

8-18-2010

Essays on Optimal Jurisdictional Size for Local Service Delivery

Juan L. Gomez

Follow this and additional works at: http://scholarworks.gsu.edu/pmap_diss

Recommended Citation

Gomez, Juan L., "Essays on Optimal Jurisdictional Size for Local Service Delivery." Dissertation, Georgia State University, 2010.
http://scholarworks.gsu.edu/pmap_diss/43

This Dissertation is brought to you for free and open access by the Department of Public Management and Policy at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Public Management and Policy Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

**ESSAYS ON OPTIMAL JURISDICTIONAL SIZE
FOR LOCAL SERVICE DELIVERY**

A Dissertation
Presented to
The Academic Faculty

By

Juan Luis Gómez Reino

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Public Policy

Georgia State University
Georgia Institute of Technology

August 2010

Essays on Optimal Jurisdictional Size for Local Service Delivery

Approved by:

Dr. Jorge Martinez-Vazquez, Advisor
Andrew Young School
of Policy Studies
Georgia State University

Dr. Jenny Ligthart
Faculty of Economics and Business
Administration
Tilburg University

Dr. Francois Vaillancourt
Faculty of Arts and Sciences
University of Montreal

Dr. James Alm
Andrew Young School
of Policy Studies
Georgia State University

Dr. Katherine Willoughby
Andrew Young School
of Policy Studies
Georgia State University

Dr. Juan D. Rogers
School of Public Policy
Georgia Institute of Technology

Date Approved: April 14, 2010

ACKNOWLEDGEMENTS

I owe my deepest gratitude to Dr. Jorge Martinez-Vazquez. Without his guidance and constant inspiration this dissertation would not have been possible. Since our first distant meeting in old Hanoi, I have enjoyed some of the most enriching professional years of my life working alongside him at the International Studies Program. I am greatly indebted to my dissertation committee and Professors Dr. Jenny Ligthart, Dr. Francois Vaillancourt, Dr. Katherine Willoughby, Dr. James Alm and Dr. Juan D. Rogers for their superb advice and generous dedication.

Thank you, Paul and Shereen, for everything. It would be hard to understate how important your friendship has been for me over these years. Thank you Alfie, for walking with me this road, all the roads.

But most importantly, this thesis is dedicated to my family, Juan, Llanos, Angel y Nanos. You fly with me wherever I go. You are worth all the efforts.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iii
LIST OF TABLES.....	iv
LIST OF FIGURES.....	v
SUMMARY.....	vi
CHAPTER 1. OPTIMAL JURISDICTIONAL SIZE FOR PUBLIC SERVICE PROVISION: A CRITICAL REVIEW OF THE STANDARD THEORY.....	1
1.1 On the optimal size of local government.....	1
1.2 A formulation of the standard theoretical model.....	12
1.3 An Extension: Incorporating Political Accountability.....	21
1.4 Accounting for different government levels.....	28
CHAPTER 2. EMPIRICAL EVIDENCE OF ECONOMIES OF SCALE IN LOCAL PUBLIC SERVICES: A META-ANALYSIS.....	33
2.1 Introduction.....	33
2.2 Economies of scale, definition and interpretations.....	34
2.3 “Stylized facts” from the review of the literature.....	36
2.4 Meta-analysis: a systematic, quantitative literature review.....	46
2.5 Meta- regression analysis.....	69
2.6 Conclusions.....	81
CHAPTER 3. JURISDICTIONAL FRAGMENTATION AND ITS DETERMINANTS: A CROSS-SECTIONAL INTERNATIONAL STUDY.....	84
3.1 Introduction.....	84
3.2 A review of the literature on government fragmentation and its implications.....	91
3.3 Empirical Model.....	106
3.4 Results.....	127
3.5 Conclusions.....	140
CHAPTER 4. A ROADMAP FOR THE ANALYSIS OF JURISDICTIONAL FORMATION AND REFORM.....	142
4.1 Introduction.....	142
4.2 A set of guiding principles for jurisdictional reform.....	145
4.3 Conclusions.....	158
REFERENCES.....	159

LIST OF TABLES

Table 1. Moderator Variables Coded from the Literature Review.....	59
Table 2. Descriptive Statistics: Cost Elasticity of Output.....	63
Table 3. Fixed Effects - Publication bias analysis.....	71
Table 4. Random Effects - Publication bias analysis.....	72
Table 5. Meta-regression results – Random Effects Estimation.....	76
Table 6. Sectoral Disaggregation – Random Effects Estimation.....	80
Table 7. Correlations: Number of jurisdictions and selected sample variables....	89
Table 8. Summary Statistics of the Dataset.....	125
Table 9. Results from the Tobit Model.....	129
Table 10. Results from the OLS Estimation.....	138
Table 11. Institutional Options for Service Delivery: A comparison.....	155

LIST OF FIGURES

Figure 1. Optimal Jurisdictional Size	5
Figure 2. Long-run U-shaped Average Cost Curve (LAC).....	8
Figure 3. Alternative shapes of the Long-Term Average Cost Curve	10
Figure 4. Alternative shapes of the Long-Term Average Cost Curve.....	11
Figure 5. Funnel Plot Analysis.	65
Figure 6. Funnel Plot Analysis – Sectoral Distribution	68
Figure 7. Sub-national Government tiers per country.....	87
Figure 8. Number of sub-national jurisdictions per country.....	88

SUMMARY

This dissertation contributes to the definition of an analytical framework for the study of optimal jurisdictional size for local service delivery. We argue that the standard economics framework for the analysis of optimal jurisdictional size importantly neglects individual preferences for political accountability. Our theoretical model shows that once we take into account such preferences, the optimal jurisdictional size for the provision of local public goods is smaller than in the standard model. We obtain empirical evidence to support our hypothesis from a sample of 197 countries. Our results show that, in fact, demand for political accountability leads to higher jurisdictional fragmentation both in terms of greater number of jurisdictions and smaller average population per jurisdiction. In addition, a meta-analysis of the empirical contributions to the study of economies of scale in the provision of local services shows that the economies of scale expected from service provision to larger jurisdictional sizes may not be present except for a handful of local services, and limited to relatively small population sizes. The results of the meta-analysis signal moderately increasing to constant returns to scale in the provision of traditional local services. In light of these results, we argue that forced jurisdictional consolidation programs across the world justified by perceptions of excessive jurisdictional fragmentation, or by the expectation of large expenditure savings due to economies of scale may have been, thus, erroneously designed. From a policy perspective, multi-layered institutional frameworks for service delivery (including cooperation and privatization among other options) may allow targeting available efficiency gains more efficiently than consolidation.

CHAPTER 1

OPTIMAL JURISDICTIONAL SIZE FOR PUBLIC SERVICE PROVISION:

A CRITICAL REVIEW OF THE STANDARD THEORY

“...the sizes of national states (or countries) are due to trade-offs between the benefits of size and the costs of heterogeneity of preferences over public goods and policies provided by government.” (Alberto Alesina & Spolaore, 2003)

1.1 On the optimal size of local government

Standard economic theory (Oates, 1972) defines the problem of optimal jurisdictional size in the provision of services as consisting of two main tradeoffs. First, the trade off existing between the welfare gains expected from smaller governments (better placed to match expenditure allocation to local preferences) and the economies of scale (or associated lower average costs) expected from the delivery of services at larger jurisdictional sizes. On that basis, equilibrium would be reached when the difference between the marginal welfare gains from more efficient provision and the marginal costs derived from foregone economies of scale is maximized.

The second critical tradeoff determining optimal jurisdictional size is that between the closer accountability offered by smaller governments to their citizens and the higher management, administrative, and information costs associated with multiple jurisdictions for service delivery. In truth, although Oates (1972) presents this disjunctive as a second trade off in his discussion, the latter trade off can be taken as a corollary to the first.

Management and administration costs can readily be assumed as part of the service production costs and, similarly, political accountability is a requirement to ensure citizens preferences are reflected in local public expenditure, increasing overall welfare. Equilibrium is reached in Oates' (1972) model when the welfare gains associated with improved tailoring of service provision to citizens' preferences are maximized relative to the costs of production.

The analytical framework proposed by Oates for the identification of the optimal jurisdictional size for local government provision departs from the consideration that the individual consumer's surplus is reduced when a collective decision must be taken to determine the level of public service provided to the jurisdiction. It is most likely that the level of the public good G provided will be either above or below the level that would have maximized individual utility thus the loss of consumer surplus incurred¹. The social welfare cost of such a collective decision on the level of G would be equal to the sum of the individual losses of welfare in Oates (1972) model.

Because we believe an individual consumer surplus approach is intuitive and helpful let us assume, following Baleiras (2001), that the representative individual derives utility from the consumption of a private good (y) and a pure public good (G) according to the following quasi-linear utility function

$$U(y_j, G) = y_j + u(G)$$

where y_j is the individual's consumption of the private good, and G is the quantity of public good produced. If the utility obtained from the consumption of the public good is

¹ Presumably for everyone except the median voter.

twice continuously differentiable ($U' > 0$ and $U'' < 0$), then changes to the goods' prices are reflected in changes to the individual's utility. With a quasi-linear utility function, the change in consumer's surplus reflects the utility impact of a price variation. The social welfare change is the result of the summation across all individuals of the changes in consumer surplus (Baleiras, 2001).

In the case of a single, representative consumer, he would choose to consume the amount of public good that maximizes $U(y_j, G)$. However, when two or more consumers are considered and the amount of public good provided is the result of a "collective decision" (i.e. direct democracy, representative political systems, etc.), then there will be a welfare loss associated to a level of public good provision that may not be optimal to any of the individuals. For some consumers, the level of G provided will be below their individual optimal level and for other individuals it will be over their optimal level. Thus, we can think of the cost curve C representing a "collective agreement" for public service delivery (see Figure 1). Following Alesina & Spolaore (2003) linear uni-dimensional model, if we note G^*_i as the individual's preferred level of G , we assume that the loss of welfare derived from the provision of a collective level of G is the same whether G exceeds or is below G^*_i by the same amount. If G is below the individual's preferred level, there is a missing input in the individual utility function. For levels of G above the preferred individual level, the increase in tax price decreases the utility of the public good².

² This is for instance the case of elderly couples without school-age children. Such couples have no preferences for public education expenditure (finances traditionally with property tax revenues in the USA), except to the extent that good schools improve property values of their districts. Beyond that point, too much public education turns into a loss of utility.

Should the decision on the level of public good provision be taken via popular vote (i.e. direct democracy), then, provided certain assumptions are met, we would expect the preferences of the median voter to be decisive. However, there are other “collective decision mechanisms” available, such as representative democracy, whereby the incentives and preferences of political representatives would also come to play, to determine such level of provision.

On the other hand, the average cost of production of the public good may be reduced as more individuals share it; this will be so if we assume the good is produced with economies of scale over a certain range. Such economies of scale may be derived from declining average fixed production costs over larger production ranges, or from the efficiency gains obtained from specialization in production that may only be efficient when output reaches a certain level. Therefore, we would expect the benefits derived from cost reduction (represented by the curve B in Figure 1) to rise sharply once additional individuals are added to the pool of beneficiaries, as corresponds to a convex relationship. The marginal increase in welfare derived from the addition of a new individual decreases eventually until it levels off. Above an equilibrium size of N^* , the sum of individual welfare losses from enjoying a G that is above or below individuals’ optimal levels may completely offset the cost-saving advantages from economies of scale. As Oates puts it, as “group size increases, the influence of any particular individual on his own level of consumption of the good diminishes (1972; p.41)”.

Thus, the optimal jurisdictional size (in population terms) is defined, in the conventional Oates’ model as the government size where the difference between the marginal social benefit obtained from average cost reduction and the increased marginal social cost

derived from a level of G that does not match the individuals' optimal choice is maximized.

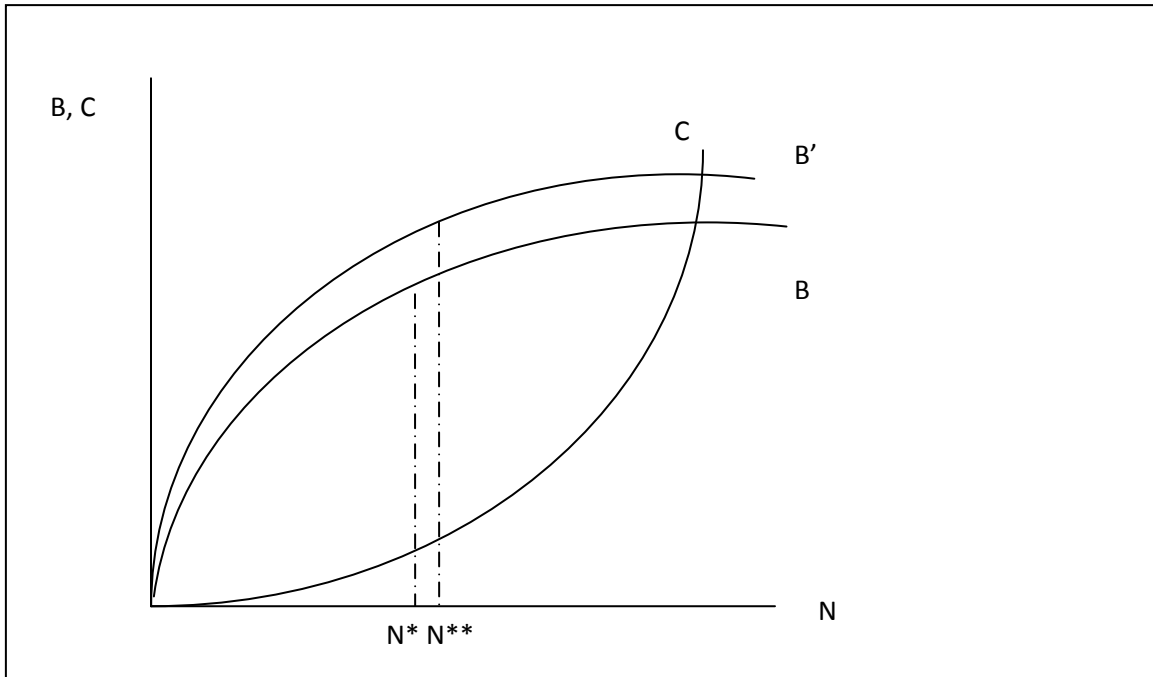


Figure 1. Optimal Jurisdictional Size

The simplicity of the conventional model is appealing. It includes in a uni-dimensional analysis both the pattern of economies of scale in the benefit curve and the losses in allocative efficiency derived from the collective determination of the level of public good. Baleiras (2001) shows this optimal jurisdictional size is Pareto efficient.

An important corollary of Oates' analysis is that different public goods achieve different optimal jurisdictional sizes of provision. Thus, a federal (or otherwise decentralized)

system of government whereby public goods and services are delivered by different levels of government would allow maximizing allocative efficiency.

The assumptions behind this model include the need for a constant and immobile population, zero transaction costs for collective decision-making and that the financing of the public good provision be efficiency-neutral (i.e. lump sum taxation). However, the assumptions of the model can be relaxed to account for these issues.

Should the public good be subject to important externalities, the benefit curve may be underestimating the social benefits accruing from the good's provision. This would be the case, for instance, of a hospital providing services to patients that reside outside its jurisdiction. Estimating social benefits only on the basis of the population of the jurisdiction hosting the hospital would underestimate overall impact and "catchment" area of benefits. In this example, the model would thus suggest that larger jurisdictional sizes are required in order to internalize the benefits/losses caused by the provision of this public good and match the jurisdiction with the benefit area to avoid under provision. This matching between different public goods with different benefit areas is known as perfect correspondence in Oates (1972)³. The analysis of the welfare implications of "discrepancies between the boundaries of a public good and those of the jurisdiction" are discussed in Olson (1969), who already points to the need to establish a wide range of government institution to reach what he calls "fiscal equivalence". The social benefit

³ Several different other terms have been used in the literature to denote this process of matching jurisdictional size to the benefit area of the provision of a public good. Breton (1965) used "perfect mapping" to conceptualize this issue. Stigler (1957) called it "fiscal correspondence" in his study of the responsibilities of different levels of government in the USA.

curve would shift outwards, increasing the optimal population size for provision to N^{**} (Figure 1 above).

The effect of relaxing the assumption of an immobile population is ambiguous. On the one hand, mobility may reduce the cost of smaller jurisdictions if consumers are efficiently sorted among different jurisdictions, but on the other hand, it may subject public good provision to congestion costs. This would lead to the implementation of tolls and charges for public good consumption as developed in the economic theory of clubs by Buchanan (1965). Lastly, the assumption of zero transactions costs for collective decision making, should it be relaxed in Oates model, may lead to larger jurisdictions and fewer governments in order to minimize the cost of government and of inter-governmental fiscal relations. The presumption is that transaction costs will be smaller for larger jurisdictions due to economies of scope that reduce the cost of policy coordination and public good delivery Boyne (1992). This argument, as we will see, does not seem to be backed up by empirical evidence in government cost savings after jurisdictional consolidation (Bish, 2001).

Oates (1972) argues that this analysis should guide policy makers and academics in answering which level of government should provide the good (i.e. *determine the level of consumption*), but that the production function of a particular good may indicate the desirability of contracting with another local government, a higher government level, or even privatizing the production of the good. The distinction between allocative and productive efficiency is thus outlined clearly by Oates. Interestingly, he briefly mentions that privatization or contracting out to other governments may “permit a different type of resolution of this trade off problem”. That, for some public goods, joint production or

privatization may allow enjoying the benefits of decentralized provision (differing levels per government) and the cost advantages from economies of scale. Thus, the basic model does offer important insights as to the efficient institutional set-up that should accommodate the provision of the public good. However, this option is not fully developed.

Assuming that the cost function of a generic public good shows increasing returns to scale, being thus U-shaped although not necessarily symmetrical, a local government may find itself unable to take advantage of the economies of scale due to reduced size (i.e. and lack of investment capacity for instance), as in point A in Figure 2.

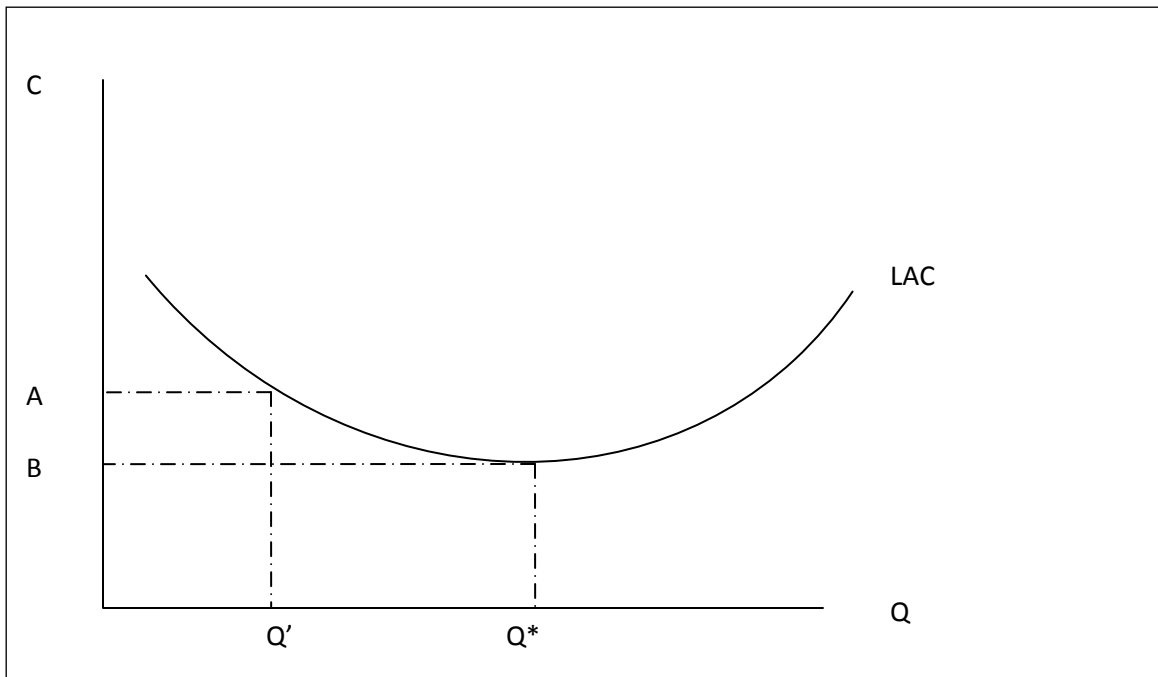


Figure 2. Long-run U-shaped Average Cost Curve (LAC)

A larger government size would allow reaching the productive capacity required to move to B, where the full extent of the economies of scale is taken advantage of. The investment required to produce the amount of public good required may be prohibitively high for a small local government, and thus it may find, if it decided to produce the good by itself, that it would be located at point Q' in Figure 2.A, a sub-optimal state.

Oates' model may arguably suggest that a larger jurisdictional size would allow reaching B and benefiting from the cost savings derived. Perhaps logically, resort to local governments' consolidation, anchored in the assumption of available economies of scale, has dominated jurisdictional reform around the world in search for more efficient local public sectors. As we have seen, once the main assumptions of the model are relaxed, the model predicts that lower levels of government fragmentation would be more efficient. Thus, allowing for externalities from the provision of public goods suggest the need for larger governments that internalize those effects. Equally, the costs of collective decision making, if assumed not to be zero, are to be reduced by larger government units that avoid duplication of functions (i.e. economies of scope).

It may also be the case that not just a single jurisdictional size achieves economies of scale, but that the latter are available in full extent once a certain benchmark has been passed, as in Figure 3. In this case, larger jurisdictional size after Q* (Figure 3) does not provide greater cost savings, but constant average costs. In any case, cost curves such as the one depicted in Figure 2 or 3 would signal limits in the extent of economies of scale. Beyond a certain level of production, long term average costs either increase or remain

constant. For the purpose of determining an optimal jurisdictional size, it would then seem that the economies of scale are constrained to certain population limits that are largely dependent on the available technology.

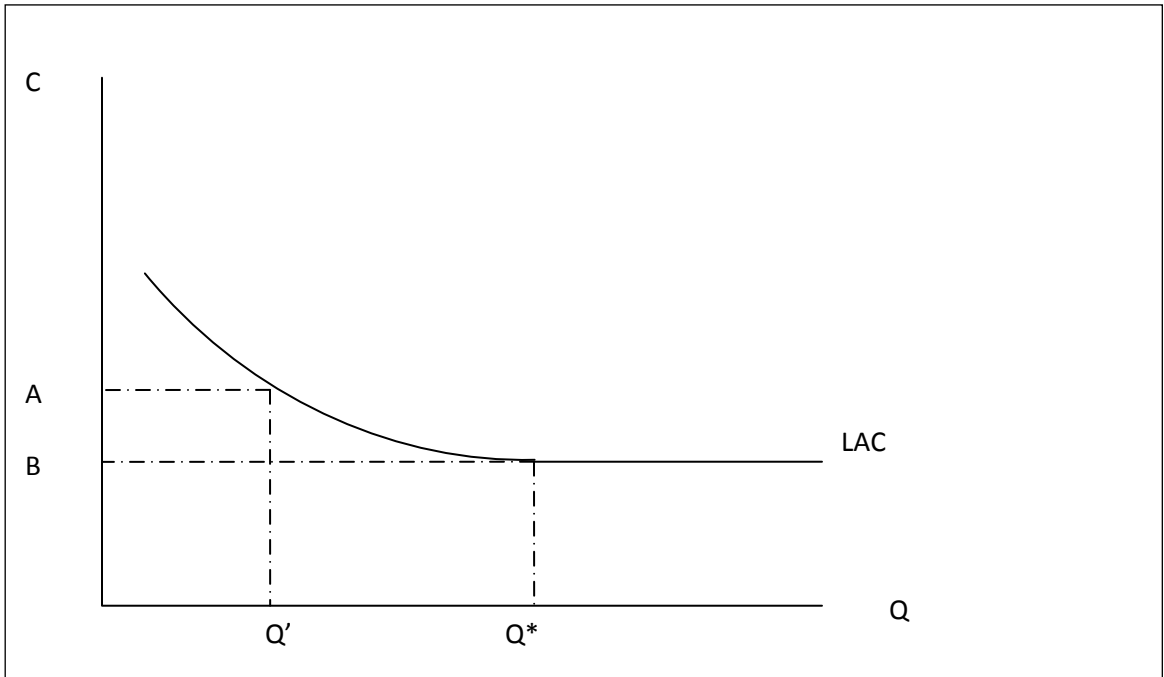


Figure 3. Alternative shapes of the Long-Term Average Cost Curve

Alternatively, the production function of the service may reflect constant costs of production independently of the level of production (as in Figure 4 for the flat LAC curve), and thus the size of the jurisdiction would not affect the unit cost of production. This is typical of labor-intensive public goods, where there is no need of large capital investment start-ups.

