Function	0.507***	(0.138)	0.431**	(0.180)	0.748***	(0.161)
Expenditure Data	0.092	(0.141)	0.163	(0.162)	0.324	(0.196)
Physical Output	-0.199	(0.149)	-0.430***	(0.152)	0.216	(0.172)
Baumol	-0.128	(0.172)	0.033	(0.286)	-0.041	(0.206)
Multiple Observations	-0.0004	(0.144)	0.567***	(0.156)	0.177	(0.184)
Population Density	-0.306***	(0.108)	-0.218*	(0.115)	-0.199	(0.180)
Number of Observations		61		58		56
Residual Variation due to Heterogeneity		94.16%		94.80%		95.73%
Proportion of Between Study Variance Explained		67.72%		62.84%		61.85%
τ^2 (Between-study Variance)		0.059		0.089		0.083

Notes: Standard errors in parenthesis. ***, **, * = statistical significance at 1%, 5% and 10%.

The significance and robustness of the results obtained in Table 5 do not seem to be affected by any particular sector. The alternate exclusion of education, water and sanitation, and garbage collection from the estimations changes the constant and sectoral coefficients substantially but the results for the other determinants remain significant. The use of the translog function continues to determine results and leads to lower estimates of economies of scale, as is the case with the use of expenditure over cost data. As expected, the variation is wider for the estimates of physical output, as this variable is substantially more correlated with the sectoral studies (e.g., education studies use population as their measure of output).[9]

6. Conclusions

Our review of the empirical literature shows that the evidence on the existence of economies of scale in the delivery of local public services is not very strong, with multiple studies reporting constant or decreasing returns to scale in a variety of services. In this paper, we use meta-analysis to systematize the wide range of empirical approaches and modeling frameworks found in the literature and help identify the determinants behind the results found.

At best, the inclusion of studies from several sectors in our analysis seems to confirm the presence of moderately increasing to constant returns to scale in the provision of local services. The potential for economies of scale seems to differ greatly, at least across three traditional services: education, water and sanitation, and garbage collection, being highest for education and lowest for garbage collection. Our analysis also offers guidelines for future empirical research in this area. Physical output and production cost data should be used, together with translog specifications for the modeling of cost functions.

The estimates of economies of scale selected for our meta-regression are those at the mean of the sample distribution of each study. As such, our analysis does not offer insights regarding the extent and length of those economies of scale. However, it is unlikely, in the context of U-shaped long average cost functions, that such economies of scale will be pervasive

[9] Our analysis also included the estimation of sector-specific meta-regressions for those sectors with sufficient observations (garbage and water supply and sanitation). However, as expected little variation was found among the most critical moderators within a particular sector and for space reasons those results are not shown here.

well beyond the average production point.

This general conclusion may be somewhat surprising to policymakers in the many countries where there has been a significant push for reforming the vertical structure of government by using forced jurisdictional consolidation programs. The evidence is so far generally weak that larger “client” bases allow reducing the average costs of production in the delivery of most local public services beyond certain a modest jurisdiction size, which many studies in this literature have estimated at 10,000 residents. Any program of jurisdictional consolidation needs to be anchored on an analysis of the potential economies of scale on the services that have been decentralized to those units.

For conducting such an analysis, this paper offers significant methodological insights: using production cost data and a translog specification function. In short, the expected savings from enlarging population size at the local government level may not be present at all and should not be automatically assumed. Conducting the proper analysis, we might find cases where jurisdictional consolidation is profitable or even appropriate for other causes, such as reducing administrative overhead or developing adequate administrative capacity and skills; however, experience suggests that forced jurisdictional consolidation may often fail to bring cost or scale advantages. When considering jurisdictional consolidation, it would be also desirable to keep in mind that the desirable economies of scale may also be obtained via alternate processes such as privatization or inter-jurisdictional cooperation. In any case, transaction costs due to the need for monitoring the quality of service provision or negotiation with partners and providers may be also relevant for all those alternatives (Baba & Asami, 2020).

