The hunt for the optimal jurisdiction size: Looking for simpler questions that can be answered

October 2013

Martin Bækgaard (martinb@ps.au.dk)

Søren Serritzlew (sorens@ps.au.dk)

&

Kim M. Sønderskov (kim@ps.au.dk)

Department of Political Science

Aarhus University

Draft, please do not cite
Introduction

The question of the optimal jurisdiction size of political systems is one of the most fundamental questions in political science. It was so in classical Greece: Plato’s The State and Aristotle’s Politics considered how city states should be organized and governed. It has remained so in the history of political philosophy (see, for instance, Dahl & Tufte, 1970: 4-16, and it still is in contemporary political science (see, for instance, Alesina’s & Spolaore’s The Size of Nations and Treisman’s The Architecture of Government). It seems, however, fair to say that definitive answers have yet to be given.

Why are we, despite centuries of attention from prominent political scientists and thinkers, in this regrettable position? The problem may be that the question is simple the wrong one. Perhaps the question of the optimal jurisdiction size is too complex to be answered. We suggest in this paper that a more modest approach is likely to be more productive: By asking a simpler question about the optimal size of the organizations providing essential government services, we can gain solid knowledge relevant to the question of efficient design of government.

The complications in seeking answers to the ambitious question of the optimal jurisdiction size of political systems are plentiful. One group of problems relates to the idea that a tradeoff exists between economic and democratic concerns. The potential tradeoff has been recognized since Plato and Aristotle (Dahl & Tufte, 1970: 4). The usual argument goes that larger jurisdictions may confer economic benefits due to economies of scale, while democracy thrives better in smaller entities because of the greater closeness from citizens to decision-makers. The evidence, however, is more complicated. Internal political efficacy is lower in larger jurisdictions (Lassen & Serritzlew, 2011), and so is participation (Oliver, 2000). However, citizens can affect extra and more important policy areas in larger jurisdictions, and larger systems reduce the likelihood that a group with particular
interests can dominate (Dahl & Tufte, 1970: 13-14). Dahl & Tufte (1970: 138) conclude that “Democratic goals conflict, and no single unit or kind of unit can best serve these goals”.

At the other side of the supposed tradeoff, Alesina & Spotaore (2003: 172) find that larger nations benefit from economies of scale, but that small countries can prosper too with free trade. However, Treisman (2007: 55) argues that scale effects are radically different for different public goods. Hence, there probably is no unambiguous optimum jurisdiction size with it comes to economic concerns. Therefore, if a tradeoff between economy and democracy exists at all, it is indeed a complicated one.

Another group of problems relates to the idea that the very question of optimal size is unequivocal. It is not. With rare exceptions, states are federal in the sense that decisions are decentralized to lower levels of government (Boadway & Shah, 2009). This implies that, in most countries, responsibilities can be and is assigned to different levels. Some types of services, such as defense, can be provided most efficiently for large jurisdictions. Others, typically those that need to be tailored to local preferences and needs (local infrastructure could be an example) are more efficiently provided in smaller jurisdictions (see Treisman, 2007: 55-56). Many other public goods and services fall in between these extremes. We can observe a great deal of variation in how these responsibilities are in practice assigned to levels of government, both spatially and over time. For example, in some countries, police is local. In other countries, it is run by the state. It may also be regional. Health care is in some countries provided by the state or federal level, in others by the regional or local level. Or it may mainly be provided by private hospitals, and privately funded. To make matters worse, most important services are often provided simultaneously by two or more levels of government. In the US, police agencies exist at the federal, state, county, and municipal levels. In health care, VHA is federal, much is private, but states have considerable leeway in determining eligibility for Medicaid.
Hence, it makes little sense to ask about the optimal jurisdiction size, even if we focus on economic concerns alone. The optimum size of a political entity depends on what services it provides. Consider, for example, Australia, where local government is only “engaged in the most minimal property-oriented services (primarily “roads and rubbish” (Boadway & Shah, 2009: 276). It may well be that the optimum size, in economic terms, is small, perhaps 5,000 inhabitants (the Australian municipalities are in fact larger than that). Imagine a municipality responsible for elementary schools, elderly care, child care, environmental regulation. The optimal size in such a system is likely to be larger.

This does, of course, not mean that the question of the economically optimal jurisdiction size should be abandoned. Rather, it suggests that we should direct more attention to questions that we can actually hope to answer. Answers to such questions will contribute to our understanding of the dilemmas involved in designing the jurisdiction size of political systems, and they will be highly relevant to policy-makers that deal with these questions in practice.