The observation of local service delivery patterns across countries shows, however, that resorting to other institutional forms of service production may also allow reaching the

optimal production level Q^* maintaining a given level of government fragmentation. This is an important issue that Oates hints at only in passing, but one with considerable implications for the “real world”, not fully considered in the model. In particular, Oates (1972) argues that, *for some services*, it may be possible to retain the benefits from decentralized provision and produce with economies of scale by resorting to privatization or service agreements among governments. It would seem, in principle, that the whole range of local services would be eligible for those arrangements, at least those that exhibit economies of scale.

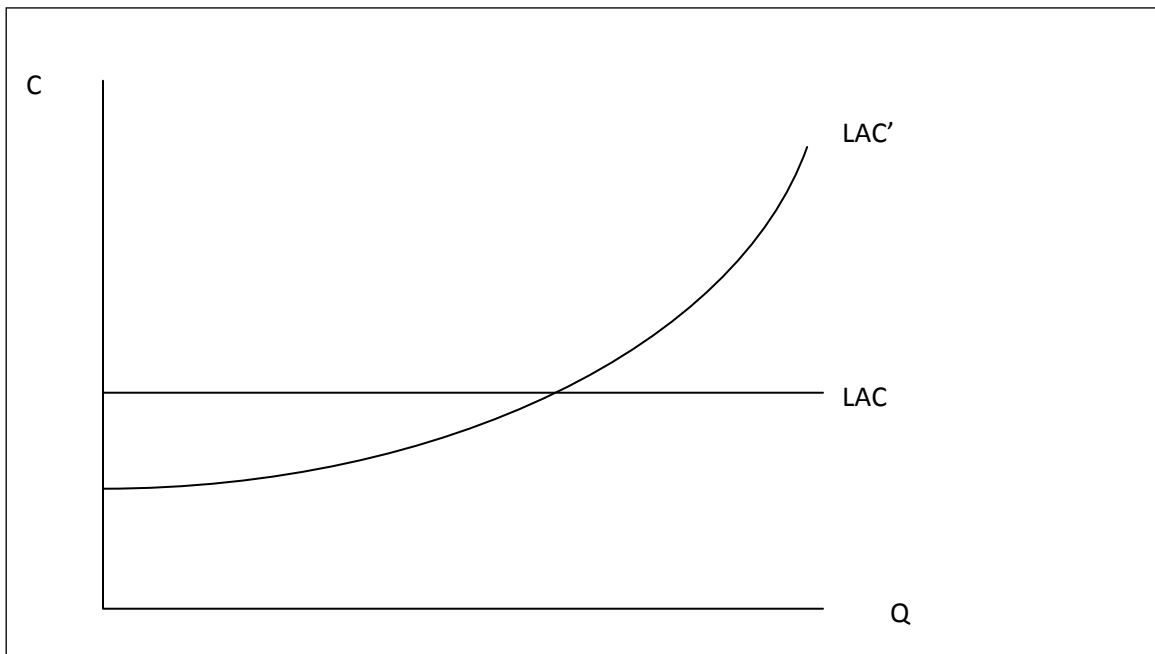


Figure 4. Alternative shapes of the Long-Term Average Cost Curve

1.2 *A formulation of the standard theoretical model*

In the mathematical appendix to his seminal work on fiscal decentralization, Oates (1972) derives the equilibrium quantity that minimizes the welfare loss incurred from providing a common level of production (as opposed to that preferred by single individuals) *for a given group* of people. His solution is however solely based on the minimization of the loss of welfare derived from the joint consumption of a level of public good that is likely not to be equal to each individual's preferences. This solution does not incorporate the other element of the trade-off, the benefits derived from the reduction in price from economies of scale and larger size. In fact, Oates does not model mathematically the optimal jurisdictional size model, but solves for the optimal level of public good provision given a certain group size. In this section, we provide a complete formulation of the trade-off, with the idea of extending this analysis to the possibilities presented by alternatives institutional forms for service delivery.

For extending the model, we will employ Buchanan's (1965) formulation of a theory of clubs (which is also the theoretical formulation of Oates' optimal jurisdictional size theory), but allowing for heterogeneous individual preferences. Our approach derives a social welfare function (from the summation of individual utility functions) that in itself contains the main trade-off between economies of scale and matching heterogeneous preferences presented by Oates. The optimal size of the group will then be a function of the public service production costs and of the level of jointly-provided public good and how the latter varies from each individual's optimal level. We also delve into Alesina &

Spolaore (2003) modeling of optimal jurisdictional size by adopting a uni-dimensional linear model whereby losses of welfare can be measured as distance between the individual's preferred consumption level of G and the level provided as a result of the collective decision.

As we discussed, we assume in our model a constant and immobile population, zero transaction costs for collective decision-making and that the financing of the public good provision be efficiency-neutral (i.e. lump sum taxation).

We assume, as before, that individuals have a quasi-linear utility function of the form:

$$(1) U_i = y_i + F_i(DG_i),$$

where DG_i is equal to the difference, in absolute value, between the level of public good provided and the level that would maximize individual welfare, for each individual. Following Oates' original formulation, the sum of the differences between the level of public good provided and that which would maximize the individuals' utility is likely to increase with the size of the jurisdiction (measured in population terms), and thus, in this initial formulation:

$$(2) DG_i = DG(G(N), G^*_i)$$

Thus, the larger the difference between the level of public good provided and the individual's equilibrium quantity (G_i^* , signaling the heterogeneous preferences), the larger the loss in welfare, and the larger the amount by which overall individual utility would be reduced. Additionally, such distance to the individual optimum will be a function of the size of the jurisdiction in population terms N , as the level of G is a collective decision that will vary with population size. For simplicity, we assume that the utility decreases in an increasing way from larger distances from the individual's optimal level of G , and that equal distances to the individual optimum, be G above or below G^* , translate into identical losses in welfare:

$$(3) \partial U_i / \partial DG_i < 0, \text{ and } \partial^2 U_i / \partial DG_i^2 > 0$$

$$(4) \partial U_i / \partial DG_1 = \partial U_i / \partial DG_2, \text{ for each } DG_1 = G_i^* + \zeta, \text{ and } DG_2 = G_i^* - \zeta,$$

We define t (the tax price paid for the public good) as equal to the average cost of provision, or total cost (C) of production divided by the size of the population served. In turn, the total cost is defined as a function of the quantity produced:

$$(5) t = C/N; \text{ where } C=C(G)$$

As the size of the jurisdiction (N) increases, the average cost of provision of the public good (G) decreases, improving the individual's welfare level. The individual's utility

from lower taxes (or average production costs), increases however in a decreasing way, represented by the positive value of the second derivative.

$$(6) \partial U_i / \partial t < 0, \text{ and } \partial^2 U_i / \partial t^2 > 0$$

This is graphically represented by curve B in our Figure 1 above. The implication of (6) above is that individual utility increases with the consumption of the private good, in a decreasing way:

$$(7) \partial U_i / \partial y > 0, \text{ and } \partial^2 U_i / \partial y^2 < 0$$

An additional member to the jurisdiction adds another person to the collective decision-making mechanism for the determination of G. Thus the sum of the distances between individual preferences of G (G_i^*) and the effective level finally provided increases, decreasing overall utility.

$$(8) \Sigma \partial U_i / \partial N < 0$$

This is graphically represented in curve C of our Figure 1 above.

The budget restriction of the individual is equal to:

$$(9) M_i = y_i + t$$

which states that the individual's income (assumed exogenously determined) is spent on consumption of the private good y and the public good G for which he pays a price equal to t .

From the individual maximization problem we obtain the individual's ideal level of public good provision (G_i^*), which is the level of public good that would maximize the individual's utility given a set of prices. In the following social welfare maximization problem, such preferred individual level is considered as exogenous.

Lastly, the transformation function for the economy is of the form $F(G, y, N) = 0$, meaning that the production cost depends on the quantity produced and on the quantity of private good required that has to be surrendered for its production. Again to adhere to Oates initial formulation, we assume that the production cost for the public good presents economies of scale up to a certain range, after which it exhibits decreasing returns to scale later, thus giving us a U-shaped cost function. This assumption is later on relaxed and its implications explored further.

Under this framework, the individual faces the following maximization problem:

$$(10) \quad \begin{aligned} & \text{Max } [y_i + U_i(DG_i(G(N), G_i^*))] \\ & \text{s.t. } M_i = y_i + t \end{aligned}$$

By assuming a quasi-linear utility function, social welfare changes are equal to the summation of changes in welfare of all the individuals due to variations in t .

Thus, if the government behaves as a benevolent planner, it will try to maximize social welfare, defined as the sum of all individual utilities (again assuming the quasi-linear form of the utility functions), solving for the optimal size of the group for service provision.

Thus we have a social welfare function (SW) of the form:

$$(11) \quad SW = \sum U_i = \sum y_i + \sum U_i (DG_i(G(N), G^*_i))$$

The constraint for the entire economy is given by the transformation function:

$$(12) \quad F(G, y, N) = 0$$

On those bases, the optimal jurisdictional size would be reached where the Lagrangean \mathcal{L} below is maximized:

$$(13) \quad \mathcal{L} = \sum y_i + \sum \{U_i (DG_i(G(N), G^*_i))\} - \lambda F(G, y, N)$$

As in Oates (1972), we will assume that the production technology is exogenous, which will contribute to determine an optimal level of production of the public good. Thus, an

increase in the number of individuals sharing the cost increases the individual utility but at a decreasing rate. This is a realistic assumption in light of empirical evidence limiting the range of economies of scale to certain production ranges, and this would define the shape of the benefit curve (B) in Figure 1.

Solving for the optimal group size involves the effects on individual (and social) utility derived from the differences between individual preferences and group provision (the larger, the larger welfare losses), and the reduction in costs from economies of scale. As expected from clubs theory, the optimal size of the group is reached when the marginal utility of adding an extra member to the group (negative in this case due to its expected influence on DG) equals the marginal cost (also negative as a cost reduction is expected from adding yet another taxpayer to the group) of producing the service.

From equation (13) we get the first order conditions:

$$(14) \quad \partial \mathcal{L} / \partial y_i = 1 - \lambda (\partial F / \partial y_i) = 0$$

$$(15) \quad \partial \mathcal{L} / \partial G = \Sigma \{ (\partial U_i / \partial DG_i) * (\partial DG_i / \partial G) \} - \lambda (\partial F / \partial G) = 0$$

$$(16) \quad \partial \mathcal{L} / \partial N = \Sigma \{ (\partial U_i / \partial DG_i) * (\partial DG_i / \partial G) * (\partial G / \partial N) \} - \lambda (\partial F / \partial N) = 0$$

Dividing (15) by (14) we obtain:

$$(17) \quad \Sigma (\partial U_i / \partial DG_i) * (\partial DG_i / \partial G) = (\partial F / \partial G) / (\partial F / \partial y_i)$$

The left hand side of this equation represents the change in total welfare derived from a change in the level of provision of the public good G, which is equal to the marginal rate of substitution between the public good (G) and the private good (Y), as $(\partial U_i / \partial Y) = I$. On the right hand side of the equation we have the marginal rate of transformation between the public and the private good. So equivalently:

$$(18) \quad \Sigma MRS_{G,y}^i = MRT_{G,y}$$

which is the well-known Samuelson condition, also reached in Buchanan's (1965) analysis of clubs. We can obtain the optimal size of N dividing equation (16) over (14):

$$(19) \quad \Sigma \{ (\partial U_i / \partial DG_i) * (\partial DG_i / \partial G) * (\partial G / \partial N) \} = (\partial F / \partial N) / \Sigma (\partial F / \partial y_i)$$

which is also the well-known membership condition in Buchanan's (1965) theory of clubs.

We can therefore re-write (19) as:

$$(20) \quad \Sigma MRS_{N,y}^i = MRT_{N,y}$$

The RHS of (20) is the change in the average cost of provision from adding additional individual to the population sharing the cost of provision, so we have that

$$(21) \quad MRT_{N,y} = \partial(C/N)/\partial N = -C/N^2$$

When we substitute in (20) and solve for N, we obtain:

$$(22) \quad N^* = \sqrt{-C/(\sum MRS^i_{N,y})}$$

Thus the optimal size of the jurisdiction is positively related to the size of the economies of scale and negatively related to the welfare costs derived from heterogeneous preferences.

As expected, the optimal jurisdictional size is a function of the production costs and the economies of scale derived from the production technology and of the proximity of the level of production of the good to the individual preferences of the citizens. This is also consistent with Alesina & Spolaore (2003) formulation where optimal jurisdictional size is inversely related to the marginal cost of distance (a measure of welfare loss from the difference between G^* and individual optimal G levels), and positively related to economies of scale.

In our formulation, we have assumed a U-shaped long run average cost curve which determines the shape of our benefit curve B. The graphical analysis presented by Oates in his Fiscal Federalism assumes such a cost structure. We could however consider the alternatives to the cost functions discussed earlier in the literature review and the way it would affect the solution to the model.

In their study of the size of nations, Alesina & Spolaore (2003) model mathematically the basic Oates framework for optimal jurisdictional size, using a uni-dimensional model whereby physical distance between the individual preferences and the effective level of public good provided denote loss of welfare from collective decision making.

This standard framework has been extended in different directions. Besley & Coate (2003) show that Oates' assumption of uniform provision of public good under a centralized system can be relaxed without altering the main model result: that under heterogeneous preferences for public goods and no spillovers, a decentralization system is superior. They also explore the outcomes in terms of public expenditure for different political economy models of legislative decision making, minimum winning coalition and cooperative legislatures.

1.3 An Extension: Incorporating Political Accountability

As discussed, the Oates model does not incorporate, in the definition of the optimal jurisdictional size, the fact that preferences for political representation/accountability may affect significantly the size of the optimal jurisdiction. The model assumes that in smaller jurisdictions the level of public good provision will be more closely tailored to citizens' preferences. This relationship is not however that straight forward, as the capacity to influence government decision may depend on other variables. Arguably, as population increases (i.e. jurisdictional size), the relative power of the individual to influence government decisions (i.e. political accountability) is diluted. Additionally, in the absence

of a decentralized system of government, local political representatives may simply be central government appointees. In such a case, local representatives would mostly be accountable to the central authorities that appointed them, and not to the citizens they would be meant to serve, regardless of the size of the jurisdiction. Secondly, even if political representatives are elected locally, they may not enjoy sufficient fiscal autonomy so as to significantly determine expenditure patterns in their local governments. This is usually the case when local governments are overly dependent on fiscal transfers from the central government, very commonly conditional in the nature of expenditure they should be used on. Thirdly, even when that autonomy exists for local level representatives, the institutional fabric may not be designed so as to allow for citizens' preferences to be conveyed regularly to the local authorities. Participatory budgets, citizens' score cards, and local assemblies are all institutions that allow consistent feedback from citizens to their local authorities on their perceptions and preferences regarding local service delivery.

An approach to modeling the role of political accountability on the optimal provision of public goods under centralized/decentralized systems of government is that of Seabright (1996). In his model, political accountability is defined as the probability that the welfare of any given locality is the decisive factor in the re-election of its government. Obviously, centralization implies a loss of political accountability in that regard. Seabright's results apply even in the case of homogeneous preferences by local citizens. His model uses localities' welfare (and not individuals') as the unit of analysis, disregarding the technology through which individual preferences are translated into localities' welfare.

Importantly, Seabright's model, although it considers public good provision across three different levels of government (national, regional and local), it assumes away the possibility of economies of scope in government action.

In an interesting contribution, Tommasi & Weinschelbaum (1999), using a common agency principal-agent model also attempt to capture the implications of a preference for citizen's control of the government in the decision to opt over more centralized versus decentralized structures of government. Tommasi & Weinschelbaum (1999) provide a good discussion of the channels through which smaller jurisdictional size may assist improved political accountability. Those include 1) by having a smaller number of voters, there is a higher individual power to affect final output; 2) higher accountability of local officials due to their physical proximity to the citizens (Ostrom, Schroeder, & Wynne, 1993); and 3) the capacity to compare government's performance across a range of local authorities (yardstick competition). Their model however assumes each individual (principal) signs a separate contract with the government (agent) as opposed to the social contract that is traditional enforced by voting under either direct or representative democracy. In addition, the choice of level of government is conducted at the start of the process, and as such is independent initially of the welfare results.

We approach this issue with a simple extension of the basic model described above that includes a new variable, political accountability (P), affecting the way preferences are reflected into a government's level of provision of the public good G. It is expected that a higher level of political accountability (assumed to be costless) is preferred by citizens, as

a signal that their preferences will be more likely incorporated into public policy (and budgetary decisions). The greater the size of the jurisdiction (N), the lower the political accountability, that is, the lower would be the capacity of citizens to affect the level of public good provided.

Therefore, with the inclusion of P:

$$(23) \quad DG = DG(G(N), G^*_i, P), \text{ whereby}$$

$$(24) \quad (\partial DG / \partial P) < 0$$

Thus it is expected that a higher level of political accountability would result in an overall lower sum of the distances between individual preferences and the level of provision of the public good.

So our new representative individual's utility function would be:

$$(25) \quad U_i = y_i + U_i(DG(G(N), G^*_i, P), \text{ and as described:}$$

$$(26) \quad P = P(N, D, FA, CP)$$

Where D represents the existence of a decentralized government with locally elected representatives, FA reflects the degree of fiscal autonomy, and CP is a measure of citizen representation in government decision-making (i.e. such as the existence of participatory budgeting systems, etc.).

An increase in jurisdictional size is related to lower individual capacity to influence the decisions of the politicians, thus reducing political accountability as defined in our model.

$$(27) \quad (\partial P / \partial N) < 0$$

In this case, the individual faces the constraint:

$$(28) \quad \begin{aligned} & \text{Max } y_i + U_i [DG(G(N), G^*_i), P(N)] \\ & \text{s.t. } M_i = y_i + t \end{aligned}$$

Where $\partial U_i / \partial P > 0$, $\partial^2 U_i / \partial P^2 < 0$, that is, an increase in political accountability is positively valued by the citizens up to a point where diseconomies of representation (in the form of perhaps red-tape, bureaucracy, etc.) offset gains from additional political accountability.

Similarly, the maximization problem faced by the benevolent planner of the jurisdiction is:

$$(29) \quad \text{Max } \mathcal{L} = Y + \sum U_i [DG(G(N), G^*_i), P(N)] - \lambda(F(G, Y, N))$$

From where we obtain the first order conditions:

$$(30) \quad \partial \mathcal{L} / \partial Y = 1 - \lambda (\partial F / \partial Y) = 0$$

$$(31) \quad \partial \mathcal{L} / \partial G = \Sigma \{ (\partial U_i / \partial DG) * (\partial DG / \partial G) \} - \lambda (\partial F / \partial G) = 0$$

$$(32) \quad \partial \mathcal{L} / \partial N = \Sigma \{ (\partial U_i / \partial DG) * (\partial DG / \partial G) * (\partial G / \partial N) + (\partial U_i / \partial P) * (\partial P / \partial N) \} - \lambda (\partial F / \partial N) = 0$$

Thus, in equilibrium:

$$(33) \quad \Sigma \{ (\partial U_i / \partial DG) * (\partial DG / \partial G) \} = (\partial F / \partial G) / (\partial F / \partial Y)$$

From where we obtain identical Samuelson condition: $\Sigma MRS^i_{G,Y} = MRT^i_{G,Y}$

How will the inclusion of our political accountability variable affect optimal size? As in the basic model, we divide (32) by (30), substitute $\lambda = 1 / (\partial F / \partial Y)$ and re-arrange terms:

$$(34) \quad \Sigma \{ (\partial U_i / \partial DG) * (\partial DG / \partial G) * (\partial G / \partial N) + (\partial U_i / \partial P) * (\partial P / \partial N) \} = (\partial F / \partial N) / (\partial F / \partial Y)$$

Or equivalently:

$$(35) \quad \Sigma MRS^{iP}_{N,y} = MRT_{N,y}$$

where the superscript P denotes the marginal rate of substitution once political accountability is included in the individual utility function.

By definition, as $(\partial U_i / \partial P) > 0$ and $(\partial P / \partial N) < 0$, then $(\partial U_i / \partial P) * (\partial P / \partial N) < 0$

As $\Sigma\{(\partial U_i / \partial DG) * (\partial DG / \partial G) * (\partial G / \partial N)$ is, by definition, negative (loss in social welfare from the addition of one more member to the jurisdiction), with the addition of $(\partial U_i / \partial P) * (\partial P / \partial N)$ we conclude that

$$(36) \quad |\Sigma MRS_{N,y}^{iP}| > |\Sigma MRS_{N,y}^i|$$

The absolute value of the loss in social welfare from one additional member to the jurisdiction is greater once we include a preference for political accountability in the individual's welfare function. Not only does utility decrease due to the addition of a new consumer's welfare loss in the form of $G - G^*_i$, but now citizens also lose individual capacity to influence the decision over G.

Solving for N, we obtain:

$$(37) \quad N_p^* = \sqrt{(-C / \Sigma MRS_{N,y}^{iP})}$$

Which leads us to conclude that $N_p^* < N^*$, that is, the optimal jurisdictional size for the provision of the public good is smaller once we introduce preferences for political accountability in the individual's utility function.

The size of the jurisdiction (N) affects welfare in a two-fold way. It does so first, by increasing the welfare costs of agreeing on a common level of provision G (greater $\sum DG_i$). The larger the group size, the larger the summation, for all citizens, of the distance between an individual's preferred level and the level of good finally provided. This effect is independent of the political system in place. Even with a direct democracy system, where the level provided is that preferred by the median voter's, the larger the size of the group, the larger will be the loss of welfare from that decision. Secondly, the larger the jurisdictional size, the larger the distance between the citizens and their representatives. Political accountability is reduced, and with it, the capacity to influence the level of provision of the good.

1.4 Accounting for different government levels

As discussed, one of the least explored dimensions of jurisdictional fragmentation is the vertical structure of government, that is, the varying number of levels of government we observe in countries around the world. Oates (1972) decentralization theorem offered a guide for the assignment of expenditure responsibilities to different government levels, but takes the vertical structure of government as given and exogenous to the model.

The determination of the efficient number of levels of government is still largely an unresolved issue which we explore in this dissertation. Logically, the applied policy analysis of the vertical structure of government has taken it as given due to the high political and financial cost of consolidating or creating a new level of government. The

analysis of social welfare in the provision of public goods by Seabright (1996) compares the outcomes of provision by central, regional and municipal levels of government, but does not discuss the efficiency in the creation of a new level of government.

We approach this issue by describing a normative, first best model where we assume an economy with three types of public goods $i=1, 2$ and 3 . For simplicity, we assume public good number 1 presents the characteristics of a pure public good, that is, is perfectly non-excludable and non-rival. Public goods 2 and 3 are impure public goods whereby the externalities from the production of public good 2 are greater than those of 3.

Additionally, we assume that in the original state, only a central and a local level of government exist. Central government is assigned responsibility over the provision of public good 1 (e.g. defense), whereas goods 2 and 3 are provided by the local government. In this framework, under what conditions would the creation of an intermediate (or regional) level of government be justified?