References

Aghion, P., Alesina, A., & Trebbi, F. (2002). Endogenous political institutions. Unpublished, Harvard University.

Ahlbrandt, R. (1973). Efficiency in the provision of fire services. *Public Choice*, *16*(1), 1-15.

Albala-Bertrand, J. M., & Mamatzakis, E. C. (2004). The impact of public infrastructure on the productivity of the Chilean economy. *Review of Development Economics*, *8*(2), 266-278.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, *8*(2), 155-194.

Alesina, A., & Spolaore, E. (2003). *The size of nations*. MIT Press.

Alt, J. E. (1971). Some social and political correlates of county borough expenditures. *British Journal of Political Science*, *1*(1), 49-62.

Alvarez, X., Caride, M., & Gonzalez, X. (2003). La gestion del servicio de recogida de basura en los ayuntamientos gallegos. *Revista Galega de Economia*, *12*(2), 1-37.

Andrews, M., Duncombe, W., & Yinger, J. (2002). Revisiting economies of size in American education: Are we any closer to a consensus? *Economics of Education Review, 21*(3), 245-262.

Andrews, R. (2013). Local government size and efficiency in labor-intensive services: Evidence from local educational authorities in England. In S. Lago-Peñas & J. Martínez-Vázquez (Eds.), *The challenge of local government size*. Edward Elgar, Cheltenham (UK), 148-170.

Aubert, C., & Reynaud, A. (2005). The Impact of regulation on cost efficiency: An empirical analysis of Wisconsin Water Utilities. *Journal of Productivity Analysis*, *23*(3), 383-409.

Avenali, A., Boitani, A., Catalano, G., D´Alfonso, T., & Matteucci, G. (2016). Assessing standard costs in local public bus transport: Evidence from Italy. *Transport Policy*, *52*(2016), 164-174.

Baba, H. & Asami, J. (2020). Estimating the minimal efficient scale and the effect of intermunicipal cooperation on service provision areas for waste treatment in Japan. *Asia-Pacific Journal of Regional Science*, *4*: 139-158.

Baleiras, R. (2001). To fragment or to consolidate jurisdictions: The optimal architecture of government. *World Bank Working Paper 2010*. http://fesrvsd.fe.unl.pt/WPFEUNL/WP2001/wp401.pdf. (March 2010)

Basab Dasgpta A.N., Narayan, A. & Skoufias, E. (2009). Measuring the quality of education and health services: The use of perception data from Indonesia. *World Bank Working Paper 2010* http://econpapers.repec.org/paper/wbkwbrwps/5033.htm. (January 2010)

Baum, C. F. (2006). *An introduction to modern econometrics using Stata* (1st ed.). Stata Press.

Baumol, W. J., Panzar, J. C., & Willig, R. D. (1988). *Contestable markets and the theory of industry structure.* Harcourt College.

Bazargan, M., & Vasigh, B. (2003). Size versus efficiency: A case study of US commercial airports. *Journal of Air Transport Management, 9*(3), 187-193.

Beaton, W. P. (1974). The determinants of police protection expenditures. *National Tax Journal, 27*(2), 335-349.

Bel, G. (2005). Un analisis de los gastos municipales por el servicio de residuos solidos urbanos. *Revista de Economia Aplicada*, *13*(38), 1-28.

Bel, G. (2009). Servicios locales, infraestructura y transporte: Dimension, escala, redes e instituciones de gobernanza. Seminario "Descentralización y Desarrollo Local". CAF. Lima, Peru.

Bel, G., & Fagueda, X. (2009). Empirical analysis of solid management waste costs: Some evidence from Galicia, Spain. Research Institute of Applied Economics. *Working Papers 2009/07.*

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

Bel, G. (2013). Local government size and efficiency in capital-intensive services: What evidence is there of economies of scale, density and scope? In S. Lago-Peñas & J. Martínez-Vázquez (Eds.), *The challenge of local government size*. Edward Elgar, Cheltenham (UK), 148-170.