In this paper, we study the economic effects of the size of government institutions. We focus on public elementary schools. Education is, of course, one of the most important services, and it is, at least partly, publicly provided in almost all developed countries. We find that this simpler question has a clear answer. The size of government institutions matters for economic outcomes.

In the next section we review the existing empirical literature on the economic effects of school size. We conclude from this review that most existing studies suffer from a serious problem of simultaneity, which may explain that results have pointed in many different directions. We then discuss how this methodological problem can be solved by using IV-estimation in a large scale quasi experiment in Danish local government. We find that there are considerable economies of scale in public schools.
School Size and Costs

It follows from the discussion above that we cannot hope for an answer to the question about the economic effects of the size of government institutions generally. What we can do is to address some more mundane and concrete questions about the economic effects of size for particular services. Public school services seem particularly useful for this purpose because the service is almost universally provided and because it is one of the primary services in most local governmental systems. Moreover, a literature has developed about economies of scale in public schools but an unequivocal answer to the question about the impact of school size on costs has yet to be provided.

Economies of scale refer to how output responds to variation in input. Increasing returns to scale in public education (positive scale effects) mean that larger schools (or, alternatively, school districts) produce with lower per pupil costs than smaller ones. There may be increasing returns to scale in public education of a variety of reasons. First, all schools and their surrounding areas need regular maintenance and cleaning which in turn entails that staff need to be hired to take care of these tasks regardless of the school size. While the total costs of maintenance and cleaning are likely to be higher the larger the school, part of the costs may be constant across schools of different sizes and hence per pupil costs may be lower the larger the school. Second, and related, all schools need an administration including a superintendent. Again, per unit costs may be considerably lower in larger schools despite the total costs being higher. This also entails that merging two or more schools may decrease the total costs of maintenance, cleaning, and administration. Third, larger schools may produce more efficiently by creating classes with more pupils such that each teacher teach relatively more students than in smaller schools (Blom-Hansen 2004: 261f).
But there are also arguments about why increasing returns to scale may not exist in public education. First, some costs may not depend on the school size but on other factors. For instance, the use of electricity to heat or cool the building may depend on the age of the school, the material of which it is built, and its location. Second, the employees may be less effective in larger schools due to, for instance, more sickness absenteeism and administrative problems of controlling and coordinating employee behavior. It is therefore largely an empirical question whether the production of public education exhibits increasing returns to scale, constant returns to scale, decreasing returns to scale, or whether a u-shaped relationship is present where increasing returns to scale exists for small school sizes and decreasing returns to scale for large schools.

The literature on size effects in primary and secondary schools is voluminous and alive which should be evident from the number of review articles that have been published within the area in recent years (Fox 1981, Andrews et al. 2002, Colegrave and Giles 2008; Leithwood and Jantzi 2009). No attempt will be made to review the literature in full here. Instead we rely on the reviews above and a number of newer studies that are not included in any of the reviews. Table 1 reports the main findings of the studies:
Table 1. Reviews and single case studies of scale effects in local government

<table>
<thead>
<tr>
<th>Study</th>
<th>Surveyed studies on scale effects</th>
<th>Main conclusion:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fox (1981)</td>
<td>Review: 35 US studies</td>
<td>Economies of scale at both the school and district level. But huge differences</td>
</tr>
<tr>
<td></td>
<td></td>
<td>between studies about when scale effects set in and if diseconomies are present</td>
</tr>
<tr>
<td></td>
<td></td>
<td>at some point as well.</td>
</tr>
<tr>
<td>Houlberg (2000)</td>
<td>Single country study focusing</td>
<td>Evidence of increasing returns to scale</td>
</tr>
<tr>
<td></td>
<td>on economies of scale of school</td>
<td></td>
</tr>
<tr>
<td></td>
<td>districts</td>
<td></td>
</tr>
<tr>
<td>Andrews et al. (2002)</td>
<td>Review: 22 US studies</td>
<td>Mixed results; no conclusive evidence on scale effects</td>
</tr>
<tr>
<td></td>
<td>elementary schools</td>
<td></td>
</tr>
<tr>
<td>Colegrave and Giles (2008)</td>
<td>Meta regression analysis</td>
<td>Evidence of increasing returns to scale until a size of approximately 1,500</td>
</tr>
<tr>
<td></td>
<td>including 22 estimates of optimal</td>
<td>pupils after which decreasing returns to scale set in.</td>
</tr>
<tr>
<td></td>
<td>school sizes (origin not</td>
<td></td>
</tr>
<tr>
<td></td>
<td>specified)</td>
<td></td>
</tr>
<tr>
<td>Leithwood and Jantzi (2009)</td>
<td>Review: 57 Studies from all</td>
<td>Mixed results; no conclusive evidence on scale effects</td>
</tr>
<tr>
<td></td>
<td>over the world</td>
<td></td>
</tr>
<tr>
<td>Zimmer et al. 2009</td>
<td>Single state study of Indiana</td>
<td>Evidence of increasing returns to scale for enrollment levels below 2,000</td>
</tr>
<tr>
<td></td>
<td>school districts</td>
<td>students after which diseconomies of scale may emerge</td>
</tr>
</tbody>
</table>