As discussed, the “trivial” answer to this question is that the regional level of government may be already defined constitutionally and thus be exogenously determined. In our model, where a benevolent planner decides the most efficient delivery options for the public goods, the decision to create an intermediate (or regional) level of government would be justified by improvements in social welfare.

Thus, assuming the utility function of the individual is now dependent on the levels of provision of the three public goods we have:

$$(38) \quad U_i = y_i + U_i(DG^1(G^1, G^{1*}_i), DG^2(G^2, G^{2*}_i), DG^3(G^3, G^{3*}_i), P)$$

$$s.t. M_i = y_i + t$$

Expression 38 simply expands our utility function as defined in (25) to reflect the existence of three public goods. Equally, the social welfare function will be equal to

$$(39) \quad SW = \Sigma y_i + \Sigma U_i(DG^1(G^1, G^{1*}_i), DG^2(G^2, G^{2*}_i), DG^3(G^3, G^{3*}_i), P)$$

which is simply the sum of the individual utilities as in our basic model. In the initial situation we had assumed that due to its pure public good characteristics, public good 1 was being produced by the central government level, while the delivery of public goods 2 and 3 was assigned to the local government level.

It is straight forward to deduce that a benevolent planner will assign responsibility to a regional (or intermediate) level of government if, once the administrative and transactions costs, plus the loss of political accountability incurred from the assignment of the responsibility over the delivery of a public good to a higher level of government are accounted for, the efficiency gains from lower average costs from economies of scale exceed the loss in allocative efficiency from a larger jurisdiction.

If we assumed that the public good 2, with greater externalities in its provision, is provided by a regional level of government instead of the local level of government, the decision will be justified in efficiency terms if:

$$(40) \quad SW^{2R} > SW^{2L}$$

Or equally, the creation of a regional level of government would be justified if the social welfare from the provision of public good 2 by a regional government (SW^{2R}), ceteris paribus, exceeds the social welfare obtained from the original distribution of responsibilities (SW^{2L}).

Again, that will mean that for the provision of good 2, the efficiency gains derived from the economies of scale obtained by the regional government overwhelm not only the efficiency losses derived from a level of provision less tailored to local preferences, but also those from lower political accountability, and the higher administrative costs derived from a new level of government. In addition, should economies of scope exist, the efficiency gains from lower average costs of production would have to make up for the loss in economies of scope as well.

This simple model may for instance be an accurate description of the composition of certain regions in Spain. After the implementation of the organic law on the organization of regional governments, regions were created by the provinces in a voluntary way. For

the so called “historical” regions of Spain, such as Catalonia and Galicia, the decision of which provinces would create the region was clear and justified in historical and cultural as well as economic and political terms. For other provinces however, the decision to join a new region was not straight forward. Several alternatives were available for those provinces, among them the option to become a single-province region.

Provinces evaluated several aspects of the decision to form a region. In the first place, joining provinces with similar preferences for public good provision was obviously an important factor in the decision to form a certain region. But also the provincial political weight in the new regional institutions was an important aspect to consider. Thus, provinces were more likely to form regions where their political leverage (and thus accountability to their citizens) was greater. All these aspects are considered in our simple model.

The obvious extension to this model, not developed in this dissertation, is to account for the positive aspects of the formation of new levels of government. This would for example involve developing a sequence for the formation of regional governments where local authorities evaluate the impact of such an event on their re-election chances among other things. In any case, our simple model allows us to test in Chapter 3 of this dissertation some basic hypothesis about the current vertical structure of government across countries.

CHAPTER 2

EMPIRICAL EVIDENCE OF ECONOMIES OF SCALE IN LOCAL PUBLIC SERVICES: A META-ANALYSIS.

“Most researchers conclude 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.” (Bish 2001. p.14)

2.1 Introduction

The standard theory of optimal jurisdictional size developed by Oates and extended later on by other authors (e.g., Alesina & Spolaore (2003)) hinges importantly on the existence of economies of scale in the provision of local public goods and services. Despite the fact that many countries have embarked on forced jurisdictional consolidation programs on the basis of insufficient economies of scale in the delivery of local public services, an empirical justification for this type of policy remains elusive. Although it would be unreasonable to expect that local governments' size match adequately the efficient scale of production for public services, it could be equally unreasonable to assume, for all services involved, that larger jurisdictions will necessarily have a more efficient scale. The ambiguous and sometimes conflicting evidence regarding the existence of economies of scale in local public service provision demands at this time a systematic quantitative analysis that summarizes and evaluates the evidence available on this issue. The goal of this chapter is to produce this quantitative review using a meta-analysis approach.

2.2 *Economies of scale, definition and interpretations*

The internationally widespread political drive for less sub-national government fragmentation is underpinned by the search for economies of scale that are supposed to reduce the average cost of output, or as Dollery & Crase (2004; p.268) put it, “a decrease in the cost per person for a given amount of service as population served increases”. 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, we can also say that economies of scale exist when we can increase the production of a good or service without increasing production costs in the same proportion.

The source of such economies of scale can be 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 additionally 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, Panzar, & Willig (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 one respectively. In elasticity terms, economies of scale exist when the cost elasticity of output (ε_y) is smaller than 1. ⁴

The literature has predominantly used this definition of economies of scale, although other contributions have also merited attention. In particular, Caves, Christensen, & Tretheway (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. Thus, in their interpretation, short-term economies of scale are defined as:

$$RTS = \frac{1}{\varepsilon_y + \varepsilon_N}$$

where ε_y is the cost elasticity of output and ε_N is the cost elasticity of 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 inputs (Z) and thus:

⁴ The definition implies that the cost elasticity of output cannot be zero, as that would lead to infinite economies of scale, an unreasonable result.

$$RTS = \frac{1 - \varepsilon_Z}{\varepsilon_Y + \varepsilon_N}$$

where ε_Z is the cost elasticity of quasi-fixed production inputs⁵. 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. Important contributions to the literature have adopted however the interpretation of economies of scale proposed by Caves et al. (1984), such as in Mizutani & Urakami (2001), Aubert & Reynaud (2005) or Filippini & Prioni (2003). A thorough review of the available empirical evidence yields, as we will see, an inconclusive picture. This is perhaps to be expected. There does not seem to be, at least *a priori*, a strong theoretical case to assume that any local public service will display similar patterns in terms of economies of scale across the wide range of existing local governments and technologies available for public service production (Dollery & Crase, 2004).

2.3 “Stylized facts” from the review of the literature

An initial review of the literature unveils a series of “stylized facts” that help shape our quantitative analysis below. This section benefits from excellent earlier reviews such as those from Boyne (1995), Andrews, Duncombe, & Yinger (2002) in the area of education, Byrnes J. & Dollery B. (2002) for Australian local governments, or more recently Bel (2009) for selected sectors.

⁵ By quasi-fixed production inputs the authors refer to the fact although in the long-run all inputs are traditionally assumed to be variable, some of them, including capital and labor for instance, can be just partially adjustable.

a. Capital vs. Labor-Intensive Services

First, from our theoretical framework, 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). In Chile, Albala-Bertrand & Mamatzakis (2004) find economies of scale in the provision of transport, sewerage, and power grid services. Bel (2005) and Alvarez, Caride, & Gonzalez (2003) show that solid waste collection and processing offers important savings in production costs derived from larger client populations in Spain. This is a finding shared equally by Callan & 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 in Canada.

Following Bel (2009), the review of recent work in the analysis of returns to scale in the operation of airports displays constant returns to scale in relation to airplanes movement and strong economies of scale in passenger traffic for smaller airports. Efficiency gains disappear however when the analysis is extended to very large, international airports, reflective of a traditional U-shape long term average cost curve (Pels, Nijkamp, & Rietveld (2003); Bazargan & Vasigh (2003)). 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 in the works of Matas & Raymond (1998) for Spain and Filippini & Prioni (2003) for Switzerland. In addition, increasing although moderate returns to scale in the Swiss urban transport are found by Farsi, Fetz, &

Filippini (2007). More conclusive is the evidence obtained from works in the area of garbage collection, where solid evidence of economies of scale are generally found (Bel, 2009).

A corollary of the above cost theory proposition is that labor intensive local services should offer less potential for economies of scale. A pioneering reference is the work of Hirsch (1959) for U.S. municipalities, discarding the existence of economies of scale in police services. Examining 44 cities and districts of Seattle's metropolitan area, Ahlbrandt (1973) does not find economies of scale in the provision of firefighting services. Similar conclusions are found in the studies of Alt (1971), Boaden (1971) and Danzinger (1978) in England and Wales. In the United States, Ostrom & Parks (1973), Dilorenzo (1981) and Gyimah-Brempong (1987) find evidence of higher production costs in firefighting and police services with the greater population size of the jurisdiction. On the other hand, Bodkin & Conklin (1971) report declining average costs of production with higher population size in these two commonly local U.S. services.

However, in the provision of public schooling -- a labor intensive service although one with potentially higher overhead costs than police or fire services for instance, economies of scale are found by Chambers (1978), Butler & Monk (1985, for relatively small districts), Callan & Santerre (1990), Duncombe, Miner, & Ruggiero (1995), Jimenez (1986), and Reschovsky & Imazeki (1997, 1999)). On the other hand, Gyimah-Brempong & Gyapong (1991) find decreasing returns to scale in the production of education in Michigan school districts. A general conclusion of these studies was that such economies

of scale were extinguished at relatively low levels of enrolment. 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 less 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, while in Maine, Deller & Rudnicki (1992) estimated the optimal size of the education district at around 2000 students.

b. Measurement, measurement, measurement

The mixed evidence for economies of scale gathered from the empirical literature may well be due to critical differences in the measures of output and production costs used in the analyses. In their review of previous works on the existence of economies of scale in Australian local government, Byrnes & 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 this branch of the literature. Most of the studies Byrnes & Dollery (2002) review assume service production is responsive to population and thus use the latter as a proxy for the former. In addition, they argue, the common use of total expenditure as a measure of cost, although reasonable, does not solve the problem of allocating administrative costs, which may be substantial to the operation.

This general critique to the body of empirical contributions in this area is an old one. Already Tiebout (1960) criticized Hirsch's seminal contributions to the literature for his use of population as a proxy for public service output. Tiebout (1960; p.444) argued reasonably that "there is no necessary relationship between population and either the output or quality of the good". In fact, larger population may lead, Tiebout argues, 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 instead of cost data for the estimation of cost functions has also been severely 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 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 & 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 & Giuliano (1984), who find diseconomies of scale in bus operation in the US if bus-miles

are used as output measure and economies of scale if revenue per passenger is used. Thus it would seem clear that the existence of economies of scale in the delivery of public services may be reasonably dependent on the output measure selected.

c. A U-shaped Average Cost Function

A third salient aspect of the initial review of the literature in studies using jurisdictions (as opposed to production units) as the focus of analysis is the concentration of economies of scale in smaller (population-wise) jurisdictions. This signals perhaps 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. 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. More recently, 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 a sample of Catalonian municipalities, Bel & Fageda (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 empirical evidence 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 finding is more evident in a third group of studies that analyzes overall expenditure patterns before and after processes of jurisdictional consolidation or inter-municipal cooperation. It is perhaps to be expected that once we aggregate all local expenditure on public goods and services the evidence of economies of scale would be even more inconclusive. This is due to the fact that, as discussed, we would be aggregating both capital- and labor-intensive services, with different potential in the reduction of average production costs due to larger volume.

Thus, in a series of studies in the early 1970s, Davies, Barton, & Williamson (1971), and Davies & 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). Abelson (1981, in Byrnes & Dollery, 2002) finds that, for 36 Australian municipalities, the average public expenditure per household does not decrease as population increases. Conversely, the analysis of the Australian Institute of Public Affairs (1993, in Byrnes & Dollery, 2002) advises of

possible economies of scale in the consolidation of 210 counties in the province of Victoria. This report, together with that of the consulting firm KPMG (1998, in Byrnes & Dollery, 2002), which suggested potential savings of up to 24% in the provision of local public services for 177 Australian local governments were critical elements in the consolidation process that took place in Australia during the 1990s (Byrnes & Dollery, 2002).

d. Modeling frameworks for the cost function

The body of empirical literature on the analysis of economies of scale in local service delivery has seen three somewhat overlapping but otherwise well differentiated stages in the modeling of production cost functions. Early works in this area used a linear function, quadratic in the measure of output, to establish the existence of U-shaped cost curves. Contributions include the early works of Hirsch (1959, 1965), Bodkin & Conklin (1971), Beaton (1974), Knapp (1982), Kumar (1983), among others. The academic standards of the early contributions are arguably weaker than those of more recent works. Many of the articles reviewed from this early stage do not provide descriptive statistics of the variables used for instance, turning the calculation of elasticity into an impossible feat. The sample of observations for the meta-analysis 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 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 contribution from Stevens (1978) in the sector of refuse collection, Duncombe et al. (1995) in education, or Christoffersen, Paldam, & Wurtz (2007) in the cleaning of schools in Denmark.

The third and more recent wave of empirical works in this area have favored heavily the use of the translogarithmic cost functions as a modeling framework. Contrary to the Cobb-Douglas, 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 recently applied to every possible service including Drake & Simper (2002) in the area of police, Fabbri & Fraquelli (2000) in water supply, or 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.

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 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. Only a small share of locally provided services may offer economies of scale from larger size jurisdictions.

As a preliminary conclusion, it may be that a more careful examination of the evidence, in particular the realization of the limited size of economies of scale, may have led to the promotion and development of alternative institutional solutions to service delivery that did not involve the forced consolidation of sub-national units of government. But in addition, as Oates' trade off suggests, even when economies of scale exist, certain inefficiencies in costs may have been worth paying for in exchange for larger improvements in allocative efficiency from smaller size local governments. However, making these judgments requires a more precise and systematic knowledge of economies of scale in the production and delivery of public goods. We next turn to how this more precise information may be produced.

2.4 *Meta-analysis: a systematic, quantitative literature review*

As we have seen above, consistent empirical evidence of economies of scale in the delivery of local public services is somewhat elusive in the literature. The wide variety of public services studied, the numerous modeling and empirical frameworks employed, and the long time frame over which the literature in this area spans has complicated reaching solid conclusions on the patterns of such economies of scale when available. Thus, there is an obvious need to systematically and quantitatively review this important body of literature so as to distill its main findings and be able to inform public policy in this area in a less ambiguous manner. In this section we describe the process proposed for the development of a meta-analysis of the literature, which we believe goes a long way in meeting those goals.

The meta-analysis methodology allows combining the results of various studies testing a particular hypothesis with the ultimate objective of deriving the average “true effect” of the issue being analyzed, while controlling for the heterogeneity of studies examined. This statistical method adds value well beyond the simple average survey method, as each study estimate is weighted by its standard error. The weighted averages thus obtained from the meta-regression are more powerful and reliable estimates of the effects being considered than a simple average.

Meta-analysis as a methodology can assist in “summarizing, reviewing and evaluating empirical research results” (Stanley, 2001; p.131). The use of meta-analysis has become increasingly widespread in economics and other social sciences, but the methodology has

been especially popular among medical researchers that required consolidating the results from empirical tests with few individuals into more powerful statistical analysis (Sutton, Abrams, Jones, Sheldon, & Song, 2000). Seminal meta-analysis in economics include those of Jarrell & Stanley (1990) on the Union-Nonunion wage gap, and Card & Krueger's (1995) on time-series studies of the impact of minimum wage on employment creation. The methodology has become very popular among economists recently. An EconLit database search revealed a total of 385 references to meta-analysis between 1985 and 2009, a considerable and multidisciplinary effort, with contributions focusing on environmental, transport, and labor topics. Although still scarce, increasing use of the methodology seems to have been made in the area of public finance.⁶The meta-analysis methodology is especially apt as a way to consolidate results from controlled experiments. It is perhaps the relatively recent use of lab studies in the area of economics that has prevented a more extensive use of the methodology. The recent wave of experimental work in economics, traditionally based in small sample studies, will be likely to raise interest in meta-analysis in the near future.

In spite of that, the topic at hand seems like a perfect fit for the application of the meta-analysis methodology. The literature spans over several decades, the results in terms of the existence (or not) of economies of scale in local public service delivery are scattered and seem to be contingent of many factors (technology, population sizes, sample, modeling framework, cost function, etc.). Thus, the main objective of our meta-analysis

⁶ These include Feld & Heckemeyer's (2009) meta-analysis of FDI and taxation, Bom & Ligthart's (2009) on the productivity of public capital, or Blackwell's (2007) analysis of tax compliance experiments, the latter two still being work in progress.

is to conduct a quantitative, systematic review of the empirical literature on the existence of economies of scale on the provision of local public services to synthesize the existing empirical evidence to date. In the next paragraphs we follow Stanley's (2001) stages for the development of our meta-analysis.

a. Identifying all relevant studies and choosing a common metric

The first stage of a meta-analysis is to identify as complete a sample of studies in the area of interest as possible. We reviewed 103 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 and the Google and Bing standard search engines, without any time period restriction. We contacted several authors for unpublished papers, dissertations, and government reports, 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 103 papers overall on the topic, the selection of pieces for the meta-analysis used the following criteria. First, we selected for the meta-analysis 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. The latter 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, 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 truly do not test for economies of scale, but 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 but different contributions that 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. The hypothesis we would like 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. It makes sense to include all local services in our study because, in most countries, jurisdictional consolidation precisely requires the delivery of *all* services to a larger population, as opposed to a small or selective set of public services.⁷

⁷ 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

Positive testing of the hypothesis would provide support to the consolidation of small local governments into larger jurisdictional units.

Our theoretical framework and hypothesis defines the cost elasticity of output as the variable of interest for the study. This statistic allows summarizing the empirical results of the literature and is therefore used as the dependent variable of the meta-analysis⁸. 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 is not available and the paper is discarded.

The selection process outlined above left us eventually with 44 studies that reported 60 values of the cost elasticity of output for different services. Of those 60 observations, 44 reported their attached standard errors and 16 did not. The availability of standard errors is essential as the meta-analysis essentially weights the observations by their variance. In their absence, other measures of study size can be used (e.g. the inverse of the degrees of freedom of the study, for instance) but are generally less satisfactory. Our empirical work

services. But in this case the challenge can become the number of observations available to conduct reliable statistical analysis.

⁸ Earlier contributions in this literature, using as the dependent variable for the study the total cost of production (in monetary terms) did not offer a realistic or theoretically sound alternative for the selection of our statistic of interest.

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 (60) using a different measure of study size. In addition to a greater wealth of information, this allow for a sensitivity test of our results.

The data set includes studies from 1978 to 2008. As discussed, several studies offer more than one observation, and we include them all in the dataset. Observations for a total of 8 services and 19 countries are included. As it is traditionally the case with cross sections, the data set of studies suffers from unobserved heterogeneity, as not all relevant moderators may have been coded.

As we expected, 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 economies of scale a measure of network length. As it was not possible to completely homogenize the statistics from the studies, our meta-regression controls accordingly for

this fact with a variable (baumol) with value one for the studies using the Baumol et al. (1988) measure of economies of scale and 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 a large number of 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 sample by many different researchers.

Fortunately, all of the studies included in our dataset use different samples, which limits this type of dependency to the greatest possible extent, making our observations 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. It is the case however 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 instances.

b. 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 on the measure of production. The presence of this vast “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 studies. With a 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” (Harbord and Higgings 2008; p.493).

In a random-effects meta-regression, the individual study estimates of the variable of interest (in our case the cost elasticity of output) 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 & Higgings (2008) that:

$$y_i | \theta_i \sim N(\theta_i, \sigma^2), \quad \text{where } \theta_i \sim N(x_i\beta, \tau^2)$$

$$\text{and thus } y_i \sim N(x_i\beta, \sigma^2 + \tau^2)$$

equivalently,

$$y_i = x_i\beta + u_i + \varepsilon_i, \quad \text{where } u_i \sim N(0, \tau^2) \text{ and } \varepsilon_i \sim N(0, \sigma^2)$$

Our dependent variable for the meta-analysis is the cost elasticity of output obtained from the studies reviewed. Regressors or independent variables include the moderators identified during the process of coding and detailed above. Over this database, we perform Variance Weighted Least Squares (WLS) regression analysis⁹ in order to attach more weight to estimations with lower standard error and thus more accurate. 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 $1/(\sigma^2 + \tau^2)$.

The unconditional (average) cost elasticity of output in our sample of 61 observations is 0.7304, that is, if service output increases by a 1 percent, the cost of provision increases by 0.73 percent, signaling some substantial economies of scale. The results of the meta-regression are presented in section 4.5 below.

In our analysis we will assess the potential for publication bias in the sample. As discussed, we defined publication bias as the higher likelihood that a study is published if it reports statistically significant results. Thus, following Bom & Ligthart (2009) we assume that:

$$\hat{\theta}_i = \theta_i + g\left(se(\hat{\theta}_i)\right) + \mu_i$$

⁹ Using the *metareg* command of Stata

where $\hat{\theta}_i$ represents the observed estimates, θ_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. In Table 1 below we test for the existence of publication bias without the insertion of any moderator variables yet. 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 & Ligthart (2009) present one of the most comprehensive analysis of publication bias to date, and their analysis is replicated below. In particular, the analysis will allow identifying the direction of the bias and select the appropriate control for our meta-regression including all relevant moderators.

Following Bom & Ligthart (2009), we estimate the equation:

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

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_p (D_n) are dummy variables with value 1 if $\hat{\theta}_i$ is positive (negative) and zero otherwise. They are interacted with the

standard errors of the estimates $\hat{\theta}_i$. 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, as we do in Tables 1 and 2 below. If both the $\alpha_p se(\hat{\theta}_i)^h D_{pi}$ and $\alpha_n se(\hat{\theta}_i)^h D_{ni}$ terms 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.

c. Identifying moderator variables: The coding process and related hypotheses

During the coding process of the empirical papers identified we identified a large number of aspects that may have determined 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 Box 1. Not all of the moderators were eventually used in the regression analysis due to their lack of significance and to ensure adequate degrees of freedom. Here we discuss the issues with greater theoretical support.

First, our coding process included the creation of a variable for the country in which the study took place. Almost half of the 60 observations eventually considered for the meta-regression were from U.S.-based studies, 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 (i.e. bus public company for instance) 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 are included at the same time.

In terms of the characteristics of the dataset for each of the studies reviewed, we created dummies denoting where a cross-section, panel, or time series was used. Following Berechman & Giuliano (1984), cross sectional data renders biased estimates as it assumes homogeneity of the observations (i.e. jurisdictions, public companies, etc.). The direction of this bias is not clear however and is a matter left to the empirical analysis. Equally, we coded the estimation method of the cost function, a variable also closely linked to the dataset structure.

With regard to 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 limits the assumptions imposed on the behavior of the dependent variable, is expected to provide more solid estimations results although, again, the direction (or sign) of the coefficient is an empirical question.

A dummy variable was created to denote, with a value of one, if expenditure (as opposed to cost) data were used as dependent variable in the analysis; this moderator is expected to bias downwards the estimates of economies of scale if expenditure data are used. In addition, we created a dummy variable denoting whether the output variable used population as a proxy (value 0) or rather a physical measure of output (i.e. gallons of water, tones of garbage, etc.).

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 that estimated cost elasticity of output would be lower with the latter than with the former, as already discussed. 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, and for cases where more than one observation was obtained from the same study. Finally, our study size variables included the elasticity's standard errors and the degrees of freedom of each study.

Importantly, our sample of studies used a great diversity of data sets, which led us to conclude there is no data dependency or sample overlap in this particular case.¹⁰ We believe that all relevant dimensions of the studies could be coded, but as more

¹⁰ For example, an area where there is potential for dataset dependency is that of comparative fiscal decentralization studies, where the use of the International Monetary Fund's Global Financial Statistics is widespread.

observations are made available, the data set should continue to evolve and improve in the future.