Berechman, J. (1983). Costs, economies of scale and factor demand in bus transport: An analysis. *Journal of Transport Economics and Policy*, 17(1), 7-24.

Berechman, J., & Giuliano, G. (1984). Analysis of the cost structure of an urban bus transit property. *Transportation Research Part B: Methodological*, *18*(4-5), 273-287.

Bernardelli, L.V., Kortt, M.A. & Dollery, B. (2020). Economies of scale and Brazilian local government expenditure: Evidence from the state of Paraná. *Local Government Studies*, *46*(3), 436-458.

Berry, F. S., & Berry, W. D. (1990). State lottery adoptions as policy innovations: An event history analysis. *The American Political Science Review*, *84*(2), 395-415.

Besley, T., & Coate, S. (2003). Centralized versus decentralized provision of local public goods: a political economy approach. *Journal of Public Economics*, *87*(12), 2611-2637.

Bikker, J., & van der Linde, D. (2016). Scale economies in local public administration, *Local Government Studies*, *42*(3), 441-463.

Bish, R. L. (2001). Local government amalgamations. *The Urban Papers*, C.D. Howe Institute, (N. 150).

Blackwell, C. (2007). A meta-analysis of tax compliance experiments. *ICePP Working Paper Series*, 07-24. http://aysps.gsu.edu/isp/files/ispwp0724.pdf. (May 2009)

Blais, A., & Young, R. (1999). Why do people vote? An experiment in rationality. *Public Choice*, *99*(1), 39-55.

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

Blom-Hansen, J., Houlberg, K., Serritzlew, S., & Treisman, D. (2016). jurisdiction size and local government policy expenditure: Assessing the effect of municipal amalgamation. *American Political Science Review*, *110*(4),

Boaden, N. (1971). *Urban policy making*. Cambridge: Cambridge University Press.

Bodkin, R. G., & Conklin, D. W. (1971). Scale and other determinants of municipal government expenditures in Ontario: A quantitative analysis. *International Economic Review, 12*(3), 465-481.

Bom, P. R. D., & Ligthart, J. E. (2009). How productive is public capital? A meta-regression analysis. *ICePP Working Paper Series*, 09-12. http://aysps.gsu.edu/isp/files/ispwp0912.pdf. (June 2009)

Bosch, N., & Suarez-Pandiello, J. (2008). Structural reform in Spain. In B. Dollery & L. Robotti (Eds.), *The theory and practice of local government reform*. Studies in Fiscal Federalism and State-local Finance. Cheltenham, U.K.: Edward Elgar Publishing Limited.

Bowler, S., & Donovan, T. (2004). Evolution in state governance structures: Unintended consequences of state tax and expenditure limitations. *Political Research Quarterly*, 57(2), 189-196.

Boyne, G. A. (1992). Local government structure and performance: Lessons from America? *Public Administration*, *70*(3), 333-357.

Boyne, G. A. (1995). Population size and economies of scale in local government. *Policy and Politics*, *23*(3), 213-222.

Brennan, G., & Buchanan, J. M. (2006). *The power to tax: analytic foundations of a fiscal constitution* (1st ed.). Cambridge University Press.

Breton, A. (1965). Scale effects in local and metropolitan government expenditures. *Land Economics*, *41*(4), 370-372.

Buchanan, J. M. (1965). An economic theory of clubs. *Economica, New Series, 32*(125), 1-14.

Burns, N. (1994). *The formation of American local governments: Private values in public institutions.* Oxford University Press, USA.

Butler, R. J., & Monk, D. H. (1985). The cost of public schooling in New York state: The role of scale and efficiency in 1978-79. *Journal of Human Resources, 20*(3), 361-382.