Based on Table 1 it seems fair to say that a clear conclusion has yet to be made. This may not be surprising since the findings in existing studies are based on one of two common approaches which do not appropriately account for simultaneity bias. The most common approach is to analyze cross-sectional or panel-level data of either schools or school districts using linear regression analysis in order to estimate a production- or cost-function (see, for instance, Chakroborty 2000; Bowles and Bosworth 2002; Robertson 2007). A common characteristic for these studies is that none of them are able to fully account for the fact that the dependent variable to some extent may affect the independent variable (Fox 1981). This may for instance be the case if school mergers take place among ineffective schools in particular.
The alternative and less common approach is to analyze per pupil cost changes in the aftermath of either school or school district consolidations (see, for instance, Duncumbe and Yinger 2007). The analysis may also in this case suffer from serious simultaneity bias because consolidations in many cases may have taken place in an attempt to cut costs. In both approaches increasing returns to scale are probably likely to be underestimated (or maybe not even identified) because consolidations of schools and school districts are likely to include ineffective schools (Duncombe, Miner and Ruggiero 1995; Dodson and Garrett 2004). But the opposite may also be true if, for instance, effective small schools are easier to merge into larger units than ineffective small schools which may for instance be the case if effective small schools to a larger extent than ineffective small schools are placed in densely populated areas where school consolidations may more easily take place. The point here therefore is that we cannot reach a conclusion about scale effects in public education without accounting for simultaneity bias. It is very telling that only one of the reviews and articles mentioned in Table 1 even mentions this central methodological problem and in this particular review simultaneity bias is mentioned as a problem that future studies should take care of rather than something that previous studies have accounted for (Fox 1981: 278). The literature on scale effects in public education thus seems to suffer from a general problem of inadequately accounting for simultaneity bias. In the next section we describe how we deal with this central methodological problem.

**Identifying the effect of changes in size on efficiency by experimental methods**

In a sense, bureaucrats are hired to, and politicians elected to, create problems of simultaneity (Blom-Hansen, Morten & Serritzlew, 2013). Government organizations must adapt to existing problems. Inefficient schools should, if bureaucrats and politicians do their job properly, be reformed in order to address the problem. This is a problem for researchers seeking to identify the effect of how government is organized, as it implies that we cannot know whether a statistical
correlation between size and efficiency reflects a causal effect of size, or rather that government organizations are reformed, expanded or reduced in response to low or high performance. This problem does not only apply to schools, but is entirely general.

Hence, the core methodological problem in the literature on the economies of scale in public school is simultaneity, or, in other words that the independent variable, size, is endogenous. The independent variable (x) is likely to be affected by the dependent variable (y) itself, because efficient schools may be more likely to merge, or with factors closely related to y, because, for instance, families may migrate to quality schools.

The problem can be tackled directly by experimental methods. Field experiments involve that values of the independent variable are randomly assigned to cases. This implies that x is in fact exogenous, and that any statistical association between x and y can be interpreted as an effect of x on y. For example, Bellé (2013) studies the effect of transformational leadership on performance in hospitals. He uses experimental methods to randomly assign nurses to groups, which were subject to different transformational leadership practices. Field experiments are feasible in cases where x can realistically be manipulated. However, this is typically impossible whenever the change required in x is expensive, controversial, or politically salient. Direct randomization is an unrealistic solution when these three characteristics apply. This will often be the case when it comes to broader questions of the effects of jurisdiction size on democracy and economic questions.

It is sometimes possible, as an alternative to direct random assignment, to identify a quasi-experiment, where changes in x, for a large number of cases, are known to be exogenous. It is difficult to imagine circumstances that would generate massive exogenous change in the size of schools (and probably most other important public organizations), and the review studies do not point to a single example. Closely related to the quasi-experiment is IV-regression. The logic of IV-
regression is to predict changes in $x$ by factors which logically cannot be affected by $y$. If it turns out to be possible to predict $x$ well by these factors, the predicted values of $x$ can be used to estimate $y$, and we can know with great certainty that the estimate can be interpreted as a causal effect of $x$.

To do this for public schools, it would be ideal to compute a measure of the chance that schools within a jurisdiction will experience a change in size. This measure, which could be interpreted as a redundancy score, should be based on objective factors that cannot be effects of $y$, and, to allow for statistical tests with leverage, there should be substantial variation in this measure across jurisdictions.