Table 1. Moderator Variables Coded from the Literature Review

Dimension	Variable	Definition	Number of Observations
Year of Survey	1970s	Value one if survey year from that decade.	18
	1980s	Value one if survey year from that decade.	6
	1990s	Value one if survey year from that decade.	21
	2000s	Value one if survey year from that decade.	15
Years of Data	YearsData	Value one if more than one year, zero otherwise.	20
Sector	Education		12
	Water and Sanitation		17
	Garbage Collection		18
	Urban Transportation		9
	Police		2
	Cleaning of Schools		1
	Transport Infrastructure		1
Country	USA		28
	Spain		10
	UK		3
	Netherlands		2
	Other countries		17
Unit of Analysis	Jurisdictions	Value one if a jurisdictional unit is focus of analysis, zero if a production unit (municipal firm, etc.)	30

Table 1. Moderator Variables Coded from the Literature Review (Continued)

Estimation Methodology	OLS		27
	SUR		22
	MLE		4
	2SLS		3
	FE		2
	GLS		1
	GMM		1
Dataset Structure	Cross Section		40
	Panel		14
	Pooled		4
	Time series		2
Cost function form	Log linear		30
	Linear		2
	Translog		27
	Quadratic		1
Expenditure Data	Expenditure	Value 1 if expenditure data used for dependent variable, 0 if cost data	18
Output Data	Physical Output	Value 1 if a measure of physical output is used, 0 if population used as a proxy for output.	53
Dummies for Elasticity	Positive Elasticity	Value 1 is the cost elasticity observed is > 0	54
	Negative Elasticity	Value 1 if the cost elasticity observed is < 0	6
Production Method	Production Method	Value 1 if the analysis distinguishes between private and public production methods	19
Definition of EOS	Baumol	Value 1 if Baumol et al (1988) definition of EOS, 0 if Caves et al (1984)	54

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 moderator 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 for production) with lower relative fixed capital investment requirements, which may have reduced somewhat 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 latter ones.

Secondly, we coded the number of years for which data was available in each study and the total number of observations. These variables attempt to measure study size and the robustness of results. Another important control variable that has been coded is the service that is 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.

d. Funnel Plot Analysis

“Funnel plots are a visual tool for investigating 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 a measure of study size takes 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 the sample of 60 observations (from 44 studies) considered in this meta-analysis does not 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) will concentrate at the top around the “true” effect value.

The sample unconditional mean is .735, reflecting some economies of scale in principle, with a minimum value of cost elasticity of output of -.245 and a maximum of 1.524. Thus, observations range from large economies of scale to sizable decreasing returns to scale, as can be seen in Table 1 below.

Table 2. Descriptive Statistics: Cost Elasticity of Output

	Observations	Mean	Std. Dev.	Min	Max
All	60	.735	.375	-.245	1.524
Education	12	.447	.522	-.183	1.524
Garbage Collection	18	.905	.178	.272	1.091
Water and Sanitation	17	.729	.319	-.245	1.086

In Figure 1 below, Chart 1 displays the first funnel plot, where all 44 observations for which we have standard errors reported are plotted. They include observations for most of the services considered and all the functional forms of the cost function 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 underline 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 11 observations out of the 44 for which standard errors are available in our sample reported a cost elasticity of output below 0.5. In this group, 6 of those observations corresponded to studies in the education sector, and 8 of them used a log-linear function to model the cost function.

Chart 1. All observations with reported standard errors.

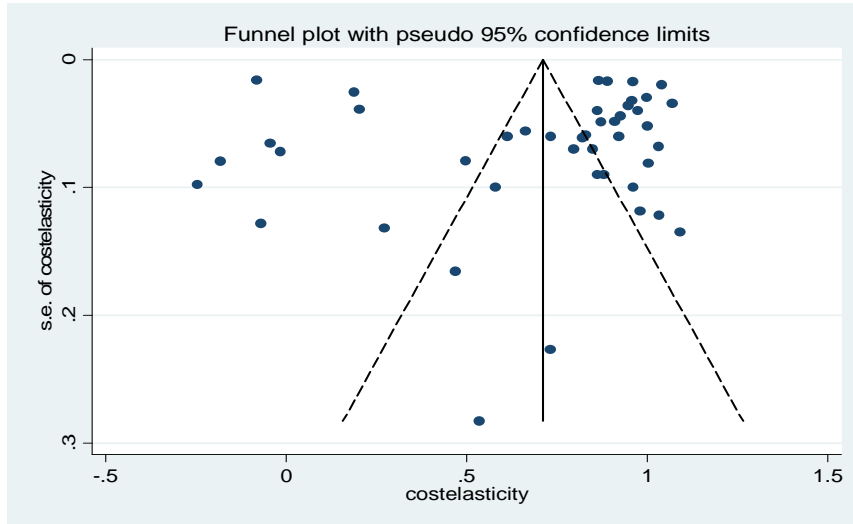


Chart 2. All obs. Weighted by Inverse of the degrees of freedom.

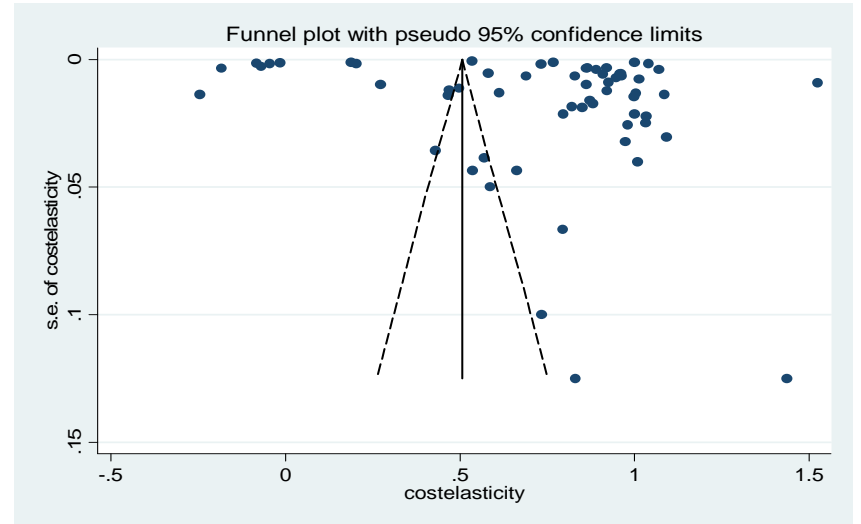


Chart 3. Observations using log-linear functions

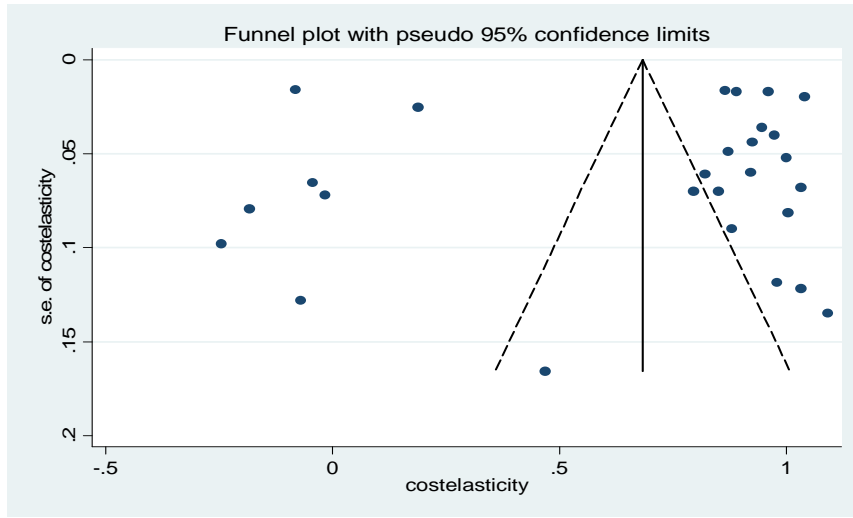


Chart 4. Observations using translog functions

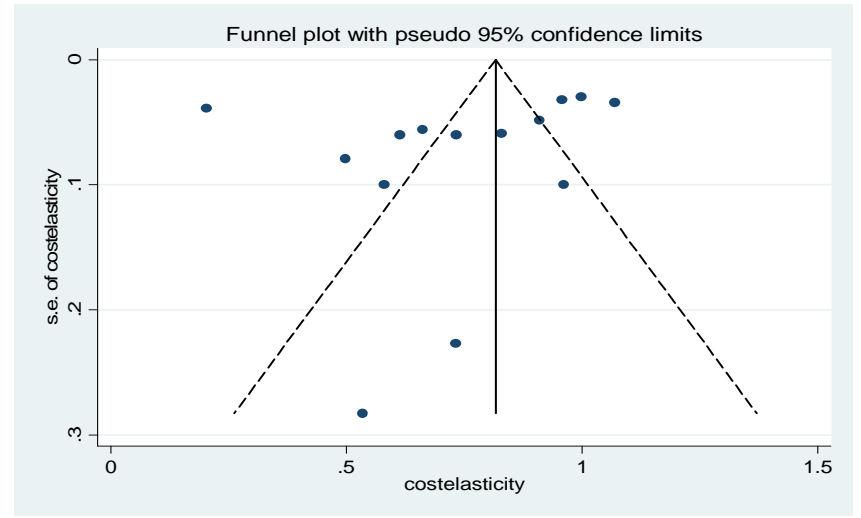


Figure 5. Funnel Plot Analysis.

The remaining 33 observations in the sample with standard errors reported included 17 observations on the garbage sector, 8 in the water supply and sanitation and 3 in the area of urban transport. The studies' unit of analysis is mostly jurisdictional units (23), 22 of them use cross sectional data, 12 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 change significantly 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 so called "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 costs. Chart 3 above 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 that 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 gather in the interval 0.5 to 1.

In terms of the sectoral distribution of observations, it would seem that the results from studies on the economies of scale of education services may be 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. These four sectors account for 55 out of the 60 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 one observation each for garbage collection and water supply report such levels of elasticity.

Chart 5. Observations on the Education sector

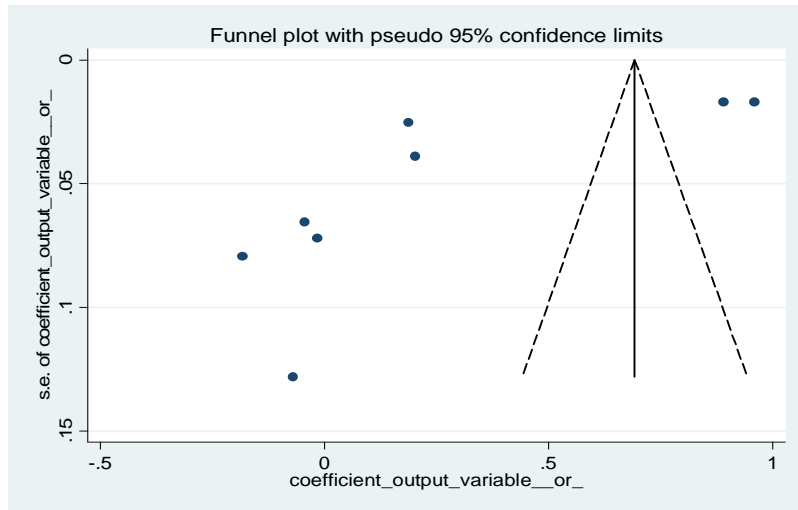


Chart 6. Observations on the Garbage Collection sector

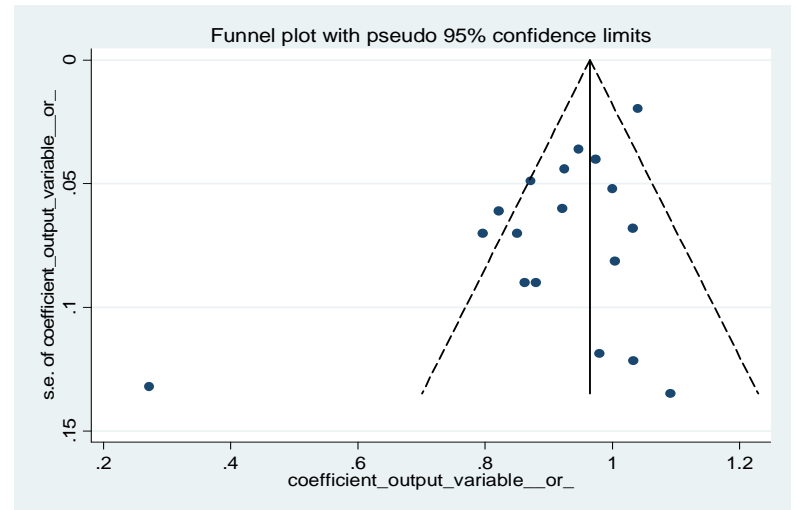


Chart 7. Observations on the Water Supply and Sanitation

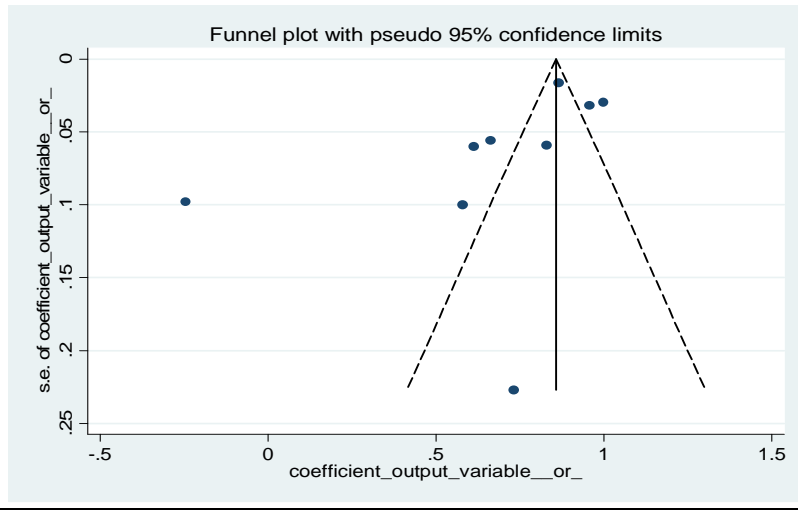


Chart 8. Observations on the Urban Transportation sector

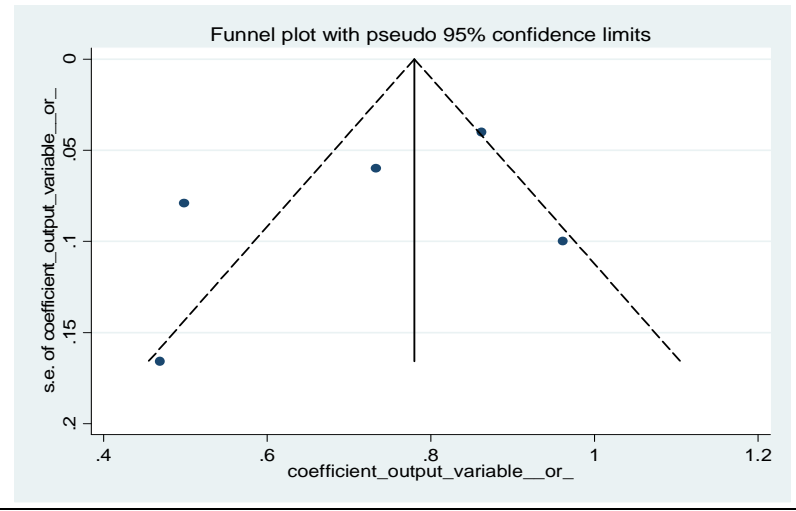


Figure 6. Funnel Plot Analysis – Sectoral Distribution.

Chart 6 above, 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 particular 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 close to 0, signaling sizable economies of scale. Among those observations, a log-linear form for the estimation of the cost function was primarily used. This may have important implications for our meta-regression analysis to which we turn next.

2.5 *Meta- regression analysis*

a. Testing for Publication Bias

In this section we test for publication bias estimating the equation;

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

First, we test for publication bias under the assumption of homogeneity among studies. Thus, we do not include in our meta-regression any of the moderator variables identified during the coding process. 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. The results are presented in Tables 2 and 3 below.

Table 3. Fixed Effects - Publication bias analysis.

	No PB	Linear PB		Non Linear PB	
		Uni-directional	Bidirectional	Uni-directional	Bidirectional
Cost Elasticity of Output	.713*** (.121)	.717*** (.142)	.723*** (.139)	.715*** (.125)	.721*** (.125)
α		-2.996 (14.94)		-17.425 (36.859)	
α_p			11.071 (16.003)		-3.381 (65.689)
α_n			-118.491*** (31.874)		-6129.188** (2947.902)
R2	0.762	0.762	0.793	0.762	0.773
N	44	44	44	44	44
Q-test of Heterogeneity (d.f. 43)	4806.57***				
I ² – Variation due to “between study” heterogeneity	99.1%				

Table 4. Random Effects - Publication bias analysis.

	No PB	Linear PB			Non Linear PB		
		Uni-directional	Bidirectional	Symmetric	Uni-directional	Bidirectional	Symmetric
Cost Elasticity of Output	.696*** (.058)	.778*** (.103)	.813*** (.071)		.721*** (.069)	.764*** (.053)	
α		-1.157 (1.200)			-3.469 (4.788)		
α_p			-.164 (.881)			-7.943 (16.348)	
α_n			-10.346*** (1.532)			-658.829*** (145.934)	
I ² Residual	0.987	0.9913	0.9859		0.9913	0.9908	
N	44	44	44		44	44	
F-test of symmetry in PB: $\alpha_p = -\alpha_n$							

In Table 2 we present the fixed effects estimates of the model assuming homogeneity between studies. Our estimate of cost elasticity of output under this assumption is somewhat smaller than the simple average for all estimates. From the results, it would seem that 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 observed from the Q-test¹¹. In addition, the I^2 test shows that 99.1% of the heterogeneity found in the sample is due to “between-study” differences.

Due to the large heterogeneity among studies (i.e. several sectors, modeling frameworks, etc.), random effects estimation is recommended, and results are offered in Table 3. The estimates reported in Table 3 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.

The estimations presented in Tables 2 and 3 however explain a small amount of the between study variation (an average of 30%). We therefore turn to the analysis of the case where we allow for observed heterogeneity between studies with the insertion of moderator variables in the meta-regression. In line with (Harbord & Higgins, 2008; p.497) and our earlier results, we do not estimate the fixed effects meta-regression, as

¹¹ 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, Marin-Martinez, Sanchez-Meca, & Botella, 2006).

such estimation assumes that “all heterogeneity can be explained by the covariates”, leading to excessive type I errors in cases (as ours) or 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, estimated by the algorithm beforehand.

In the first column of Table 4 we present the random effects estimation of the model with observed heterogeneity. The first column in Table 4 presents estimates without testing for publication bias, while the second and third columns 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 (column 2 of Table 4), it would seem that such estimation presents the better fit for the model. The coefficient on the constant variable, the conditional mean of the sample, is 1.026, signaling constant economies of scale for works published in the 2000s, in sectors other than education and garbage collection, that used predominantly the Cobb-Douglas form approach to the cost production function, used cost data as opposed to expenditure figures and population as a proxy for output. The non-linear publication bias test yields a conditional mean of the cost elasticity of output of 0.829, signaling some 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 2000s as reference group, estimates of economies of scale seem to have been greater in the 1980s (a period in which the sophistication of the analyses increases considerably with the generalization of log linear function estimations and the first contributions using translog cost functions). This effect however falls considerably later on the estimates of the cost elasticity of output are similar in size between the 1990s and the 2000s. Despite the robustness in the sign of this relationship, the results are not statistically significant. We thus find no solid support for the hypothesis that modern production methods, enjoying “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	.710*** (.259)	1.026*** (.239)	.829*** (.250)
1970s	.175 (.122)	.067 (.106)	.123 (.117)
1980s	-.319 (.220)	-.305 (.186)	-.326 (.207)
1990s	-.071 (.145)	-.084 (.124)	-.098 (.138)
Education	-.888*** (.223)	-.719*** (.202)	-.811*** (.216)
Garbage Collection	.328* (.166)	.247* (.144)	.285* (.157)
Translog Cost Function	.397** (.151)	.243* (.139)	.328** (.145)
Expenditure Data	.547*** (.148)	.434*** (.129)	.519*** (.140)
Physical Output	-.305** (.145)	-.285** (.126)	-.285** (.137)
Baumol	-.197 (.170)	-.216 (.145)	-.206 (.162)
Multiple Observations	.269* (.148)	.170 (.128)	.211 (.142)
α_p		-1.504* (.831)	-16.6 (14.4)
α_n		-6.16*** (1.598)	-285.09** (127.3)
Number of Observations	44	44	44
Residual Variation due to Heterogeneity	96.74%	96.87	96.87%
Proportion of Between Study Variance Explained	60.95%	73.51%	66%
τ^2 (Between study Variance)	0.056	0.038	0.049
Prob > F	0.0000	0.0000	0.0000

Our initial hypotheses regarding sectoral distribution of economies of scale are 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 and collection sector (an assumed capital intensive sector) displayed lower higher cost elasticities of output, leading to smaller economies of scale although the results were not statistically significant. Results for the water and sanitation sector were not found significant either in alternative model specifications.

Greater sophistication in the modeling of production costs leads to smaller estimates of economies of scale. The use of the translog function has in the literature, all things equal, led to higher estimates of cost elasticity of output. Acknowledging that the translog functions offers substantial advantages in the study of economies of scale, we must 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.

The bias introduced in the analysis of economies of scale by the use of inadequate measures of output is made obvious from the results. The use of expenditure data, as opposed to cost data, increases substantially the estimates of cost elasticity of output, thus

reducing the perceived potential of economies of scale. As 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 proved 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 showed no impact from the use of different definitions of economies of scale (i.e. Baumol or Caves), which is a positive sign of the consistency of our sample and offers some relief as to the possible distortion introduced by the heterogeneity in the measurement of our dependent variable. The results 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 variables were included in earlier model specifications but were found not significant. Country dummies were consistently found not significant. Their introduction as moderators in some cases was even pernicious as they could display high correlation with other moderators (for instance, 9 of the 10 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. Equally non significant

proved to be the variables identifying the structure of the dataset used in the study (i.e. cross-sections, panel, etc.), and the estimation method (i.e. OLS, SUR, etc.). Among other reasons, high correlation was found as expected 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. The results presented in Table 4 were robust to the inclusion of these variables.

It would seem that our results are greatly determined by the studies obtained from the area of education. From Table 1 we can observe that the mean cost elasticity of output in the sector of education is the lowest among the three main sectoral groups of observations. From Chart 5 in Figure 2 above we also notice that most of the observations reporting negative, or low cost elasticity of output are obtained from studies in the education sector. Thus, in order to test the robustness of our results to sectoral patterns, we estimate alternative model specifications excluding alternatively each of the three sectors for which the largest number of observations is obtained (i.e. education, garbage collection and water and sanitation). For simplicity, only selected variables are reported.

Table 6. Sectoral Disaggregation – Random Effects Estimation

	(1) Without Education	(2) Without Water and Sanitation	(3) Without Garbage
Constant	.631** (.256)	.645** (.246)	-.648 (.448)
(1) Garbage Collection (2) Education (3) Water and Sanit.	.553*** (.182)	-1.171*** (.150)	.021 (.192)
Translog Cost Function	.468*** (.162)	.442*** (.130)	.440** (.189)
Expenditure Data	.383** (.167)	.623*** (.124)	.457** (.224)
Physical Output	-.155 (.148)	-.389*** (.113)	.617*** (.202)
Baumol	-.078 (.169)	.095 (.190)	-.231 (.203)
Multiple Observations	-.070 (.179)	.492*** (.119)	.612** (.227)
Number of Observations	48	43	42
Residual Variation due to Heterogeneity	97.84%	84.15%	89.97%
Proportion of Between Study Variance Explained	39.57%	78.29%	58.86%
τ^2 (Between study Variance)	.048	.033	.073

As we can see, the significance and robustness of the results obtained in Table 4 do not seem to be affected by any particular sector. The exclusion of, alternatively, education, water and sanitation and garbage from the estimations changes substantially the constant and sectoral coefficients, but other determinants of the results remain significant. The use of the translog function continues to determine results and leads to lower estimates of economies of scale, as it is the case with the use of expenditure over cost data. As expected, variation is wider when the coefficient on physical output area considered as we alternatively eliminate sectors from the analysis, as the variable is substantially more correlated with the sectoral studies (i.e. education studies use population as their measure of output).