Byrnes J., & Dollery B. (2002). Do economies of scale exist in Australian local government? A review of the research evidence. *Urban Policy and Research, 20*, 391-414.

Callan, S. J., & Santerre, R. E. (1990). The production characteristics of local public education: A multiple product and input analysis. *Southern Economic Journal, 57*(2), 468-480.

Callan, S. J., & Thomas, J. M. (2001). Economies of scale and scope: A cost analysis of municipal solid waste services. *Land Economics*, *77*(4), 548-560.

Card, D., & Krueger, A. B. (1995). Time-series minimum-wage studies: A meta-analysis. *American Economic Review*, *85*(2), 238-243.

Caves, D. W., Christensen, L. R., & Tretheway, M. W. (1984). Economies of density versus economies of scale: Why trunk and local service airline costs differ. RAND. *Journal of Economics*, *15*(4), 471-489.

Chambers, J. G. (1978). Educational cost differentials and the allocation of state aid for elementary/secondary education. *Journal of Human Resources*, *13*(4), 459-481.

Christoffersen, H., Paldam, M., & Wurtz, A. H. (2007). Public versus private production and economies of scale. *Public Choice*, *130*(3-4), 311-328.

Cunha Marques, R., & De Witte, K. (2011). Is big better? On scale and scope economies in the Portuguese water sector. *Economic Modelling*, *28*(2011), 1009–1016.

Danzinger, J. (1978). *Making budgets*. London: Sage.

Davies, B., Barton, A., Williamson, V., & McMillan, I. (1971). *Variations in services for the aged.* London: Bell and Sons.

Davies, B., & McMillan, I. (1972). *Variations in children's services among British Urban Authorities.* London: Bell and Sons.

Deller, S. C. (1998). Local government structure, devolution, and privatization. *Review of Agricultural Economics, 20*(1), 135-154.

Deller, S. C., & Rudnicki, E. (1992). Managerial efficiency in local government: Implications on jurisdictional consolidation. *Public Choice*, *74*(2), 221-231.

Dilorenzo, T. J. (1981). The expenditure effects of restricting competition in local public services: The case of special districts. *Public Choice*, *37*(3), 569-578.

Dollery, B., & Crase, L. (2004). Is bigger local government better? An evaluation of the case for Australian municipal amalgamation programs. *Urban Policy and Research, 22*(3), 265-275.

Dollery, B., & Fleming, E. (2006). A conceptual note on scale economies, size economies and scope economies in Australian local government. *Urban Policy and Research*, *24*(2), 271-282.

Dollery, B., & Robotti, L. (2009). *The theory and practice of local government reform*. Edward Elgar Publishing.

Dowding, K., & Mergoupis, T. (2003). Fragmentation, fiscal mobility, and efficiency. *The Journal of Politics*, *65*(04), 1190-1207.

Downs, A. (1957). An economic theory of political action in a democracy. *The Journal of Political Economy*, *65*(2), 135-150.

Drake, L., & Simper, R. (2002). X-efficiency and scale economies in policing: A comparative study using the distribution free approach and DEA - PB - Routledge. *Applied Economics*, *34*(15), 1859.

Duncombe, W., Miner, J., & Ruggiero, J. (1995). Potential cost savings from school district consolidation: a case study of New York. *Economics of Education Review*, *14*(3), 265-284.

Duncombe, W., & Yinger, J. (2007). Does school district consolidation cut costs? *Education Finance and Policy*, *2*(4), 341-375.

Easterly, W., & Levine, R. (1997). Africa's growth tragedy: policies and ethnic divisions*. *Quarterly Journal of Economics*, *112*(4), 1203-1250.

Edelman, M. A., & Knudsen, J. J. (1990). A classic economies of size analysis on average school costs: An Iowa case study. *North Central Journal of Agricultural Economics*, *12*(1), 99-108.

Ellickson, B. (1971). Jurisdictional fragmentation and residential choice. *American Economic Review*, *61*(2), 334-339.

Fabbri, P., & Fraquelli, G. (2000). Costs and structure of technology in the Italian water industry. *Empirica*, *27*(1), 65-82.

Farsi, M., Fetz, A., & Filippini, M. (2007). Economies of scale and scope in local public transportation. *Journal of Transport Economics and Policy*, *41*(3), 345-361.

Feld, L. P., & Heckemeyer, J. H. (2009). *FDI and taxation: A meta-study.* http://www.cesifo.de/DocCIDL/cesifo1_wp2540.pdf.

Feld, L. P., Kirchgässner, G., & Schaltegger, C. A. (2003). Decentralized taxation and the size of government: Evidence from Swiss state and local governments. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=478261. (November 2008)

Filippini, M., & Prioni, P. (2003). The influence of ownership on the cost of bus service provision in Switzerland--An empirical illustration. *Applied Economics*, *35*(6), 683-690.

Gendzwill, A., Kurniewicz, A. & Swianiewicz, P. (2020). The impact of municipal territorial reforms on the economic performance of local governments. A systematic review of quasi-experimental studies. *Space and Polity*, *25*. 10.1080/13562576.2020.1747420

Gomez, J. L., & Willoughby, K. (2008). Performance informed budgeting in U.S. state governments. *Presupuesto y Gasto Público*, *51*, 287-303.

Grayson, T. (2005). *What would Kentucky be like without its large number of counties*. Kentucky Land Office. http://www.sos.ky.gov/land/journal/articles/bryant/counties.htm. (February 2010)

Grosskopf, S., Margaritis, D., & Valdmanis, V. (1995). Estimating output substitutability of hospital services: A distance function approach. *European Journal of Operational Research*, *80*(3), 575-587.

Gyimah-Brempong, K. (1987). Economies of scale in municipal police departments: The case of Florida. *Review of Economics and Statistics*, *69*(2), 352-356.

Gyimah-Brempong, K., & Gyapong, A. O. (1991). Production of education: Are socioeconomic characteristics important factors? *Eastern Economic Journal*, *17*(4), 507-521.

Handry, M., & Martinez-Vazquez, J. (2008). Understanding sub-national government fragmentation in Indonesia and options for reform: Background for a "Grand Strategy" for Pemekeran. Unpublished.

Harbord, R. M., & Higgings, J. P. (2008). Meta-regression in Stata. *The Stata Journal*, 8*(*4), 493-519.

Hirsch, W. Z. (1959). Expenditure implications of metropolitan growth and consolidation. *The Review of Economics and Statistics*, *41*(3), 232-241.

Hirsch, W. Z. (1965). Cost functions of an urban government service: refuse collection. *The Review of Economics and Statistics*, *47*(1), 87-92.

Hortas-Rico, M., & Solé-Ollé, A. (2010). Does urban sprawl increase the costs of providing local public services? Evidence from Spanish municipalities. *Urban Studies*, *47*(7), 1513-1540.

Hortas-Rico, M., & Salinas, P. (2014). Determinación de la escala mínima eficiente en la provisión de bienes públicos locales. *Revista de Economía Aplicada*, *66*(XXII), 35-65.

Hortas-Rico, M., & Ríos, V. (2019). Is there an optimal size for local governments? A spatial panel data model approach. *Regional Studies, 54*(7), 958-973. 10.1080/00343404.2019.1648786

Huedo-Medina, T., Marin-Martinez, F., Sanchez-Meca, J., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *CHIP Documents*, Center for Health, Intervention, and Prevention (CHIP). University of Connecticut.

Jarrell, S. B., & Stanley, T. D. (1990). A meta-analysis of the union-nonunion wage gap. *Industrial and Labor Relations Review*, *44*(1), 54-67.

Jimenez, E. (1986). The structure of educational costs: Multiproduct cost functions for primary and secondary schools in Latin America. *Economics of Education Review, 5*(1), 25-39.