In order to provide a more complete picture, our estimation analysis also included initially the estimation of sector-specific meta-regressions for those sectors with sufficient observations (garbage and water supply and sanitation). As expected however, little variation is found among the most critical moderators within a particular sector, which turns the analysis worthless in our particular context. As our main goal is to estimate the potential for economies of scale in the case of jurisdictional consolidation, sector specific results were not critical to our conclusions.

2.6 Conclusions

Jurisdictional consolidation policies across the world have been justified by invoking the existence of economies of scale in the delivery of local public services that would allow

reducing average costs of production for larger “client” bases. Our first review of the empirical literature shows however that evidence of such economies of scale is inconclusive, with multiple studies reporting constant or decreasing returns to scale in a variety of services. Our meta-analysis systematizes the wide range of empirical approaches and modeling frameworks found in this literature and helps identify the study 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, being highest for education and lowest for garbage collection. The analysis also offers guidelines as to future empirical endeavors in this area. Physical output and production cost data should be used, together with translog approximations to 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 as to the extent and length of those economies of scale. However, it is unlikely that, in the context of U-shaped long average cost functions, such economies of scale will be pervasive well beyond the average production point. From the point of view of jurisdictional consolidation and assuming a situation of population growth, consolidation that brings production levels near the lowest long term average cost levels should expect

the gains from economies of scale to slowly disappear assuming no changes to production technology in the short term.

CHAPTER 3.

JURISDICTIONAL FRAGMENTATION AND ITS DETERMINANTS:

A CROSS-SECTIONAL INTERNATIONAL STUDY

“The need for more counties had a pragmatic value in the early days of the commonwealth. People lived in isolated communities or on farms far removed from the county seats. (...) Citizen’s naturally wanted to be at least within a day’s ride of their country seat.” (Commonwealth of Kentucky, 2005)

3.1 Introduction

Considering the limited (and service-specific) evidence on economies of scale on local service delivery available from the empirical literature, it is surprising to observe how often programs of jurisdictional consolidation across the world have been justified upon that assumption. As it was perhaps to be expected, such forced consolidation programs have yielded mixed results (Dollery & Robotti, 2009). The international experience seems to show that countries around the world are opting instead for adapting to high levels of jurisdictional fragmentation while encouraging the creation of new institutional modes for service delivery (e.g. special districts, inter-municipal cooperation, home owners associations, etc.).

The observed proliferation of new collective decision making mechanisms for local service delivery may be an indirect reflection of the limited size and specificity of the available economies of scale. It may also be symptomatic of the fact that, even when

economies of scale can be obtained, a preference for political accountability, as discussed in the theoretical framework outlined in Chapter 1, leads to a smaller (i.e. less than optimal from an economies of scale perspective) average size of local governments. The latter concept of political accountability is defined as the voters' capacity to influence the election (or re-election) of local representatives, and presumably this capacity is greater in smaller jurisdictions.

As pointed out, forced consolidation has often failed to live up to the efficiency gains expected from the start of those policy decisions. Experiences in some countries (e.g., Canada) show that total government costs after consolidation may actually have increased, at least in the short term (Sancton, 2008). Several reasons are behind those results. For a start, it is difficult to retrench public servants due to labor agreements, which would lead to duplication of staffing in the new, consolidated jurisdiction. Secondly, once the municipalities are consolidated, it is generally the case that the salaries of the public servants are brought up (or equalized) to the higher level observed pre-merger in the merging municipalities (Deller, 1998), which would further increase overall labor costs. In any case it would appear that the relationship between fragmentation and inefficiency is not yet fully established (Dowding & Mergoupis, 2003).

In addition, municipal or regional jurisdictional mergers can face considerable political opposition and institutional friction, even if one could reasonably expect efficiency gains from the merger and larger jurisdictional size. This may explain why, for example in Peru, a 2008 recent law on municipal mergers, which provides large fiscal incentives to consolidation in the form of grants to the jurisdictions merged over a period of 15 years,

has not been able to usher a single merger process. In another example, in Spain, a highly fragmented country with over 8,000 municipal governments, jurisdictional consolidation has also encountered strong political opposition at the local level, even when no distinguishable preferences over service delivery are observed among the merging municipalities or even when good evidence of potential for economies of scale exist (Bosch & Suarez-Pandiello, 2008).

The cases of Peru, Spain, and perhaps other countries, offer anecdotal evidence that forces other than the traditional economic theory postulates might be influencing the determination of the level of jurisdictional fragmentation in a country. To be sure, these other causes are likely to be complex and likely also to be intertwined, with reasons ranging from simple political turf and selfishness of local politicians, to more serious issues of representation and accountability and heterogeneity. Advancing in this knowledge is the main theme of this chapter and to do so from an international perspective. To our knowledge, to date, there does not exist a rigorous study analyzing the cross-country determinants of fragmentation in the way this issue has been previously analyzed for some particular countries.

Our challenge in this chapter is to search for common determinants of fragmentation across countries that may help better explain variations in the vertical structure of government in terms of the numbers of tiers of governments (i.e. the number of sub-national level governments in a country) and the number of jurisdictions in each tier (i.e. total number and also relative to population and area). At the outset, the country fixed effects of such an analysis are expected to be large and powerful, but not necessarily overwhelming. In addition to measuring fragmentation using the traditional indicators in

the literature (i.e. number of governments relative to population and surface), we will evaluate a dimension of fragmentation that has gone relatively unexplained: the number of tiers of government in a given country.

The levels of jurisdictional fragmentation around the world, as it is to be expected, vary widely. In Chart 3.1 below we show the number of sub-national levels of government (in addition to the central government tier). As it can be observed, 10 countries in our sample of 197 report having four tiers of sub-national level governments, while 50 report three. More than 50 percent of the countries in our sample have two sub-national levels of government, including countries vastly different in terms of population, ethnic composition, etc.

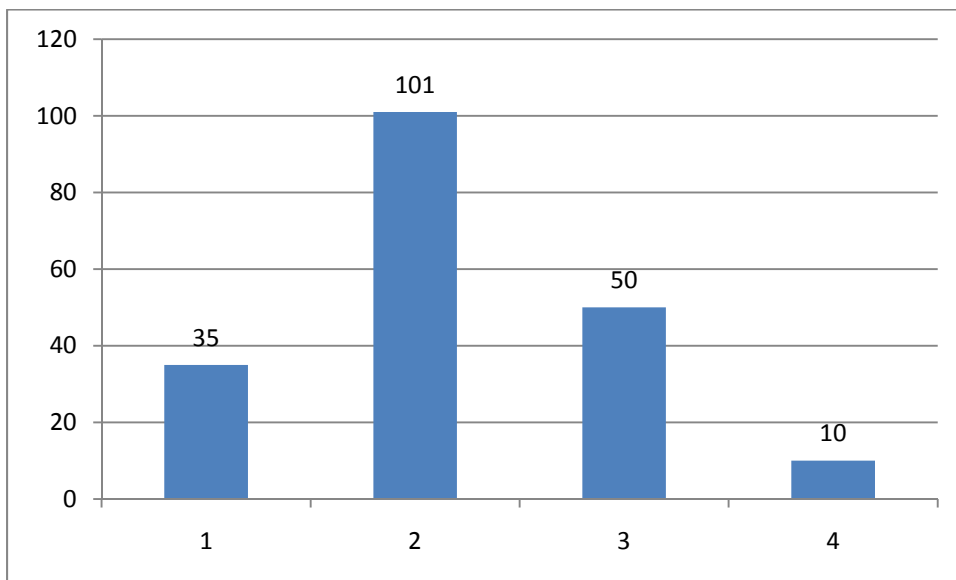


Figure 7. Sub-national Government tiers per country.

Equally, the absolute number of sub-national jurisdictions per country ranges widely in our sample, from a minimum of 0 in Kiribati (only a central level government) to a maximum of over 240,000 in India (including the Gram panchayat level of government). The median value is situated at 194. Chart 3.2 offers additional information on the distribution of the sample in terms of the absolute number of jurisdictions.

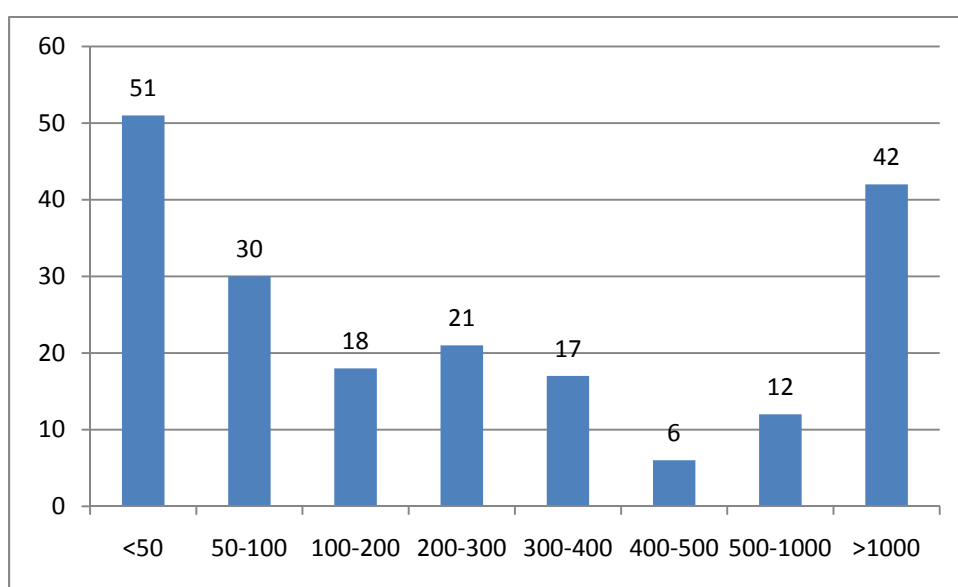


Figure 8. Number of sub-national jurisdictions per country.

A total of 42 countries in our sample report more than one thousand sub-national jurisdictions including all sub-national tiers of government. A similar number of countries (51) report less than 50 jurisdictions in total. As it is suggested by the correlations presented in Table 3.1, the total number of jurisdictions is highly correlated to the country's population, not so much with the geographical area. Overall levels of

income inequality, GDP pc and human development do not seem to be correlated with the level of jurisdictional fragmentation.

Table 7. Correlations: Number of jurisdictions and selected sample variables.

	Number Jurisdictions	Population	Area	Gini Coefficient	GDP pc	HDI
Number of Jurisdictions	1					
Population	0.74	1				
Area	0.25	0.44	1			
Gini	-0.04	0.02	0.09	1		
GDP pc	-0.02	-0.06	0.08	-0.48	1	
Human Development Index	-0.01	-0.03	0.09	-0.41	0.76	1

From our theoretical model we have concluded that Oates' (1972) traditional model of optimal jurisdictional size would seem to predict larger average jurisdictional size (and thus lower fragmentation) than our extended model which adds preferences for political accountability to the utility function of individual residents.

But before we get into the chapter, we must recall that Oates' model does not predict government size, but *group size* (i.e. in a way Oates' model equates population with jurisdiction size). In this sense Oates' model may be misleading when Oates (1972) identifies groups with local governments (or municipalities), and later on argues that economies of scale in production can be achieved via other institutional arrangements (service contracts, cooperation, etc.). Moreover, Oates' model refers exclusively to the cost (in efficiency terms) of collective decision making among groups, and it does not

incorporate the cost of transactions from inter-government coordination or cooperation. Thus Oates's model is helpful in that it outlines some of the most important tradeoffs of the process of assignment of functions across government levels, but it does not explain the observed level of government fragmentation since it does not deal with governments, but with groups.

The question we still want to look at is thus: what are the determinants of the current level of fragmentation observed across countries and how does this level relate to the predictions of our standard model? The relevance of this analysis lies in the often-cited inefficient size of local governments, especially in developing countries. But once we better understand what may be behind the level of fragmentation, would we remain surprised about the current status quo with the observed local government fragmentation across so many countries?

The rest of this chapter is organized as follows. In section 2 we offer a review of the literature on government fragmentation, while section 3 outlines the empirical models proposed to approach the analysis of jurisdictional fragmentation. Section 4 presents the results from our estimations and concluding comments are offered in section 5.

3.2 A review of the literature on government fragmentation and its implications

The level of government fragmentation has important implications for an array of issues. For instance, from a public choice perspective it is argued that more efficient governments (essentially defined as smaller government budgets) are expected from increased competition among local governments (Niskanen, 1975; Brennan & Buchanan, 1980). There is some empirical evidence supporting this hypothesis. In particular, Sjoquist (1982) finds that expenditures per capita in the southern metropolitan regions of the USA decrease as the number of jurisdictions increases. Pressure to reduce government spending seemed to be behind recent drives for municipal amalgamation in Canada as well (Sancton, 2008).

The theoretical foundations for the potential benefits to be derived from sub-national government fragmentation can be traced back to Tiebout's (1956) seminal proposition that citizens “vote with their feet” and chose the mix of services and taxes that best represent their preferences. Assuming adequate supply of local governments (which assumes away the problem of fragmentation but it implies it), equilibrium is reached when each citizen maximizes the welfare obtained from the mix of goods and services provided by the local government, net of tax payments. Thus, a sufficient number of governments offering distinguishable tax and expenditure packages would be required for this efficient sorting of individuals. From this point of view, certain level of fragmentation would be efficiency-enhancing, since a diversity of preferences requires a diversity of government service packages.

Empirical evidence on this issue is, again, mixed. Feld, Kirchgässner, & Schaltegger (2003) find no evidence that fragmentation affects the size (in terms of expenditure budgets) of Swiss cantons. In fact, earlier contributions of the literature have tended to suggest that augmenting the number of governments may lead to increases in the overall size of the public sector. This is likely to occur especially if economies of scale in public administration are foregone with greater decentralization (Oates, 1991), or if citizens' control over local bureaucracies translates into a wider range of services being provided (Zax, 1989). Among other reasons, this "larger public sector" results have been justified under the assumption of a lower quality of bureaucrats at the local government level, leading to poorer expenditure management and higher costs for services (Prud'homme, 1995). On the other hand, a recent cross-country study by Martinez-Vazquez & Yao (2009) argues that the total public sector employment grows with fiscal decentralization, as increases in employment at the sub-national level more than offset declines at the central government level.

From an equity point of view, higher fragmentation has been suggested to lead to the separation of tax bases among rich and poor. Richer areas (such as suburban residential areas) have a clear incentive to separate from impoverished urban areas that are subsidized with the revenues collected from their tax bases. The end result of such motivated fragmentation may be higher inequalities in the provision of services. A highly fragmented local government level may also lead to more self-centered governments in term of policy issues; reducing the incentives for coordination among jurisdictions to maximize overall welfare levels (Lewis, 1998). Earlier contributions seem to show, however, that higher income voters may be supportive of redistributive programs if they

expect to obtain long term utility gains (such as overall increases in property values in the jurisdiction) (Jorge Martinez-Vazquez, 1981).

The concept of fragmentation is not easily operationalized. The literature has favored a measure of the number of local governments relative to the size of the territory (either in terms of area or population) (Nelson, 1990). However, the range of overlapping authorities over the same territory (as ultimately, the institutional form chosen to deliver the public good or service may take several forms) adds an extra complication to the estimation of local government fragmentation levels. Optimal government size, Ostrom, Tiebout, & Warren (1999) argued, may depend on the nature of the public good provided and not just on the preferences of local consumers, leading to “polycentrism”, or a system of multiple collective decision-making mechanisms. Socio-spatial studies have termed this process “state rescaling”, a redefinition of the role of government at all levels, including the involvement of non-governmental agencies in public service delivery, the allocation of further competencies through decentralization to sub-national levels of government and the assumption of former national competencies by supra-national institutions. In line with this argument, Martinez-Vazquez & Timofeev (2008) find that the trend towards decentralization observed over the last decades has not affected, on average, the size of government, and offer empirical evidence of significant and positive relationships between globalization, decentralization and government size.

Whereas there is a shortage of contributions exploring the determinants of the observed high fragmentation of local government across the world,¹² the literature is rich in

¹² In addition, most of them are U.S. based, with some experiences from Canada and Australia as well.

contributions exploring the effect of fragmentation on economic growth, urban sprawl and other topics of economic interest. Nelson & Foster (1999) argue that in the framework of metropolitan areas in the US, as decision making becomes more fragmented (more local governments are present), growth in personal income declines. This is related to the fact that annexation of suburban areas is largely behind the increase in income growth of metropolitan areas, which are favored by their consolidation with affluent suburbs. Measures of sprawl used in the literature, very concentrated in the U.S. experience on the other hand, include population density (as an indication of low-density development), the percentage of dwellings in single-unit detached houses, or housing units per square kilometer. In one of the few studies that analyses the plausible reverse causation between fragmentation and urban sprawl, Razin & Rosentraub (2000) conclude that residential sprawl impacts positively on fragmentation, but fragmentation does not have a predictable effect on sprawl. Important results from their analysis include the confirmation that higher population density is a significant predictor of lower fragmentation, and that the age of the metropolitan area is a prominent variable predicting fragmentation¹³.

Empirical analyses of the level of fragmentation of the vertical structure of government are scarce perhaps due also to the difficulties in operationalizing our variable of interest..

¹³ Discussion on the causality for this effect is however insufficient in Razin and Rosentraub's analysis. Their age variable measures metropolitan areas that were already large in 1940, and thus less affected by the impact of the widespread use of automobiles on urban planning. Denser areas however would be expected to be less fragmented in light of previous findings; whereas Razin and Rosentraub find that older (and thus denser in their context) areas are more fragmented.

However, this topic has not gone unattended in the literature. We review below the available evidence on the determinants of government fragmentation.

a. Do heterogeneous local preferences lead to greater fragmentation?

As discussed in the previous chapter, evidence in support of the economies of scale that would suggest the need for larger size governments is hard to find. Can we assume that the other element of the tradeoff proposed by Oates (1972), that efficiency gains are attained from tailoring local public good provision to local preferences, is granted? If so, are heterogeneous preferences behind the observed fragmentation of local government?

The measurement of preferences has traditionally been more difficult than the measurement of costs. However, there have been some successes in the attempt to operationalize the heterogeneity of preferences in the economic literature. In his seminal contribution to the study of the determinants of government fragmentation in the U.S.A., Nelson (1990) shows that the number of jurisdictions is positively correlated with the degree of heterogeneity of individual preferences, measured by income dispersion and age dispersion (i.e. more income and age dispersion leads to a higher number of governments). This finding is well aligned with the decentralization postulates laid out by Oates (1972). If the efficiency gains from better tailoring public good provision to smaller groups make up for the loss of economies of scale, then we would expect higher levels of group fragmentation (and thus smaller government sizes). Nelson's analysis uses all types of governments from the U.S. Census, including both elected governments and special districts. He finds that the latter empirical relationship is stronger in the case of special districts than for general purpose governments, and that it may lend support to

the standard argument that heterogeneity of local preferences leads to smaller jurisdictions. Contrary to his expectations, greater racial homogeneity (assumed to be a reflection of lower preference heterogeneity) leads in his analysis to higher fragmentation, a surprising result.

Stansel (2005) finds a direct relationship between the level of fragmentation and economic efficiency; he finds a positive relation between the number of local governments in the US relative to population, and the level of economic growth. To the extent that local growth is achieved by the attainment of production and allocative efficiency gains from greater decentralization, more heterogeneous preferences would lead to preferences for greater government fragmentation. However, the theoretical linkages between economic growth and the decentralization of expenditure still await more conclusive empirical evidence (see Martinez-Vazquez & McNab, 2003). These results would be in line also with Musso (2001) who, using a sample of Californian cities, finds that more affluent communities (with income as a proxy for diverse preferences) in fast-growing counties are more likely to form new cities. Conversely, Burns (1994) argues that changes in the level of access and quality of local services do not explain local government formation, but instead tax avoidance and racial exclusion are found to be the most significant determinants. In line with Burns, Martinez-Vazquez, Rider, & Walker (1997) find that increasing racial heterogeneity of a state population increases the number of school districts, supposedly in order to satisfy their preference for disassociation.

A different line of work, important to our analysis, has linked heterogeneity in preferences to further decentralization. Shelton (2007), using data from a wide sample of

countries finds that heterogeneity in preferences (measured by an index of ethno-linguistic fractionalization) leads to further decentralization, measured as the share of local government expenditure in total government size. Decentralization is greater, Shelton shows, in the education and health sector, where one is more likely to find higher fragmentation of preferences due to social and demographic factors. In any case, greater decentralization (especially measured as the share of local governments on total public expenditure) does not necessarily translate into greater fragmentation of local government.

b. Economies of scope and demands for local accountability

The second major trade-off critical in the definition of the optimal jurisdictional size is that between administrative costs and local government accountability.¹⁴ If economies of scope are present, the joint delivery of, for instance, solid waste collection and water services by a supra-local level of government offers cost savings over their separate provision by two or more local governments. Evidence of economies of scope exists, especially for private sector production, but it is scant with reference to local government production processes. Sharing production inputs was shown to be a source of scope economies in the health care services sector (Grosskopf, Margaritis, & Valdmanis, 1995; Dollery & Fleming, 2006). Equally, Callan & Thomas (2001), in their estimation of a multi-product cost function for municipal waste services, find evidence of significant cost

¹⁴ It is argued that local government production functions present economies of scope, that is, that the output from the joint production of local public goods or services is greater than the output obtained with two separate processes, using the same amount of input (Panzar & Willig, 1977).

savings from the joint provision of recycling and disposal services; whereas Wolff (2004) argues that substantial economies of scope are found in the integrated management of River Basin systems in the US, as opposed to functional specialization. Although scarce, this empirical evidence extends across very different local government services, providing some ground to expect lower fragmentation of government due to the cost savings offered by joint production.

Can demand for greater political accountability lead to higher fragmentation? Several dimensions of the concept of political accountability need to be distinguished. As discussed in Chapter 1, we may define political accountability as the voter's capacity to influence the election (and actions) of their local representatives (Seabright, 1996). Such ability is expected to affect the political responsiveness to local preferences in the mix of public goods and services provided. This responsiveness is largely dependent on fiscal authority aspects such as whether powers to spend and tax have been devolved to sub-national governments, and most importantly on whether elections are being held at the sub-national tiers of government and on the quality of the election systems. Thus, for heterogeneous preferences to lead to distinguishable mixes of public goods provided across jurisdictions, certain institutional foundations must be in place linking citizens with policy makers.

The expected greater responsiveness of representatives from smaller jurisdictions to the preferences of their constituents is not an aspect explored by Oates in its theory of optimal jurisdictional size. His model is anchored in a "direct democracy system",

whereby local citizens “vote” effectively on the level of public good to be provided. No political economy issues related to representative democracy are at play thus in Oates’ contribution.

If representative democracy models are the setting for public good provision, then the principal-agent aspects of the relationship between constituents and representatives come into play. In this regard, following Tommasi & Weinschelbaum (1999), there may be four channels through which smaller jurisdictions are able to exert improved control over their political representatives (and thus ensure their priorities are acted upon). First, in line with Olson's (1965) theory of collective action, smaller jurisdictions reduce the incentives to free-riding (which is rendered more visible in smaller groups). Second, the existence of yardstick competition introduces additional benchmarks for political performance not easy to manipulate by local representatives. Linked to this aspect, policy diffusion models (Berry & Berry, 1990) would argue that local experimentation and the diffusion of best practices across jurisdictions may also introduce incentives and benchmarks for government performance.