Jordan, M. M., & Hackbart, M. M. (1999). Performance Budgeting and Performance Funding in the States: A States Assessment. *Public Budgeting & Finance*, *19*(1), 68-88.

Kaufmann, D., Kraay, A., & Mastruzzi, M. (2008). Governance matters VII: Aggregate and individual governance indicators 1996-2007. http://econpapers.repec.org/paper/wbkwbrwps/4654.htm. (December 2009)

Kenny, L. W., & Schmidt, A. B. (1994). The decline in the number of school districts in the U.S.: 1950-1980. *Public Choice*, *79*(1/2), 1-18.

Knapp, M. (1982). Economies of scale in local public services: The case of British crematoria. *Applied Economics*, *14*(5), 447-453.

Kumar, R. C. (1983). Economies of scale in school operation: Evidence from Canada. *Applied Economics*, *15*(3), 323-340.

Lago-Peñas, S., & Martínez-Vázquez, J. (2013). *The challenge of local government size: Theoretical perspectives, international experience and policy reform.* Edward Elgar, Cheltenham (UK).

Lewis, P. G. (1998). *Deep roots: Local government structure in California*. San Francisco: Public Policy Institute of California.

Lowery, D., & Lyons, W. E. (1989). The impact of jurisdictional boundaries: An individual-level test of the Tiebout model. *The Journal of Politics*, *51*(1), 73-97.

Mariano Tommasi, & Federico Weinschelbaum. (1999). A principal-agent building block for the study of decentralization and integration. *Universidad de San Andres, Departamento de Economia.* http://ideas.repec.org/p/sad/wpaper/20.html. (October 2008)

Martinez-Vazquez, J. (1981). Selfishness versus public 'regardingness' in voting behavior. *Journal of Public Economics*, *15*(3), 349-361.

Martinez-Vazquez, J., & McNab, R. M. (2003). Fiscal decentralization and economic growth. *World Development*, *31*(9), 1597-1616.

Martinez-Vazquez, J., Rider, M., & Walker, M. B. (1997). Race and the structure of school districts in the United States. *Journal of Urban Economics*, *41*(2), 281-300.

Martinez-Vazquez, J., & Timofeev, A. (2008). Fiscal perspective of state rescaling. *International Center for Public Policy*, 08-06. http://aysps.gsu.edu/isp/files/ispwp0806.pdf.(March 2009)

Martinez-Vazquez, J., & Yao, M. (2009). Fiscal decentralization and public sector employment: A cross-country analysis. *Public Finance Review*, *37*(5), 539-571.

Matas, A., & Raymond, J. (1998). Technical characteristics and efficiency of urban bus companies: The case of Spain. *Transportation*, *25*(3), 243-263.

McDavid, J. C. (2000). Alternative service delivery in Canadian local governments: The costs of producing solid waste management services. *Canadian Journal of Regional Science*, *23*(1), 157-174.

Mehay, S. L. (1981). The expenditure effects of municipal annexation. *Public Choice*, *36*(1), 53-62.

Melkers, J., & Willoughby, K. (2005). Models of performance-measurement use in local governments: Understanding budgeting, communication, and lasting effects. *Public Administration Review*, *65*, 180-190.

Melkers, J., & Willoughby, K. (2001). Budgeters' views of state performance-budgeting systems: Distinctions across branches. *Public Administration Review*, *61*(1), 54-64.

Miyazaki, T. (2018). Examining the relationship between municipal consolidation and cost reduction: An instrumental variable approach. *Applied Economics*, *50*(10), 1108-1121. DOI: 10.1080/00036846.2017.1352077

Mizutani, F., & Urakami, T. (2001). Identifying network density and scale economies for Japanese water supply organizations. *Papers in Regional Science*, *80*(2), 211-230.

Morgan, D. R., & Mareschal, P. (1999). Central-city/suburban inequality and metropolitan political fragmentation. *Urban Affairs Review*, *34*(4), 578-595.

Musso, J. A. (2001). The political economy of city formation in California: Limits to Tiebout sorting. *Social Science Quarterly*, *82*(1), 139-153.

Nelson, A., & Foster, K. (1999). Metropolitan governance structure and income growth. *Journal of Urban Affairs*, *21*, 309-324.

Nelson, M. A. (1990). Decentralization of the subnational public sector: An empirical analysis of the determinants of local government structure in metropolitan areas in the U.S. *Southern Economic Journal*, *57*(2), 443-457.

Nelson, M. A. (1992). Municipal amalgamation and the growth of the local public sector in Sweden. *Journal of Regional Science*, *32*(1), 39-53.

Niaounakis, T., & Blank, J. (2017). Inter-municipal cooperation, economies of scale and cost efficiency: An application of stochastic frontier analysis to Dutch municipal tax departments. *Local Government Studies*, *43*(4), 533-554. DOI: 10.1080/03003930.2017.1322958

Niskanen, W. A. (1975). Bureaucrats and politicians. *Journal of Law and Economics*, *18*(3), 617-643.

Oates, W. E. (1972). *Fiscal federalism*. NY: Harcourt Brace Jovanovich.

Oates, W. E. (1991). Searching for Leviathan: An empirical study. *Studies in fiscal federalism*, 415-424.

Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups.* Cambridge, Mass. & London: Harvard University Press.

Olson, M. (1969). The principle of 'fiscal equivalence': The division of responsibilities among different levels of government. *American Economic Review*, *59*(2), 479-487.

Ostrom, E., & Parks, R. (1973). Suburban police departments: Too many and too small? In L.H. Masotti & J.K. Hadden (Eds.), *The urbanization of the suburbs*. Beverly Hills, CA: Sage Publications. 367-402.

Ostrom, E., Schroeder, L., & and Wynne, S. (1993). *Institutional incentives and sustainable development, infrastructure policies in perspective.* Westview Press.

Ostrom, V., Tiebout, C. M., & Warren, R. (1999). The organization of government in metropolitan areas: A theoretical inquiry. In M. McGinnis (Ed)., *Polycentricity and local public economies: Readings from the Workshop in Political Theory and Policy Analysis*. University of Michigan Press.

Panzar, J. C., & Willig, R. D. (1977). Economies of scale in multi-output production. *The Quarterly Journal of Economics*, *91*(3), 481-493.

Pels, E., Nijkamp, P., & Rietveld, P. (2003). Inefficiencies and scale economies of European airport operations. *Transportation Research: Part E: Logistics and Transportation Review*, *39*(5), 341-361.

Piñero, J. M., Pérez, C., & Conthe, J. (2021). Tamaño óptimo de los municipios españoles calculado a través del coste efectivo de los servicios. *Instituto de Estudios Fiscales, Papeles de trabajo* 1/2021.

Prieto, A.M., Zofio, J.L. & Álvarez, I. (2015). Cost economies, urban patterns and population density: The case of public infrastructure for basic utilities. *Papers in Regional Science*, *94*(4), 795-816. 10.1111/pirs.12096.

Prud'homme, R. (1995). The dangers of decentralization. *The World Bank Research Observer*, *10*(2), 201-220.

Razin, E. (2000). The impact of local government organization on development and disparities--A comparative perspective. *Environment and Planning C: Government and Policy*, *18*(1), 17-31.

Razin, E., & Rosentraub, M. (2000). Are fragmentation and sprawl interlinked? North American evidence. *Urban Affairs Review*, *35*(6), 821-836.

Reschovsky, A., & Imazeki, J. (1997). *The development of school finance formulas to guarantee the provision of adequate education to low-income students.* National Center for Education Statistics. Developments in School Finance.

Reschovsky, A., & Imazeki, J. (1999). Does the school finance system in Texas provide students with an adequate education? Paper prepared for presentation at the Annual Meeting of the American Education Finance Association, Seattle, Washington, March 18-20, 1999.