A third channel, most important to us, is the fact that physical proximity to local representatives allows easier access to them (i.e. reduced transaction costs) even when demands are not related to policy changes but simply to quick, expedite action on issues of citizens’ interest. County formation in several of the U.S. States, for instance, has been particularly determined by this. Historical records from Kentucky and Georgia for instance would show that county boundaries were drawn ensuring no citizen resided more

than a day's ride from the county seat. Considering the strong historical inertia of jurisdictional formation, constituents' proximity to local governments may have been the single most important determinant of jurisdictional fragmentation.

In a game theoretic setting, physical proximity increases the probability of interacting in multiple venues, which allows for opportunities for punishment (Tommasi & Weinschelbaum, 1999) and thus introduces incentives for politicians to comply to citizens demands¹⁵. It is important to distinguish among these possible determinants of jurisdictional fragmentation if we attempt to explain the mounting anecdotal evidence of citizens' resistance to jurisdictional consolidation in the presence of nearly identical preferences for public goods and certain potential for economies of scale.

These aspects may be more important than any perceived gains on the (statistical) importance of individual votes in smaller jurisdictions. Even in small constituencies, the probability that a single vote will be decisive is minute. In light of the small statistical significance of a single vote, a large literature has developed attempting to explain why individuals vote at all (especially considering the cost of voting is not negligible), the so-

¹⁵ A preference for proximity to institutions governing the management of common services may be behind, for example, the widespread creation of home-owner associations in the U.S. The latter are truly miniature local private governments providing services traditionally under the responsibility of county governments (such as water supply, garbage collection, etc.). Although the creation of such associations is due often to legal requirements to urban site development, they also arise spontaneously and in any case, have expanded enormously over the last 20 years, representing now nearly 20% of the American dwellers. Across the board, the services provided by these associations, and the characteristics of such services do not differ greatly from those previously offered by elected governments, and although differences can be found at the margin.

called “voting paradox”. Explanations have included a desire to maintain democracy (Downs, 1957), or even the exercise of a sense of duty (Riker & Ordeshook, 1968). The general conclusion of this literature is that despite the weight carried by the alternative explanations, the paradox remains largely unresolved (Blais & Young, 1999).

Such demand for political accountability may not even translate into greater citizen participation in government, even though proximity to representatives may reduce the cost of collecting information about local policies. Although direct citizen involvement in participatory planning, budgeting or evaluation process may not depend on the level of government analyzed, the use of citizen satisfaction surveys does seem to assist budget development processes (Jordan & Hackbart, 1999; Melkers & Willoughby, 2005), informing governments’ decision on new expenditure programs (Riverbank & Kelly, 2006). Along those lines, there seems to be substantial evidence from international surveys that residents show higher levels of satisfaction from the services received from local governments than those received from the central government (CIS, several years, or Dasgupta, Narayan, & Skoufias (2009) for the case of Indonesia).

The literature on performance budgeting provides an alternative avenue to explore how accountable and responsive local governments are to citizen’s preferences and whether size affects the level of accountability. By explicitly defining indicators and performance benchmarks in their budgets, governments not only change the technology of accountability, but may offer a critical instrument to evaluate how close government actions are to local preferences, even though one can argue as to whether that was a main

goal of the introduction of performance based budgeting processes. A recent review of the literature in this area shows that the widespread implementation of performance budgeting measures at different levels of government in the US does not seem to have translated into significant changes in inter or intra-sectoral allocations (Gomez & Willoughby, 2008). This may reflect a lack of flexibility in adjusting expenditure patterns to local preferences due to largely committed, politics-driven budgets. But, it could also be a reflection of how time-consuming it is for citizens to collect and process government budgetary information. In a pioneering study, Melkers & Willoughby (2001) found that budgetary officials at the State level were more positive about the impact of results-based budgeting on states' budgetary appropriations in the U.S. if the performance measurement was a requirement of the legislature (and not just of the executive).

There is no clear evidence that the impact of performance measurement on budget formulation is greater as we move down the government hierarchy. (Melkers & Willoughby, 2005; p.18) find in a recent contribution to this literature that "less than 50 percent of county and city officials agreed that performance measures were a vital decision aid on budgetary issues". They find that the use of performance measures is more common at the county than at the municipal level. We cannot conclude, however, that the latter would be an indication that smaller governments are less accountable to their citizens. In fact, this finding could support the opposite view that because smaller governments are more accountable to their citizens, there is less need for the implementation of performance measures.

In summary, a review of the literature suggests that the case for economies of scope is relatively well documented, although the largest contribution of the literature in this area is concentrated on private sector production processes. We are not able to establish at this point a connection between higher citizen participation and greater involvement in budgetary processes at lower levels of government and smaller size of jurisdictions; the literature in this area is in a very initial state of development.

c. Fragmentation versus equity in local government structures

In the absence of proper fiscal equalization mechanisms, fragmentation of the local government structure may make the equitable delivery of public services difficult due to the accompanying fragmentation of fiscal bases (Warner & Hefetz, 2002). On the other hand, government fragmentation may allow for better access of the poor and rural population to services tailored to their needs, thus making the system of government more equitable.

Academic contributions to this debate have focused on the analysis of metropolitan areas and especially on the fiscal comparison between central metropolitan and sub-urban areas. As Razin (2000; p.28) puts it, changes in local government organization “re-shape the rules of the game of local development and influence inter-local disparities”. On that line, Schneider (1986), in his study of metropolitan disparities in access to services, does not find support for the hypothesis that fragmentation results in higher service inequality.

Neither do Morgan & Mareschal (1999) who find no effect of metropolitan fragmentation on fiscal health.

Contributions in this area have focused on cross-sectional comparisons in expenditure per capita versus income per capita in the major metropolitan areas of the U.S. A general conclusion of this literature is that fiscal fragmentation may have been favored as a strategy to avoid inter-jurisdictional redistribution of local fiscal bases (Ellickson, 1971; Lewis, 1998).

d. Institutional, Demographic, and Geographical issues

Institutional aspects that may restrict or even encourage further fragmentation have been also the object of empirical analysis. Nelson (1990), again focusing on the U.S. case, finds that tax and expenditure limitations (TEs) increase the number of special districts, which may be created as a way to circumvent tax limits posed upon state and county government by their citizens. He does not find however evidence that other self-imposed limits to local debt, or home rule¹⁶ clauses have any significant impact on the number of governments. Contrary to Nelson's findings, Bowler & Donovan (2004; p.194) most recently qualified this finding, arguing that, "absent the pressure of ballot initiatives, TEs do not lead (in the U.S.) to an increase in the formation of new local governments". In the same vein, Lewis (1998) argues that the implementation of Proposition 13, which imposed severe limits to the rise of property taxes in California, did not lead to an

¹⁶ Home rule is defined in (Nelson, 1990) as the allocation of significant autonomy to local governments in carrying out local functions.

increase in the level of local government fragmentation in the form of additional special districts.

The overall size of the public sector may be an important determinant of the fragmentation of local government systems. This hypothesis has not been fully explored to date perhaps due to the endogeneity involved in its testing. For instance, Kenny & Schmidt (1994) find that state aid to school districts is an important determinant of the great consolidation of school districts observed between 1950 and 1980. They argue that a larger role of the state in the financing of schools reduces the capacity of districts to differentiate the education services provided and sort themselves on the basis of average income in their jurisdictions, so consolidation ensues.

Outside main economic arguments, jurisdictional fragmentation may also (and perhaps most importantly) be affected by a myriad of institutional reasons such as the form of the state (federal vs. unitary), the form of government systems (monarchy versus republic), a history of decentralized government or secession of certain regions, cultural and ethnic issues, civil or armed conflicts, and so on. More recently, and particularly with regard to urban municipal consolidation processes, the desire to compete in a “global city” environment may have prompted metropolitan consolidation processes¹⁷.

Demographic and geographical variables have traditionally been used as controls in the empirical analysis of fragmentation. Metro area population and land area are positively correlated and statistically significant determinants of the number of governments in Nelson (1990), a result partially supported in Bowler & Donovan (2004). The hypothesis

¹⁷ See, for example, the case of Toronto in Sancton (2004).

behind the inclusion of these variables is that large demographic or geographical areas may be more difficult to manage and thus fragmentation would be efficiency enhancing. Nelson's analysis includes additionally US-specific institutional explanatory variables not applicable to cross-country analysis.¹⁸

As discussed above, numerous factors may partly explain the level of government fragmentation encountered in a particular country. Our challenge is now to develop an empirical model that allows exploring the relative influence of the explanatory variables discussed, a task approached in the next section.

3.3 Empirical Model

In our theoretical model, we have argued the number of jurisdictions is a function of several factors, including importantly, the cost of production of services (and thus the potential for economies of scale), the heterogeneity of preferences, and preferences for political accountability, which, all things equal, led us to an equilibrium solution where the optimal jurisdictional size is smaller than in the case where the latter variable is not considered. In addition, we have discussed additional determinants of jurisdictional fragmentation, including institutional, geographical and demographic variables.

Generically, we can represent the relationship between optimal jurisdictional size and these set of relevant variables as:

¹⁸ These include whether a referendum or majority approval is required for a territory to be annexed to a city or for the consolidation of two or more jurisdictions.

$$N = N(C, U(DG), P, \bar{Z}),$$

where, using the notation from Chapter 1, N is the jurisdictional size, C represents a measure of economies of scale, $U(DG)$ depicts the level of heterogeneity of preferences as defined in Chapter 1, P represents preferences for political accountability and Z is a vector of variables including institutional, geographic and demographic factors affecting jurisdictional size.

As discussed, the expected sign of the partial derivatives is as follows:

$$\frac{\partial N}{\partial C} > 0, \frac{\partial N}{\partial U} < 0, \text{ and } \frac{\partial N}{\partial P} < 0$$

Our first task in the definition of our empirical model is to further clarify our dependent variable. As already noted above, operationalizing the level of government fragmentation is a complex task. The literature has favored absolute measures, such as the total number of governments, over scaling indicators, such as the average population or extension by jurisdiction. Conversely, the number of local governments for a certain population size could be used as a relative measure of fragmentation. Nelson (1990) uses in his study of U.S. government fragmentation the number of governments per metro area, while Bowler & Donovan (2004) use the number of governments (again both general and special-purpose governments) in a state at the time of a census. In exploring the effect on fragmentation of Proposition 13 in California (which importantly limited revenue collections from the property tax), Lewis (1998) also uses total number of governments as his variable of interest.

The use of absolute measures of government fragmentation (i.e. total number of governments) may be a reasonable empirical strategy when a certain level of jurisdictional homogeneity can be assumed within the sample, as in country case-studies. Our aim however is to explore these relationships within the context of as large as possible a sample of countries. This therefore requires a relative measure of fragmentation that helps homogenize to the extent possible the individual values of our dependent variable.

First, we should clarify what is understood as “government” or jurisdiction in this context. In short, we include in our calculations all levels of government with service delivery responsibilities and all jurisdictions within each level. The rationale for such a criterion is that we are interested in exploring whether heterogeneity of preferences for public services may lead to further fragmentation so as to allow a better matching of preferences with service delivery. Arguably, a country may be divided in a large number of jurisdictions, but in the absence of elected governments with some authority over their budgets, public service delivery may not differ much in that system from a fully centralized service delivery system. The level of authority over local expenditure and revenue sources is, however, very hard to measure. In fact, fiscal decentralization experts around the world have criticized this fact as a major obstacle in the analysis of the impact of fiscal decentralization on economic growth, or service delivery. Acknowledging that the level of fiscal autonomy of sub-national government around the world varies wildly, it is to be expected that a jurisdiction with a locally elected government will have larger

authority over its budget than a similar territorial unit with de-concentrated units of the central government. Our analysis will control accordingly for this.

Our estimation strategy is to try different specifications of the dependent variable. First, in line with previous contributions to the literature, we will use the total number of jurisdictions (from elected levels of government) as our dependent variable. However, as just discussed, the heterogeneity of jurisdictions and different sizes across the world requires the use of a relative indicator of fragmentation that allows homogenizing the values of our dependent variable. Thus, as derived from our theoretical model in Chapter 1, we will use as dependent variables the ratio of population and area size to the total number of jurisdictions in the subsequent model specifications. To that extent, all jurisdictions from elected levels of government (including the central government) are added, and the values of the above mentioned ratios are calculated for each country.

Lastly, we will also consider a rarely discussed dimension of fragmentation in the literature: the number of levels of government. As it is the case with other indicators of government fragmentation, the number of levels of government in a given country is likely to be correlated with the geographical size of the country and its population size. This is an aspect of fragmentation largely ignored in the literature, and thus we provide a first exploration into this issue.

3.3.1 Hypotheses

Economies of Scale and Fragmentation

In line with the standard model of optimal jurisdictional size, we would expect that the potential for economies of scale to be a deterrent to high government fragmentation. As we saw in Chapter 2, the literature has traditionally approximated the measurement of economies of scale in various ways. Earlier studies used population as a proxy for economies of scale, while more recent contributions have used either expenditure or production cost data. As the results from our meta-analysis show, population can be arguably a rough measure of potential economies of scale, leading to their underestimation.

In the context of our analysis of jurisdictional fragmentation, using population as our measure of economies of scale introduces additional problems. For instance, considering China and Luxemburg, we would have to expect (if population were our proxy for economies of scale) lower relative levels of government fragmentation in China than in Luxemburg, since China's large population would signal to the largest international potential for economies of scale. That is, obviously, not a reasonable assumption. Population would seem to be therefore naturally and directly related to the level of jurisdictional fragmentation, and it is discarded as a valid proxy for economies of scale. This approach is also consistent with Nelson's (1990) analysis of jurisdictional

fragmentation in metropolitan areas of the U.S., whereby population is included solely as an environmental, control variable.

From the meta-analysis conducted in Chapter 2, education, urban transportation, garbage collection and water and sanitation, in that order, were the services that displayed (from the studies analyzed) the largest potential for economies of scale. Ideally, one would want to explore the functional assignments of our sample of countries and ascertain in which of them are these functions decentralized to sub-national levels of government as an indication of potential for economies of scale. Regrettably, few countries have formal statements of expenditure assignments and even where they exist, they are often no more than broad lists of functional responsibilities with great overlap across tiers of government. In addition, collecting data on the cost elasticity of production from service delivery at sub-national levels of government for a sample of close to 200 countries is, simply, unrealistic.

We are left with few straight-forward options for variables approximating the potential for economies of scale in public service delivery. The most comprehensive analysis of the determinants of fragmentation in the U.S., that of Nelson (1990) does not include a measure of economies of scale, a clear signal of the difficulties encountered in operationalizing this variable even in the context of a single country case-study. The Global Competitiveness Index (Schwab, 2010) may offer a possible avenue. Among the index sub-components, a measure of technological readiness is included. As discussed, the sources of economies of scale are mostly technological in nature. They may originate

in gains from more efficient division of labor (internal economies of scale), or can be derived from an expansion in the industry where the company operates, leading to increased leverage power with suppliers. Lastly, they may also be due to the use of more specialized inputs of production. The latter technological dimension is measured by the GCI Technological readiness sub-index, which approaches the capacity of an economy to adopt new technologies in order to improve the productivity of the national private sector¹⁹. The use of this measure implies assuming that if such technologies are available to the private sector, it would be more likely that the public sector is also introducing them in its production processes, thus increasing their productivity²⁰. As a downside, the introduction of technological advances may allow cost reductions even in specialized production, thus making more efficient the tailoring of public good productions to local preferences.

Heterogeneous Preferences and Fragmentation

The second set of variables aims to test the hypothesis that heterogeneous preferences lead to greater jurisdictional fragmentation. This hypothesis is derived from Oates's standard postulate that greater preference heterogeneity leads to smaller optimal group size for service delivery. Heterogeneous preferences are operationalized with measures of income, race and age dispersion. In terms of income dispersion, higher values of the Gini

¹⁹ The measure is a composite one, including values on aspects such as availability of new technologies, firm-level technology absorption, legal ICT framework, FDI and technology transfer, mobile and internet coverage, and personal computers per capita.

²⁰ Alternatively, to be fair, the public sector may suffer from Baumol's disease, and witness rise in salaries in response to productivity increases in the private sector.

income inequality index should lead to preferences for territorially tailored social policies, and thus a higher the level of jurisdictional fragmentation where local governments can influence policy design and implementation.

Equally, a measure of ethnic dispersion (taking into consideration race, language, and religious dimensions) is used to approximate heterogeneous preferences derived from varying ethnic compositions.²¹ Greater ethnic diversity is also expected to translate into a more fragmented government system that allows reflecting the preferences of the minorities. First, we used a recent contribution by Alesina, Devleeschauwer, Easterly, Kurlat, & Wacziarg (2003), which provides recalculated measures of the Easterly & Levine (1997) data and disaggregated indexes of fractionalization, which are used in the estimations. In order to conduct a sensitivity analysis of results, checking the consistency of results with alternative measures, we also construct a Herfindahl ethnic and linguistic fractionalization index on the basis of data on ethnic composition collected from Population Statistics (see Table 3.1 below). Both indexes were highly correlated and their alternative inclusion on the model estimations rendered no significant differences.

We include an additional measure of heterogeneity of preferences as an index of age dispersion. The rationale behind it is that senior citizens are assumed to display well

²¹ Martinez-Vazquez et al. (1997) use the share of minorities over the total population in their analysis of the impact of race dispersion on school district consolidation.

differentiated sets of preferences (i.e. for a start, they may favor higher expenditures in health and lower expenditure in education as they do not have school-age children²²).

The fractionalization Herfindahl indexes (for both ethno linguistic and age fractionalization) used in this analysis have the form:

$$Fractionalization_j = 1 - \sum_{i=1}^N g_{ij}^2$$

with $0 \leq g_{ij} \leq 1$

where g_{ij} is the share of group i in country j in unit terms. Thus, the larger the value of the index, the larger is the fractionalization of the country, and thus we assume the more likely a fragmented system of government will be in place to cater to those heterogeneous preferences. At the other extreme, where there is only one group, and thus the share of the group over the total population is 1 (or 100 percent), the fractionalization index would be zero. Thus, positive signs are expected in both coefficients for the age and ethnic-fractionalization variables.

Measuring Political Accountability

Our variable of interest, political accountability, is operationalized in various forms. As discussed, the existence of elected government representatives at the sub-national levels

²² On the other hand, housing values are generally linked in the U.S. for instance to school quality. Or seniors may have grandchildren attending local schools, which would balance the set of preferences assumed by this indicator.

of government (as opposed to appointed by central government authorities) is a critical indicator of political accountability. Sub-national officials can be removed from office if policies are not reflective of the majority preferences, the very political link that is required for fiscal decentralization to succeed. However, the existence of sub-national elections is a necessary but not sufficient condition for accountability to be exercised. Several pseudo-democracies and even dictatorships around the world hold sub-national “elections”. In these countries, the range of candidates is severely limited to those of the party in power or political allies. No credible political alternatives are offered to citizens to choose from. To take account of the democratic status of countries around the world, we interact our dummy variable measuring the existence of sub-national elections with an index of legislative competitiveness from the Political Institutions Database (World Bank, 2006).

In addition, we include in our model the World Bank’s Governance Index sub-component of “voice and accountability”. This indicator (with a minimum value of 0 and a maximum of 5) aims to measure the “extent to which citizens of a country are able to participate in the selection of governments (Kaufmann, Kraay, & Mastruzzi, 2008)”. The governance indicators collected by the World Bank include an additional indicator, “Government Effectiveness”, which aims to measure the capacity of the local bureaucracy to provide quality public services, and its independence from political pressures. Both measures cover aspects of our variable of interest (political accountability) and offer the opportunity of reducing the endogeneity of the model by using them as a proxy for the latter. Arguably, the “voice and accountability” indicator represents more closely the

capacity to elect and exercise control of the citizenry over their political class, the very essence of accountability.

A third dimension of accountability that is explored is the authority of local representatives over tax collection and spending policies. It may be the case local authorities are elected but still budgets are “conditional” in their use so that no spending discretion is allowed. Additionally, local authorities may not enjoy revenue raising capacity at the margin, which also limits their ability to respond to local needs. We measure this dimension in two ways, first, including a variable measuring the share of total expenditure conducted at the sub-national level, a traditional indicator of fiscal decentralization. Second, we include a dummy variable with a value of 1 if the sub-national governments have authority over taxing, spending or legislating (World Bank, 2006). Unfortunately, this variable is only available for a limited number of countries, reducing the number of observations importantly.

Other variables that may importantly influence the level of jurisdictional fragmentation include the share of public expenditure out of GDP (for countries where government employment is the main source of family income in rural areas for instance). This variable presents however a clear problem of endogeneity, as more fragmented systems of government are more likely to show also a higher government share in GDP, and is therefore not used in our final model specification. Alternatively, the existence of constitutional provisions guaranteeing the provision of social services as a right of the citizenry (with obvious fiscal and governmental implications) may greatly affect the level

of public sector expenditure. Unfortunately, the latter variable could not be obtained as our search for constitutional texts did not render recent versions in English that could be coded for a large number of countries.

Additional Institutional Hypotheses

A wide array of institutional variables may affect the level of fragmentation of a country, as we have previously discussed. The available political institutions data offers several alternatives to test institutional hypothesis in this context. Arguably, the impact of institutional variables is likely to extend beyond any single year and as such, a cross-sectional analysis for a single year will not capture the full implications of maintaining or changing any critical aspect of the institutional fabric of a country. Our results in this area should therefore be taken with caution.

First, we test the impact of having a presidential, parliamentary or assembly elected president on jurisdictional fragmentation. Second, we test whether the existence of a nationalistic party in power (arguably an advocate of unitary systems) leads to lower levels of fragmentation²³. As additional tests to the quality of democracy, we introduce a variable measuring whether there is a constitutional limit on the number of years an executive can serve before elections are called. Ideally, following Nelson (1990), we would have liked to include in our analysis whether provisions exist for minimum

²³ As an example of the caveat introduced above, if a nationalistic party has just been elected, despite of its ideological leanings, it might not be able to affect the jurisdictional organization of the country unless it keeps in power for an extended period of time. From that point of view, the variable is likely not to be significant in a cross-sectional analysis.

population sizes for new jurisdiction creation. We were unable to collect this variable for a large number of countries where legislation is not easily accessible.

Finally, anecdotal evidence seems to suggest that movements from autocratic to democratic systems of government release pent up pressure for jurisdictional fragmentation. Indonesia may be a good example where the advent of democracy unleashed a process of fragmentation (*pemekaran*) that translated into an increase in the number of sub-national level jurisdictions in search for democratic spaces of representation. The end of armed conflicts may also lead to such a process of new jurisdiction formation. We measure these dimensions with a single dummy variable with value one when either a significant transition to a more democratic system of government (measured by an increase in the Polity IV project index of 3 or more points) occurred in the country over the last 10 years.

Geographic Control Variables

Arguably, country size is related to government fragmentation. Difficulties in reaching isolated populations from the political center are a strong incentive to create sub-national levels of government that can deliver public services. We use country area (in square kilometers) to measure country size. Additionally, a measure of geographic accessibility is required, as countries with small land area may still be highly fragmented if they are of mountainous territory or display other natural features that difficult accessibility. Such lack of accessibility would lead to a preference for smaller governments closer to the

citizens. Absent a better measure of accessibility, we use the ratio of the highest to the lowest altitude in a country as a proxy for accessibility. Ideally, we would like to use many altitude points in every country to evaluate their dispersion, but such information could not be obtained.

Alternative variables to consider in order to control for the way geographical characteristics may affect jurisdictional fragmentation could include the geographical position of a country, such as whether it is landlocked or coastal and thus we tested the impact of a dummy variable measuring this dimension. We also include a variable denoting the continent of location of the country and test for the impact of latitude, a variable that has proved highly significant in the analysis of economic growth patterns.