Riker, W. H., & Ordeshook, P. C. (1968). A theory of the calculus of voting. *The American Political Science Review*, *62*(1), 25-42.

Riverbank, W., & Kelly, J. (2006). Performance budgeting in the municipal government. *Public Performance and Management Review*, *30*(1), 35-46.

Sancton, A. (2008). Structural reform in Canada. In B. Dollery & L. Robotti (Eds.), *The theory and practice of local government reform.* Studies in Fiscal Federalism and State-local Finance. Cheltenham, U.K.: Edward Elgar Publishing Limited.

Schneider, M. (1986). Fragmentation and the growth of local government. *Public Choice*, *48*(3), 255-263.

Schwab, K. (2010). *The global competitiveness report 2009–2010.* World Economic Forum.

Seabright, P. (1996). Accountability and decentralisation in government: An incomplete contracts model. *European Economic Review*, *40*(1), 61-89.

Shelton, C. A. (2007). The size and composition of government expenditure. *Journal of Public Economics*, *91*(11-12), 2230-2260.

Sjoquist, D. L. (1982). The effect of the number of local governments on central city expenditures. *National Tax Journal*, *35*(1), 79-87.

Sole-Olle, A., & Bosch, N. (2005). On the Relationship between authority size and the costs of providing local services: Lessons for the Design of intergovernmental transfers in Spain. *Public Finance Review*, *33*(3), 343-384.

Stanley, T. D. (2001). Wheat from chaff: Meta-Analysis as quantitative literature review. *The Journal of Economic Perspectives*, *15*(3), 131-150.

Stanley, T., & Doucouliagos, H. (2007). Identifying and correcting publication selection bias in the efficiency-wage literature: Heckman meta-regression. *Working Paper - Economic Series 2007*, 2007/11.

Stansel, D. (2005). Local decentralization and local economic growth: A cross-sectional examination of US metropolitan areas. *Journal of Urban Economics*, *57*(1), 55-72.

Sterne, J. A., & Harbord, R. M. (2004). Funnel plots in meta-analysis. *The Stata Journal*, *4*(2), 127-141.

Stevens, B. J. (1978). Scale, market structure, and the cost of refuse collection. *Review of Economics and Statistics*, 6*(*3), 438-448.

Sutton, A., Abrams, K., Jones, T., Sheldon, D., & Song, F. (2000). *Methods for meta-analysis in medical research.* Chichester: John Wiley.

Tiebout, C. M. (1956). A Pure Theory of Local Expenditures. *The Journal of Political Economy*, 64(5), 416-424.

Tiebout, C. M. (1960). Economies of scale and metropolitan governments. *The Review of Economics and Statistics*, *42*(4), 442-444.

Tran, C., Kortt, M., & Dollery, B. (2018). Population size or population density? An empirical examination of scale economies in South Australian local government, 2015/16. *Local Government Studies, 45*(5)*.* 10.1080/03003930.2018.1501364

Turley, G., McDonagh, J., McNena, S., & Grzedzinski, A. (2018). Optimum territorial reforms in local government: An empirical analysis of scale economies in Ireland. *The Economic and Social Review*, *49*(4), 463-488.

Warner, M., & Hefetz, A. (2002). Applying market solutions to public services: An assessment of efficiency, equity, and voice. *Urban Affairs Review*, *38*(1), 70-89.

Warner, M. E. (2006). Inter-municipal cooperation in the U.S.: A regional governance solution? *Urban Public Economics Review/Revista de Economia Publica Urbana*, *6*, 221-240.

Wolff, P. (2004). *Economies of scale and scope in river basin management*. The Pacific Institute, Oakland California.

World Bank. (2006). *Political institutions database*. http://econ.worldbank.org.

Zax, J. S. (1989). Is there a Leviathan in your neighborhood? *American Economic Review*, *79*(3), 560-567.