3.3.2 Estimation Methodology and Data

For our estimation strategy, we first approach the analysis of the determinants of the number of levels of government in a given country. Subsequently, we test the main hypothesis of model relative jurisdictional fragmentation across levels. The rationale for such an approach is the understanding that, although certain simultaneity in the process is obvious, presented with the opportunity to define the territorial structure of a country (for instance in the Constitution drafting processes), or at least to approach its reform, the logical sequential order of events is to first define the number of levels of government and, on that basis, determine the mechanisms for creation of new jurisdictions (or their merger) within each level. For example, the 1978 Spanish Constitution only defined the

different levels of government (regional, provincial, municipal). The Organic Law regulating this constitutional aspect established that in instances where regions are formed by just one province, the provincial level of government is sub-summed into the regional one as there is perfect overlap. The number of regions eventually created, and the number of municipalities existing currently in Spain is the result of an institutional process where most forces considered in this paper were at play.²⁴

As discussed in the previous section, the number of levels of government in a given country is a largely unattended dimension of jurisdictional fragmentation. By selecting the number of tiers of sub-national levels of government as the dependent variable, we will observe that while some countries have two tiers or even more, there are some other countries that have only one tier or none. Thus Tobit estimation appears to be an adequate estimation approach (Tobin 1950). Because the variable is censored at 0, in its standard mathematical expression, in terms of a latent variable y^* (Baum, 2006):

$$y_i^* = X_i' \beta + \mu$$

$$y_i = \begin{cases} 0 & \text{if } y_i^* \leq 0 \\ y_i^* & \text{if } y_i^* > 0 \end{cases} ,$$

where X_i' is a row vector of explanatory variables, β is a vector of parameters to be estimated, and $\mu \sim N(0,1)$ is the error term.

²⁴ To date for instance, the Constitution in Spain has not been amended to include the names of the regions in its articles.

In our sample, this variable either shows zeros for those countries with no sub-national levels of government or a positive integer between 1 and 4. Arguably, there are other alternatives to operationalize this aspect of jurisdictional fragmentation. For instance, we could have collapsed all positive values and turned this censored variable into a binary one, but at the cost of losing important information on the different number of tiers of sub-national governments. Alternatively, we could have counted all levels of government in the variable, including the central one, thus allowing for the use of an ordered probit model fit, another maximum likelihood estimation method. However, the cardinal nature of the variable suggests that ordered probit estimation is not the right approach²⁵. Lastly, truncated data estimation methods were not appropriate since the data generation process did not present this characteristic. We could perhaps convert the variable into a truncated one by discarding all observations with a value of zero (Baum, 2006), but again at the cost of degrees of freedom and loss of information.

Since this is the first exploration of the determinants of the number of sub-national levels of governments in a given country we maintain the same set of explanatory variables proposed in our theoretical framework, and therefore estimate the following equation:

$$\hat{L}_i = \alpha_0 + \beta_1 C_i + \beta_2 U_i + \beta_3 P_i + \sum_{j=1}^M \phi_j Z_{ji} + \mu_i$$

²⁵ We would expect that the number of sub-national tiers of government in a given country also depends on the number of tiers that already exist. It would become more difficult to add another tier if a relatively high number of sub-national levels of government already exists. Thus, the variable is non-ordinal in nature.

where L is the number of levels of government, C_i represents economies of scale, U_i reflects the impact of heterogeneous preferences on over all welfare, P_i is our measure of political accountability, and Z_{ji} is a vector of other institutional and geographical variables. The marginal effects on the observed value L will be equal to the individual coefficients times the probability that the latent variable value is between a certain interval. In practice, our interpretation of results will focus of the statistical significance of the coefficients and the direction (sign) of the relationships found.

Once we have explored the determinants of government fragmentation into different tiers, we turn to the analysis of the more traditional aspects of jurisdictional fragmentation. Summary statistics for the data used in this analysis are provided in Table 3.1 below. The data have been collected from a wide variety of sources. As the database used for this empirical part of the analysis is a cross-section on nearly 200 countries, OLS estimation will be applied to the model specifications using the total number of jurisdictions, the average population per jurisdiction, and the average area per jurisdiction as alternative dependent variables.

For simplicity, we will assume a linear functional form for this relationship and thus the equation to be estimated is:

$$\hat{N}_i = \alpha_0 + \beta_1 C_i + \beta_2 U_i + \beta_3 P_i + \sum_{j=1}^M \phi_j Z_{ji} + \mu_i$$

where N is the number of jurisdictions (or the average population or area per jurisdiction, depending on the specification), C_i represents economies of scale, U_i reflects the impact of heterogeneous preferences on overall welfare, P_i is our measure of political accountability, and Z_{ji} is a vector of other institutional and geographical variables.

At the outset, and considering the multiple avenues through which the levels of fragmentation of a country can be determined, our model may leave out, due to lack of available data, critical explanatory variables and thus suffer from omitted variable bias that leads to higher standard errors. To explore this aspect we will conduct standardized Ramsey tests. Moreover, the size of our sample (effectively around 143 observations) limits the number of variables that can be included in the analysis. However, regional dummy variables are introduced to absorb common variation in regional fragmentation patterns.

We had also initially assumed a linear form for our model, which may not accurately reflect the relationship between the variables. In order to test the linearity of the relationship hypothesized, we conducted graphical analysis, plotting the residuals against the suggested predictors in early model specifications and exploring any possible deviations from linearity. This analysis suggested the need to transform several of the variables into logarithmic form, and to opt for a quadratic form in the relationship between population and jurisdictional fragmentation.

Both the White and Breusch-Pagan tests showed a certain amount of heteroskedasticity as it was to be expected from the nature of our sample. Accordingly, robust standard errors were calculated to correct the heteroskedastic errors.

Table 8. Summary Statistics of the Dataset

Variable Name	Max	Min	Mean	Standard Deviation	Obs.	Description	Source	Year
Population	1,330,044,605	5,079	30,724,686	123,328,416	218	Total Country Population	International Database, US Census Bureau	2008
Domestic Market Size	6.93	1.3	3.78	1.18	132	Sum of GDP plus imports, minus exports, normalized on a 1-7 scale	World Economic Forum	2008
Techno Readiness	6.15	2.19	3.80	1.09	132	Composite Index of Technological Capacity	World Economic Forum	2008
Age014	50.00	13.40	28.89	10.45	218	Percentage of Population under 15 years of age	International Database, US Census Bureau	2008
Age1564	79.90	47.80	63.51	6.75	218	Percentage of Population between 15 and 64 years of age	International Database, US Census Bureau	2008
Age65+	22.80	0.90	7.60	5.11	218	Percentage of Population over 64 years of age	International Database, US Census Bureau	2008
AHV	0.67	0.46	0.51	0.03	218	Age Herfindahl Country Value		
AleELF	0.93	0	0.44	0.26	187	Ethno linguistic Fractionalization Herfindahl Index	Alesina et al (2003)	
AleLin	0.92	0.00	0.39	0.28	192	Language Fractionalization Index	Alesina et al (2003)	
AleRel	0.86	0.00	0.44	0.23	202	Religion Fractionalization Index	Alesina et al (2003)	
Country Area	16,995,800	2.00	597,974	1,774,385	218	Total Country Area in square kms	International Database, US Census Bureau	
Gini	74.61	24.00	40.88	9.91	151	Income/Expenditure Inequality	U.N. Wider Database	Several Years
ConIneq	1.00	0			151	Dummy variable with value 1 for Consumption inequality data		
GDPpc	85,382	298	13,259	15,704	182	Purchasing Power Parity GDP Data	United Nations HDR	2007
VAI	4.07	0.19	2.49	1.00	205	Voice and Accountability Index	World Bank	2007
GEI	4.91	0.15	2.48	1.00	209	Government Effectiveness Index	World Bank	2007
SubDem Status	2	0	1.1	0.882	190	Interaction term; Sub-national Elections*Democratic Status (0=Dictatorship, 1=Pseudo Democracy, 2=Electoral Democracy)	Own Calculations	2009
SubDem Status2	0	7	5.08	2.93	171	Interaction Term: Sub-National Elections *Index of Legislative Competitiveness (0=No legislature – 7=Multiple parties got seats in Parliament)	World Bank and Own Calculations	2006

Table 8. Summary Statistics of the Dataset (Continued)

Finite Term	1	0	0.894	0.308	170	Existence of a Finite Term for the Legislature	World Bank	2006
Change Polity4	1	0	0.21	0.41	171	Dummy variable with Value 1 if increase in the Democratic Polity IV project index of 3 or more points occurred in the country over the last 10 years.	Polity IV	2008
Level2	118	2	17	18	211	Number of Jurisdictions in Level #	Several	Several Years
Level3	14,000	5	442	1,352	164	Number of Jurisdictions in Level #	Several	Several Years
Level4	240,073	-	7,237	31,144	62	Number of Jurisdictions in Level #	Several	Several Years
Level5	42,008	432	14,822	17,954	10	Number of Jurisdictions in Level #	Several	Several Years
Pop/Juris	577,674	542.9	56,736	86,057	212	Average Population per jurisdiction	Own Calculations	Several Years
Area/Juris	76,501	0.18	1,981	6,261	212	Average Area per Jurisdiction	Own Calculations	Several Years
Sublevels	4	0	1.944	.772	196	Number of Sub-national Levels of Government	Several	Several Years

3.4 Results

We now turn to the analysis of the results from the estimation of the first model of jurisdictional fragmentation using the number of sub-national levels of government of a given country as the dependent variable. In Chapter 1 of this dissertation we argued that the creation of a new level of government would be justified if overall welfare is improved (considering losses from economies of scale, gains from tailoring preferences to citizens, and preferences for political accountability) by assigning the provision of public goods and services from the central (or local) level of government to a new sub-national level.

The hypothesized determinants of the number of sub-national levels of government coincide therefore with those used in the above analysis of jurisdictional fragmentation. As we discussed, we use maximum likelihood estimation (i.e. a tobit regression model) to conduct this analysis. The results of the tobit regression model are shown in Table 3.2 below.

From our results, it would seem that the vertical structure of government might be solely related to “size” variables and not to other institutional or preference-related aspects. As naturally expected, we find a strong positive relationship between population and fragmentation, defined as the number of sub-national tiers of government. Equally, a large country area seems to increase the probability of having more sub-national levels of

government. Both results are robust to the different model specifications, even after regional dummy variables are included.

Of the set of variables measuring heterogeneity of preferences, it would not seem that any of them offers significant insights as to their impact on this aspect of fragmentation. We could not find either any significant relationship between our set of institutional variables and the number of levels of sub-national government. This is somewhat surprising since we had expected that for example more ethnic fractionalization would have led to a higher probability of having regional level governments representing those differences. It would seem that, at least in our sample, geographical effects, captured partly by the regional dummies, are more important determinants of this aspect of jurisdictional fragmentation.

Table 9. Results from the Tobit Model

	Dep. Variable: Number of Sub-national Levels of Government		
	Model 1	Model 2	Model 3
Population	.091** (.041)	.086** (.042)	.092** (.042)
Area	.064** (.031)	.060* (.031)	.055* (.029)
AleELF	-.361 (.224)	-.364 (.228)	-.281 (.229)
Gini	-.004 (.005)	-.004 (.005)	.006 (.006)
ConIneq	.144 (.139)	.141 (.145)	.196 (.165)
GDPpc	-.058 (.058)	-.048 (.063)	-.049 (.062)
AHV	-.024 (1.785)	.188 (1.894)	.022 (1.817)
Sub National Elections*Dem Status 2		.006 (.021)	.001 (.022)
VAI		-.037 (.057)	-.084 (.058)
Oceania			-.089 (.183)
South America			-.383* (.205)
Central America			-.535*** (.157)
North America			-.684*** (.217)
Asia			-.386** (.174)
Africa			-.379 (.213)
Constant	.673 (1.221)	.682 (1.269)	.733 (1.192)
Number of Observations	144	141	141
Pseudo R-squared	0.1358	0.1259	0.1684
Log Pseudo-Likelihood	-114.65	-113.25	-107.75
Sigma	.536 (.034)	.540 (.033)	.519 (.034)
Probability > F	0.000	0.000	0.000
Notes: MLE estimation. Robust standard errors reported in parenthesis. Population, GDPpc and Area in logarithmic form. Europe is the reference group for the regional dummy variables. ***, **, and * denote significance at the 1, 5 and 10% level respectively			

Estimation results for the next three measures of jurisdictional fragmentation are presented in Table 3.3 below. As discussed, we use three measures of jurisdictional fragmentation: total number of jurisdictions, population average and area size average per jurisdiction. For each of these variables, we estimate three model specifications. In the first one (Models 1, 4 and 7 respectively), we test the standard economic hypotheses of fragmentation, that is, the impact of economies of scale and of heterogeneity of preferences in the level of jurisdictional fragmentation.

Ramsey tests were run on each of the regressions, searching for possible omitted variable bias, with negative results.²⁶ Initial plotting of the regressions residuals against the main variables of interest showed departures from the linearity assumption, and thus transformation of certain independent variables were required. In particular, we hypothesized a quadratic relationship between fragmentation and population levels; we also transformed area and per capita GDP into a logarithmic form, and also took the logs of our dependent variables. Correlation was found among some variables (especially between population and area as it was to be expected), leading to a certain amount of multicollinearity. Most of our coefficients are however individually and jointly significant, and very robust to different model specifications, both in terms of their size and signs, so this aspect did not seem to be relevant.

We applied the Breusch-Pagan test to our initial model specification, with the results suggesting a certain amount of heteroscedasticity existed, confirmed by the

²⁶ We used both `linktest` and `ovtest` Stata commands, with both tests supporting the hypothesis of no omitted variables in the model.

plotting of residuals against some of the quasi-ordinal variables in the sample (such as Sub-national Democratic status). As a result, our final estimations calculated and reported robust standard errors.

Our second model specification includes proxy variables attempting to measure demand for political accountability. From our theoretical framework, we derived the result that, all things equal, demand for political accountability should translate into a smaller optimal jurisdictional size, and thus lead to higher fragmentation. Results of this specification can be seen in Models 2, 5 and 8. Finally, our third model specification (for which results are presented in Model 3, 6 and 9 below) included a series of geographical and institutional control variables meant to test the robustness of the findings.

Regrettably, the variable approximating potential for economies of scale (technological readiness) proved insignificant and did not add to the overall explanatory power of the model. Individual significance tests confirmed this finding and the variable was discarded from the final model specification. This result left us without a proxy variable for economies of scale, a fact already common to other studies on the determinants of fragmentation. As we discussed, the difficulties in approximating this dimension seem insurmountable at this time. Ideally, as in Chapter 2, we would like to input into the analysis the “true value” of economies of scale for local services (that is, the true value of the cost elasticity of production) in a given country derived from a meta-analysis such as the one conducted in the previous chapter. That is not an available option right now. The proxies considered (population, technological readiness, market size) all

present serious identification issues or are simply not significant in the empirical analysis. The absence of this variable would mean that the variation in the level of jurisdictional fragmentation across countries due to the different potential for economies of scale is left in the error term, reducing the explanatory power of the model. However, in terms of the remaining set of hypotheses, we still draw important conclusions, outlined below.

First, the results show that, as expected, population and area are naturally and positively correlated with the level of jurisdictional fragmentation of a country. Population seems to be non-linearly related to jurisdictional fragmentation. Higher population leads to increasing levels of fragmentation, but at a decreasing rate. This relationship holds both for absolute measures of fragmentation (total number of jurisdictions) or relative ones (ratio of area to number of jurisdictions). The maximum level of fragmentation is reached at a population level of around 1 billion people, a level that only India and China have reached. An alternative model specification was tested whereby population was included as explanatory variable in the model in a logarithmic (and not quadratic) form. The estimation of this specification showed that a 1 percent increase in population led to a 0.68 percent increase in the total number of jurisdictions, a less than proportional increase. The relationship between area and fragmentation is equally significant. A 1 percent increase in country area leads to a .3 percent increase in the total number of jurisdictions in the full model specification.

The set of variables measuring heterogeneity of preferences also offers statistically significant and sizable results. Our measure of ethno-linguistic fragmentation

suggests, perhaps counter-intuitively, that greater dispersion on ethnic and linguistic groups leads to lower jurisdictional fragmentation, both in terms of smaller number of jurisdictions and higher average area. The result may a priori seem surprising, but it is very robust across model specifications, and even when an alternative measure of ethno-linguistic fragmentation is used. It is reasonable to assume that ethno-linguistic differences would lead to heterogeneous preferences, but the data would seem to show that the end result may be a lower level of jurisdictional fragmentation, both in terms of the total number of jurisdictions and larger average size. A possible explanation for this result is that ethno-linguistic fractionalization is more prevalent in poorer, low-growth countries with lower quality of government (Alberto Alesina et al., 2003). Already (Aghion, Alesina, & Trebbi, 2002) find that ethnic fractionalization is inversely related to quality of democracy²⁷. It follows that countries with lower quality of government (less democratic societies) and high ethno-linguistic fractionalization are less likely to create spaces for democratic representation with additional levels of government or new jurisdictions. This is an aspect for which we already control for in our model however²⁸, so we are left without a better explanation for the consistency of this relationship.

Our second variable approximating heterogeneous preferences, the age dispersion index, displays the hypothesized sign, but it is not robust across specifications. It would seem that greater age dispersion leads to a smaller average jurisdictional size, but we find

²⁷ A result robust in our sample even when we control for country latitude.

²⁸ In fact we find a negative correlation between the Alesina et al. measure of ethnic fractionalization and the variables approximating quality of democracy in our sample.

no statistically significant evidence of the variable affecting the total number of jurisdictions or average population size.

Larger per capita GDP also leads to lower levels of fragmentation in our sample. To the extent that overall increases in per capita income tend to homogenize preferences for public good provision, this result may be a reasonable one to expect. Although significant, the relationship between jurisdictional fragmentation and inequality shows that increases in inequality may lead to lower jurisdictional fragmentation overall, a result that goes against what we expected. The results are not robust however to the insertion of regional dummies in the model.

Our variables approximating preferences for political accountability are significant and display the hypothesized sign. The interaction term formed by the existence of sub-national elections and a measure of democratic status or legislative competitiveness proved to be significant and robust to different model specifications. The sign of the coefficient is positive, indicating that increased demand for political accountability leads to higher jurisdictional fragmentation, both in terms of a greater number of jurisdictions or smaller average area per jurisdictions. It could be argued that this relationship may be endogenous, and that in more jurisdictionally fragmented countries citizens are more participative and more likely to demand enhanced political accountability. That may not be the case however. As we discussed, (Lowery & Lyons, 1989) showed no evidence of better information levels of citizens at lower levels of government. In any case it would seem that the construction of the interaction term would

dispel doubts as to the possible endogeneity between these two variables. The use of an independent assessment of democratic quality and of an index of legislative competitiveness also independent from the territorial organization of the country should diminish the possible problem of double causation. The World Bank Voice and Accountability Index, a “measure of the extent to which citizens are able to participate in the selection of governments”, displayed identical sign as the earlier measure of democratic quality but it was only statistically significant when average area per jurisdiction was selected as the dependent variable.

We explored several other avenues through which institutional aspects related to demand for political accountability could have impacted the level of jurisdictional fragmentation. First, we included the number of sub-national levels of government as an explanatory variable. A larger number of jurisdictions is likely to be found in countries with more levels of government. The reason behind is that intermediate levels of government are able to deliver services to local or municipal governments unable to do so due to their small size. Lack of resort to such intermediate levels of government (i.e., county governments in the U.S., regional districts in British Columbia, or provincial governments in Spain) would introduce incentives to local mergers, and thus reduce jurisdictional fragmentation. This relationship proved to be highly significant. As hypothesized, more levels of government led to an increase in the overall number of jurisdictions. The results, however, did not suggest that a similarly strong statistical relationship existed between levels of government and the average jurisdictional size.

Secondly, we tested a set of several institutional aspects that may have reasonably affected the level of fragmentation and are proxies for democratic quality. These included; 1) the number of years of the executive in office (signal of quasi-democratic systems when its value exceeds 8); 2) the constitutional existence of a finite term in office before new elections are called; 3) whether a nationalist party is in power (more likely to advocate for a less decentralized state) and 4) the parliamentary system of the country. Of this set of variables, only the existence of a finite term in office seemed to significantly affect the total number of jurisdictions, but not their average size. It must be again mentioned that some of the explanatory variables we use may exert their influence over the process of creation or merger of jurisdictions over time. This influence may take more time than one year in which we are observing the data and therefore our cross-sectional analysis is limited in this respect.

Thirdly, we introduced in the analysis available indicators measuring the size of the public sector in the GDP, the share of sub-national expenditure over the national total, and a variable measuring whether fiscal authority existed over expenditure or tax policies at the local level (for which only 29 observations were available). We could not find significant relationships between these variables and the level of jurisdictional fragmentation and were eliminated from final model specifications. Fourthly, the variable approximating built up demands for jurisdictional fragmentation that may have been released via recent advances to the democratic quality of system of government proved to be non-significant.

Finally, in order to at least partly to account for the country fixed effects from our sample, we introduced regional dummy variables. Using Europe as the group of reference (arguably the most fragmented region together with North America), the dummy variables showed significant and lower levels of fragmentation. Their inclusion did not affect the significance of our main variables for the different hypothesis and added importantly to the overall explanatory value of the model.

Table 10. Results from the OLS Estimation

	Dep. Variable: Log of Total Number of Jurisdictions			Dep. Variable: Log of Area to Jurisdictions Ratio		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Population (Millions)	.020*** (.003)	.021*** (.003)	.019*** (.003)	-.006** (.003)	-.006** (.003)	-.006* (.003)
Population Squared	$-.1 \times 10^{-4}$ *** (0.02×10^{-4})	$-.1 \times 10^{-4}$ *** (0.02×10^{-4})	$-.1 \times 10^{-4}$ *** (0.02×10^{-4})	0.04×10^{-4} * (0.02×10^{-4})	0.04×10^{-4} * (0.02×10^{-4})	0.04×10^{-4} * (0.02×10^{-4})
Area	.384*** (.061)	.266*** (.060)	.292*** (.057)			
AleELF	-1.899*** (.583)	-1.380** (.539)	-1.392** (.565)	2.851*** (.711)	2.553*** (.692)	2.251*** (.745)
Gini	-.024** (.011)	-.024*** (.009)	-.006 (.011)	.033** (.015)	.0362** (.014)	.007 (.016)
ConIneq	-.235 (.279)	-.407 (.250)	-.211 (.273)	.166 (.345)	.141 (.323)	-.056 (.348)
GDPpc	-.140 (.146)	-.249** (.128)	-.324*** (.138)	.181 (.192)	.401** (.186)	.413* (.212)
AHV	-4.590 (4.513)	-6.348 (4.095)	-5.529 (4.317)	-11.641* (6.507)	-14.552** (7.041)	-11.792 (7.119)
Sub National Elections*Dem Status 2		.145*** (.035)	.137*** (.038)		-.132** (.055)	-.123** (.059)
VAI		.118 (.121)	.099 (.127)		-.333** (.157)	-.333* (.170)
Sub Levels		.907*** (.199)	.806*** (.201)		-.238 (.198)	-.235 (.203)
Oceania			-1.531*** (.387)			2.614*** (.822)
South America			-.875** (.431)			1.892*** (.521)
Central America			-.805*** (.369)			.305 (.565)
North America			.206 (.713)			.856 (.728)
Asia			-.568 (.388)			.701 (.572)
Africa			-.935** (.442)			1.229** (.624)
Constant	1.553 (2.467)	6.131** (2.762)	6.208** (2.841)	7.949** (3.612)	-4.936 (4.137)	-3.118 (4.236)
Number of Observations	144	141	141	144	141	141
R-squared	0.58	0.67	0.71	0.26	0.31	0.39
Probability > F	0.000	0.000	0.000	0.000	0.000	0.000

Notes: OLS estimation. Robust standard errors reported in parenthesis. GDPpc and Area in logarithmic form.
Europe is the reference group for the regional dummy variables.
***, **, and * denote significance at the 1, 5 and 10% level respectively

Table 10. Results from the OLS Estimation (Cont.)

	Dep. Variable: Log of Population to Jurisdictions Ratio		
	Model 7	Model 8	Model 9
Area	-.015 (.055)	.017 (.054)	.026 (.059)
AleELF	1.178 (.547)	.826 (.529)	.884 (.562)
Gini	.013 (.011)	.017 (.012)	.003 (.015)
ConIneq	.157 (.275)	.241 (.247)	.026 (.264)
GDPpc	-.062 (.139)	.071 (.137)	.187 (.146)
AHV	2.635 (4.474)	3.932 (4.177)	4.362 (4.281)
Sub National Elections*Dem Status 2		-.143*** (.037)	-.128*** (.041)
VAI		-.116 (.133)	-.072 (.131)
Sub Levels		-.427** (.195)	-.357* (.204)
Oceania			.789** (.382)
South America			.677 (.492)
Central America			.724* (.396)
North America			-.598 (.751)
Asia			.794* (.417)
Africa			.948* (.498)
Constant	8.569*** (3.067)	8.362*** (2.958)	6.724** (2.941)
Number of Observations	144	141	141
R-squared	0.13	0.23	0.28
Probability > F	0.005	0.000	0.000
Notes: OLS estimation. Robust standard errors reported in parenthesis. GDPpc and Area in logarithmic form. Europe is the reference group for the regional dummy variables. ***, **, and * denote significance at the 1, 5 and 10% level respectively			

3.5 Conclusions

In this chapter, we have tested the traditional hypotheses of the literature on optimal jurisdictional size (economies of scale and heterogeneous preferences), but also the implications of preferences for political accountability on the international levels of jurisdictional fragmentation observed around the world.

We have, for the first time to our knowledge, analyzed a largely unexplored dimension of jurisdictional fragmentation, the determinants of the number of levels of government, together with more traditional measures of fragmentation used in the literature. Overall, the vertical structure of government, at least in our sample, might be mostly related to “size” variables and not to other institutional or preference-related aspects. Both population and area size are positively correlated to the number of government tiers of a country. This result is common to all model specifications and dependent variables used as measures of fragmentation.

Our main result, in line with the conclusions of our theoretical model, is that preferences for political accountability would lead to smaller jurisdictional size and a larger number of governments. These results are robust across different model estimations. They also lend strong support to our contention that accountability needs to be added to the list of critical dimensions in the theory of optimal jurisdictional size, even though this aspect has not yet been fully incorporated into the standard economic theory.

Additionally, we find support to the hypothesis that the higher the number of government tiers, the more likely it is that jurisdictional fragmentation will be high. Small local governments have less of an incentive to consolidate into larger jurisdictions if the immediately higher government tier provides, assuming subsidiary responsibility, the services the former are unable to due to their limited capacity.

Although there is a strong theoretical case to expect that more heterogeneous preferences should lead, all other things equal, to more jurisdictional fragmentation, the set of variables used to approximate this aspect display mixed results. In some cases, although the variable proposed may be certainly reflective of heterogeneous preferences, other interfering variables may affect the causality of the relationship sought. High ethno-linguistic fragmentation, for instance, seem to be related to lower levels of fragmentation due to political economy issues related to the quality of democracy in countries that show high ethno-linguistic fractionalization. As we have argued, ethno-linguistic fractionalization is more prevalent in poorer, low-growth countries with lower quality of government. Thus, less democratic societies with high ethno-linguistic fractionalization are less likely to create spaces for democratic representation with additional levels of government or new jurisdictions.

Our analysis underlines also the need to continue to explore new avenues for the measurement of economies of scale on local service delivery, a critical aspect that the empirical literature, including our contribution, has not been able to address adequately.

CHAPTER 4
A ROADMAP FOR THE ANALYSIS OF JURISDICTIONAL FORMATION
AND REFORM

“...it is still too early to conclude that an optimum size has been established for metropolitan and local governments.” Breton (1965)

4.1 Introduction

Is a country's vertical structure of government efficient? Should it reduce the number of small jurisdictions via consolidation and jurisdictional merger? Under what circumstances should the country consider the creation of a new government tier? These and similar questions are posed regularly in countries around the world, regardless of their development level, cultural ties, religious leanings or cultural and institutional fabric. The answers to such questions are not straightforward. As in many other aspects of public policy, we have to conclude that there is no single ideal level of jurisdictional fragmentation. Delving in the empirical and applied policy literature in this area, we may however be able to suggest a process to develop a solid technical base to assist policy formulation. Thus, in this final chapter, we aim to provide an analytical framework that guides the formulation of jurisdictional reform by synthesizing the empirical results of this dissertation and summarizing the lessons learnt from the international experience.

The international experience shows that countries around the world have defined a vast number of criteria, both qualitative and quantitative, for regulating the formation of new jurisdictions (or the merge of existing ones). The former offer enhanced flexibility (and are perhaps more vulnerable to political manipulation) and have therefore been favored commonly over more transparent quantitative measures (Handry & Martinez-Vazquez, 2008). The review of these criteria offers a first assessment of how theoretical considerations related to the determination of optimal jurisdictional size have been translated into policy. It also represents a benchmark to analyze to what extent current levels of jurisdictional fragmentation fit the ideal “efficient” patterns.

Among the most widely used criteria for regulating new jurisdiction formation, most countries use different forms of population criteria. These vary significantly even within a given geographic region (for example, from the 50,000 people requirement in Ecuador to the 5,000 requirement of Paraguay in Latin America) and may be determined either by the national government (for regional and municipal tiers) or, in federal or quasi-federal countries, by regional levels of government (for municipal or local levels, such as in Argentina, Spain, Australia, etc.). In addition, some countries have required that the formation of a new jurisdiction be subjected to minimum land area criteria. For instance, in Colombia, a new municipality must not represent more than one third of the territory of the municipality from which it emerges. In addition, it is normally required that the territory of the new jurisdiction has territorial continuity.

The review process for new jurisdiction formation also includes generally the analysis of the socio-economic conditions of the localities and the population support for the request, thus evaluating to a certain extent the existence of heterogeneous preferences. In Latvia for instance, the Council for Administrative Territorial Reform conducts a full investigation for new jurisdiction formation that includes these issues. Additionally, higher tiers of government tend to evaluate the fiscal and administrative feasibility of the future jurisdictions. In Mexico, the law requires that the evaluation of requests for new jurisdiction formation includes a feasibility analysis that demonstrates the new jurisdictions will have sufficient fiscal capacity to run their administrations.

At the origin of demands for new jurisdiction formation we also find a preference for political accountability or accessibility to local government representatives. Among the criteria used to evaluate the representatives' capacity to attend to their constituencies when considering new jurisdictions we find population density, average distance to the new jurisdiction capital, land area, or even the state of local communications infrastructure as a proxy for travel time to the capital center (Mexico).

In a recent report on the process of jurisdictional fragmentation observed in Indonesia, Handry & Martinez-Vazquez (2008) distil the main criteria that should lead the consideration of new jurisdiction formation: 1) the production/cost efficiency criterion; 2) the representation/political responsiveness/accountability criterion; 3) the financial/fiscal capacity criterion; and 4) the administrative capacity criterion. In this chapter we combine those criteria with the results of the empirical analysis conducted in chapters 3

and 4 of this dissertation to provide an analytical framework for the evaluation and implementation of jurisdictional reforms. As such, we do not summarize here the conclusions already contained in the empirical analysis, but apply them, together with the lessons learnt from other contributions, to the definition of a set of lessons or principles that may assist the formulation of jurisdictional consolidation/fragmentation policies around the world.

4.2 A set of guiding principles for jurisdictional reform

1. Clarify the objectives of the jurisdictional reform.

Although it may appear to be a straightforward proposition, any jurisdictional reform must depart from the definition of clear policy objectives. This is especially important in light of the implicit trade-offs between some of the policy goals generally considered for this type of reform. In the past, jurisdictional reforms have been pursued, among other reasons, to: 1) achieve savings in the average cost of local public service production; 2) improve the allocative efficiency of government by better matching local preferences with jurisdictional borders; 3) create new democratic spaces of representation; and 4) assist economic competition in a globalized world (especially in the case of large metropolitan centers).

For example, when jurisdictional reforms are initiated in the context of fiscal crisis, the overarching objective is likely to be the reduction of average costs of service provision.

This, if it leads to consolidation in order to achieve economies of scale, may be at odds with alternative objectives such as ensuring political accountability or attending to heterogeneous preferences. Thus a clear list with the prioritization of objectives can assist better policy formulation.

However, the traditional policy tradeoffs encountered when approaching jurisdictional reform may not be insurmountable. In theory, it may be viable to ensure a greater level of achievement of these independent policy objectives if the policy alternatives considered are not solely limited to jurisdictional consolidation or fragmentation. Recent international experiences show that there is an array of institutional mechanisms for local service delivery that can achieve efficiency gains in the production of public services without compromising political accountability or the tailoring of public services to local preferences. The latter will be further discussed below.

2. Avoid dogmatic and uniform approaches: One size does not fit all.

Jurisdictional reform must start from the assumption that the expected efficiency gains from consolidation or further fragmentation may not be shared across the territory by all existing jurisdictions. The analysis of the empirical literature on economies of scale in public service delivery shows that the population level at which such savings in the average cost of production is exhausted varies widely by service and by sample of jurisdictions.

The pioneering work of Hirsch (1959) unveiled that economies of scale in fire services were exhausted after a 100,000 population level was reached. In Bodkin & Conklin (1971) the reduction in the average production costs of police and fire services was limited to population sizes between 5,000 and 10,000 citizens, whereas for water supply and public works it extended to population sizes of around 150,000 people. For solid waste collection, McDavid (2000) determines full economies of scale are reached in localities with up to 5,000 inhabitants in Canada, while Stevens (1978) optimal population size for refuse collection in his USA sample is reached at 20,000 people. Finally, in Spain, Bel (2005) also finds that population sizes of around 20,000 exhaust the potential economies of scale in waste collection.

Such population limits are even more restrictive in the area of education. As discussed above, Duncombe et al. (1995) show that only the consolidation of districts with less than 500 students in the state of New York offers potential for the reduction in average costs. In a similar study, Edelman & Knudsen (1990) in Iowa conclude that economies of scale are concentrated in student populations between 800 and 900 students. Lastly, in Maine, the optimal student population size for a school district is estimated by Deller & Rudnicki (1992) in around 2000 students.

Thus optimal jurisdictional sizes for service delivery may vary by service, and even within the same service, due to other factors such as geography, may not be identical across a given national territory. Furthermore, as production and delivery technology change, it may affect the expected potential for economies of scale dramatically. In such

a fluid environment, flexible approaches that allow multiple institutional set ups for service delivery are required to respond to the different characteristics of a country's jurisdictions.

3. Start by clarifying sub-national functional assignments.

In light of the different potential for economies of scale associated with the traditional local public services, a clear understanding of the sub-national expenditure assignments is required in order to estimate the net impact of jurisdictional reform on the average cost of service production. Consolidation for instance would not offer great efficiency gains if the set of services being delivered by local governments are such (and will remain to be such) that the optimal jurisdictional sizes for their production are small.

Expenditure assignments may not even be homogeneous across a given country's government tiers. Asymmetric assignments, whereby increased expenditure authority is allocated for a larger number of services to jurisdictions with proven managerial capacity are common around the world. This adds yet another caveat to the expected gains to be obtained from consolidation for instance, since the consolidation policy could be promising (in terms of reductions in average production costs) for just a given sub-set of sub-national governments and not others of the same tier.

4. Examine each case individually under transparent guidelines and identify the potential for economies of scale.

In light of the significant number of caveats that may affect the achievement of the objectives defined for jurisdictional reform (either consolidation or fragmentation), individual attention to each case is important to establish whether such objectives are likely to be achieved. Clear national (or sub-national, if regions for instance are tasked with deciding on this issue) guidelines and criteria, both quantitative and qualitative, should be made available to jurisdictions considering altering their territorial limits.

In particular, the evaluation process must aim to offer substantive information and analysis to the citizenry as to what can be expected from the consolidation (or division) process. For example, details of the expected savings (or improvements in the quality) of the provision of services could help curb opposition to consolidation. Over the last couple of decades, the methodological analysis of economies of scale has improved considerably. From the modeling of production costs using translog functions to the use of data envelopment analysis (DEA) to explore relative productive efficiency, advances in the analysis of economies of scale allow estimating their potential within reasonable margins. These and other methodologies can provide extremely useful information as to whether the desired gains in cost savings from jurisdictional reform are truly available and should become “due diligence” analytical requirements.

5. Preferences for political accountability lead to smaller jurisdictional sizes and reforms must adequately incorporate this aspect.

Even in situations where relatively homogeneous preferences exist, together with certain potential for economies of scale, consolidation may be hindered by citizen's preferences to remain close to their local representatives. The inertia of existing institutional arrangements and the identity associated with them cannot be underestimated when considering jurisdictional reforms.

Our empirical analysis of jurisdictional fragmentation patterns across countries has controlled for the best available indicators of heterogeneous preferences and still finds that preferences for political accountability affect very importantly the results. Given the presence of such preferences, one expects to find average jurisdictional sizes below what the mere analysis of economies of scale and heterogeneous preferences would indicate.

Thus, citizen's resistance to consolidation in a given jurisdiction may also be the exercise of a preference for proximity to their local representatives, an important variable to be added to the economic analysis of optimal jurisdictional size. Acknowledging that preferences across jurisdictions may often be quite homogeneous, and that economies of scale may exist to a certain scale if consolidation ensued, citizens may still be willing to pay a higher price for services (in the form of taxes) for the geographical proximity to their representatives. In such situations, alternative avenues for service delivery that

allow reaching economies of scale while respecting current fragmentation levels will be required.

6. Existing institutions and regulations can jeopardize reaching efficiency gains in the short to medium term.

The expected administrative savings from jurisdictional consolidation may be compromised from the start. Regulations in public sector employment (or the power of public sector unions) for instance may lead to upward equalization of local public salaries in the event of consolidation, increasing administrative costs. Also, restrictions to retrenchment in public sector employment may limit short term efficiency gains in this area.

The relative importance of these aspects will vary importantly from one country to another depending on the regulatory framework for public sector employment. If, for instance, a large percentage of employees in sub-national administrations are under fixed term contracts, it may be easier to adapt to the expectedly lower human resources requirements of a consolidated administration. In the long run, all factors of production could be altered, but such analysis must be incorporated in the estimation of the cost savings (increases) expected from consolidation (fragmentation).

7. Intermediate levels of government may pose a limit to local consolidation.

The existence of intermediate tiers of government with subsidiary responsibility over the delivery of services that local governments cannot provide due to their limited capacity may take away incentives to jurisdictional mergers.

The scenario is relatively common in several countries around the world: small municipal governments unable to meet their expenditure assignments that relinquish the provision of certain services to the immediately higher tier. From a political economy perspective, these situation contributes to justify the role of the intermediate levels of government, while diminishes the incentives for municipal mergers. Our empirical analysis has shown this variable to be a highly significant determinant of current levels of fragmentation, thus lending support to a generalization of this result for a large sample of countries. On the other hand, the existence of intermediate levels of government may be perfectly justified because of the size of the country and the externalities across local jurisdictions in the provision of certain public services.

8. Multi-layered institutional frameworks may allow for efficiency gains without loss of political representation.

The international experience shows that, around the world, new institutional set ups for public service delivery may be allowing jurisdictions to reach economies of scale without the need for changes to jurisdictional sizes. The experience of special districts, inter-

municipal cooperation, or even home owners associations may offer alternatives to exploiting economies of scale where they exist.

A scenario where a local government delivers some services directly, contracts out others and cooperates yet with other local governments in the joint delivery of some will increasingly become the rule and not the exception across the world. Thus, a country's legal framework should allow local governments to opt, case by case, for the institutional structure more adequate to their characteristics and needs.

The realization that not all services offer similar potential for economies of scale, added to strong citizen's preference for political accountability, should lead policy makers to depart from the traditional policy choice between consolidation and fragmentation, and include new alternatives for service delivery. In Table 4.1 we discuss the main advantages and disadvantages of three of these alternatives; direct service provision, privatization and inter-jurisdictional cooperation.

In summary, the economies of scale pursued with consolidation can be reached through alternative avenues (Warner, 2006). Privatization and cooperation both allow additionally maintaining identical levels of government representation, as they focus on the provision of those services with potential for cost savings. On the other hand, both privatization and cooperation may incorporate high transaction costs into the production of services. The definition of contracts either among different jurisdictions or with private companies is an imperfect process that takes away control from the local administrations. However, as

they are contracts, they can be regularly re-negotiated or re-designed, a less plausible option in the case of consolidation. In any case, as experience in contracting accumulates and trust is built among the jurisdictions and with their private contractors, these transaction costs are likely to diminish over time.

Table 11. Institutional Options for Service Delivery: A comparison

	Advantages				Disadvantages		
	Economies of Scale	Economies of Scope	Political Accountability	Private Sector Development	Transaction Costs	Regulatory Costs	Loss of Political Accountability
Direct Provision via Consolidation	Available depending on the service and the size of the market.	Available potentially as it allows managing a larger number of services with a lower resource endowment.	Fully retained	Does not contribute to local private sector development.	Should not increase.	Should not increase.	Loss of institutional representativeness. Greater distance between local government and its citizens.
Privatization	Available depending on the service and the size of the market.	Available potentially as it allows concentrating resources in other public services.	Allows maintaining level of political representation but loss of control over service delivery.	Potentially contributes to the development of the local private sector if used as suppliers/producers.	High initial contractual transaction costs. Diminish in the long run.	High, especially in production process that are highly technological in nature.	Potential loss of control over production and quality levels. High political costs if prices increase without control.
Inter-jurisdictional Cooperation	Available depending on the service and the size of the market.	Available potentially depending on the production method selected for the service.	Allows maintaining an almost identical level of institutional representation.	Potentially contributes to the development of the local private sector if used as suppliers/producers.	Potentially high, diminish in the long run as experience and trust builds.	None, unless cooperation leads to subcontracting to private sector.	Minimal, and recoverable if agreements expire or are cancelled.

Source: Martinez Vazquez y Gomez (2008)

The formation of special districts, a mostly U.S. and French phenomenon, provides an alternative to inter-jurisdictional cooperation, where jurisdictions create a new institutional body to manage and finance a particular local service. This model may be considered as a last stage in the evolution of inter-jurisdictional cooperation, where such arrangements are institutionalized to the point of creating a truly new government unit.

The share of services being provided by any of these institutional mechanisms is expected to vary over time. As Warner (2006) argues, the share of services jointly provided via cooperation in the USA has been diminishing, especially in rural areas over the last two decades, due to structural factors related to the ability of rural localities to compete in an increasingly market-based governmental system. It is expected that similar forces will be at play at different stages in countries around the world, leading us to multiple equilibria in overall sub-national government structures.

9. Consolidation should be incentivized, not forced.

There is little rationale for forced jurisdictional consolidation unless it reaches some of the goals defined above. If the potential to achieve those goals exists, then policy should facilitate the voluntary agreement of local governments to merge. Political opposition from the electorate may jeopardize the success of the experience to the point of making it fail.

If the efficiency gains can reasonably be achieved via other service delivery mechanisms, then consolidation should be suggested as a voluntary option and incentivized from the higher government tiers if considered an adequate policy path. Forced consolidation may encounter such a popular opposition that may in fact compromise the efficiency gains expected from the process. As with other incentives schemes, the net positive benefits expected from the process of consolidation should be independent from the incentives provided. If only fiscal incentives make consolidation attractive, as they are likely to expire over time, they may not lead to substantive reform. Additional grants to assist the building-up of the administrative capacity of the jurisdictions merging, or for the construction of transport and communications infrastructure that better integrate the merging jurisdictions, are reasonable incentives to offer. However, the sole provision of short to medium term fiscal incentives, devoid of solid, long-lasting efficiency gains from jurisdictional consolidation in the form of economies of scale or others, is a policy likely doomed to fail.

10. Widespread advocacy on the benefits of consolidation is required.

Linked to the earlier point, voluntary consolidation, where it needs a referendum passing the new jurisdictional structure, requires intense and widespread advocacy to communicate to the population the benefits expected from the measure. If the cost savings expected and the improvement in service delivery are not obvious to the citizens, the policy will not be understood or supported. Consolidation is a policy that constantly

must be linked to the long term national development plans. Otherwise it may be perceived as a political game detached from the preoccupations of the common citizen.

4.3 Conclusions

The relatively low attention that such an important topic as the vertical structure of government has received in the public finance literature is somewhat surprising. The theoretical and empirical work developed in this dissertation is an attempt to contribute to advance our knowledge in this critical area for public sector performance.

A first general conclusion drawn from our analysis is that the optimal level of jurisdictional fragmentation seems to be in permanent evolution across countries depending on the available technology for public services production, on the citizen's preferences for political accountability, and on how the socio-economic, ethno-linguistic and geographical conditions of a given country affect the heterogeneity of local preferences for public goods and services.

In light of such fluidity in the optimal level of government fragmentation, it is advisable not to assume that “cookie cutter” policies will render homogenous and solid results. In particular, jurisdictional consolidation processes around the world have been regularly advocated without an adequate analysis of their potential benefits. We have argued in this dissertation that there is a need to avoid dogmatism in jurisdictional reform policy and opt instead for a flexible array of policy options that may allow reaching in theory

Private contracting, inter-jurisdictional cooperation, or the creation of special districts offer institutional mechanisms for service delivery that may allow reaching economies of scale (when they exist) without compromising local citizen's preferences for political accountability. It is becoming the norm, and not the exception, that local governments will deliver directly some services, privatize others, and cooperate with other jurisdictions in the joint production of the rest. Again, the share of services provided under each institutional alternative is likely (and has been shown) to vary depending on ideological preferences of the local representatives, changes to production technologies, or structural factors that reduce the ability of certain localities to cooperate and lead them back to direct provision (Warner, 2006).

This does not necessarily mean that all vertical structures of government are adequate. High jurisdictional fragmentation may present enormous challenges to efficient service delivery if it is coupled with low managerial capacity, weak tax bases, and inefficient political accountability processes. If the benefits from consolidation overwhelm those that could arguably be obtained from alternative policies, strong advocacy on those expected results should be undertaken, together with voluntary consolidation programs that are adequately incentivized. Equally, the process of new jurisdiction formation should be designed including both quantitative and qualitative minimum requirements in scale of production, fiscal sustainability, political accountability and administrative capacity.

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 Economía*, 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.
- 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.
- Baleiras, R. (2001). To fragment or to consolidate jurisdictions: the optimal architecture

- of government. *Working Paper*. <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.
- 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.
- 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.
- 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, 21 pages. <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.
- 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, 42 pages. <http://aysps.gsu.edu/isp/files/ispwp0912.pdf>. (June 2009)
- Bosch, N., & Suarez-Pandiello, J. (2008). Structural reform in Spain. In *The Theory and*

- Practice of Local Government Reform*, by Brian Dollery and Lorenzo Robotti (eds.), 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.
- 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.
- Dollery, B., & Robotti, L. (2009). *The Theory and Practice of Local Government Reform*. Edward Elgar Publishing.
- 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, 271-282.
- 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.

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.

- Huedo-Medina, T., Marin-Martinez, F., Sanchez-Meca, J., & Botella, J. (2006).
Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Center for
Health, Intervention, and Prevention (CHIP) CHIP Documents. 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.
- Jiminez, 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.
- 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.
<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. E., & Willoughby, K. G. (2001). Budgeters' Views of State Performance-Budgeting Systems: Distinctions across Branches. *Public Administration Review*, 61(1), 54-64.
- 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.
- 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 in L.H. Masotti and J.K. Hadden (eds) *The Urbanization of the Suburbs* (pp. 367-402). Beverly Hills, CA: Sage Publications.
- Ostrom, E., Schroeder, L., & 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 *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.
- Prud'homme, R. (1995). The dangers of decentralization. *The World Bank Research*

- Observer*, Vol. 10, No. 2, pp. 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 *The Theory and Practice of Local Government Reform*, by Brian Dollery and Lorenzo Robotti (eds.), 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. Deakin University. Faculty of Business and Law. SCHOOL OF ACCOUNTING, ECONOMICS AND FINANCE School Working Paper - Economic Series 2007 SWP 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.

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.