On the face of it, this should be quite easy. It is reasonable to expect that the redundancy score is a function of two exogenous factors: How many schools are geographically close to it, and how large are they. The problem, of course, is that these variables rarely change, and when they do, it will often be a result of changes in the school structure, which will then produce a problem of endogeneity once again.

To see how this problem can be solved, imagine a system with five schools, as in Figure 1 (A). Let us assume that, for some reason, half of the pupils were only allowed to attend to schools a or b, while the other half could only attend schools c, d, or e. Figure 1 (A) shows the distances between the schools. For school a, the only competition comes from school b. The distances is, say, 4 km. For school c, competition comes from d and e. The distance to each of these schools are, say 5 and km, respectively.
Now imagine that, by exogenous reform, pupils were allowed to choose any school. This change would imply that many additional distances become relevant, as shown in panel B. For instance, school b now faces competition from school c, which is only 2 km away. We should expect that it would be much more likely after the reform that either school b or c (perhaps the smallest of them) would close. If the reform were truly exogenous, some function of the changes in distance could be used to create an ideal instrument. This instrument would be computed from changes in redundancy scores and, hence, unaffected by y.

The comprehensive redrawing of municipal borders associated with the Danish municipal reform of 2007 provides exactly this. In the next section we introduce the Danish structural reform of 2007, and explain how it can be used to find instruments for changes in the average size of public schools.

The Danish structural reform as a source of exogenous variation in school size
Denmark has 98 municipalities, which are responsible a broad range of welfare services, including public schools. Ten years of primary education is compulsory in Denmark. Parents can elect public or private schools (home teaching is also an option, but it is used only be few). About 80% of children are enrolled in a public school. The public schools are regulated by law (stipulating what
subjects must be taught, and regulating exams and other assessments), but they are governed by the municipalities, which are responsible for supervision, quality and availability, and determines the school structure. This means that the municipalities decide the number and size of public schools. The municipalities also fund the public schools from general local income taxes (OECDa, 2011: 25). As of 2012, the average school size of the 1,318 existing public schools was 426 pupils.\(^1\) Average annual costs per pupil is among the highest in OECD (OECD, 2013: 162-164), and amounts to approximately DKK 60,000, corresponding to about EUR 8,000.

Approximately 42 percent of public expenditure is spent by the municipalities. They are governed by elected city councils. The mayor is elected among local councilors. The structural reform of 2007 changed the map of Danish: 271 municipalities\(^2\) were by law amalgamated into 98. The reform was swift (see, for example, Blom-Hansen et al, 2013; Mouritzen 2010). In 2004 a government commission recommending amalgamations. Six months later, the structural reform was agreed upon in the parliament (Regeringen and Dansk Folkeparti 2004, 9-12). According to the agreement, municipalities with less than 20,000 citizens should be merged with neighbors. Smaller units could be allowed if they agreed with a larger neighboring municipality on certain aspects of service provision. Within these restrictions, the municipalities were free to agree how to amalgamate. Case studies of individual amalgamation processes suggest that particularly questions of local identity, internal cohesion, the likely party political composition of the future municipality, homogeneity in service and wealth and, not least, ambitions of becoming an influential player in the future municipal structure were important determinants (Mouritzen 2006). A quantitative study of potential determinants shows that commuting patterns is the best predictor of amalgamation decisions (Bhatti and Hansen 2011).


\(^2\) Including Ærøskøbing and Marstal, which were amalgamated into Ærø, effective January 1\(^{st}\) 2006.
Figure 2 (panel A) shows the central part of peninsula Jutland in Denmark, before the reform. The municipal structure was quite fragmented.

Figure 2: Changes in municipal borders in central Jutland

Panel A

Panel B

Panel C

Panel D

Note: In 2007, Grauballe Skole (upper left) had 171 pupils, Gjern Skole (upper right) had 360, Søring Skole (lower right) had 158, Voel Skole (center) 199, and Resenbro Skole (lower left) had 143 pupils.

Panel B shows the same geographical area after the reform. The municipal structure is much more consolidated. For example, post-reform municipality of Silkeborg consists of four pre-reform municipalities (Silkeborg, Gjern, Them, and Kjellerup).

We are interested in scale effects in the public school system, or, in other words, in the effect of average school size on spending in a jurisdiction. More specifically, we need an exogenous instrument variable which can predict changes in average school size. We use the reform to obtain
exactly that, based on the idea that the 2007 structural reform implied exogenous change in school redundancy.

Panel D of Figure shows the location of five schools in the municipality of Silkeborg in 2007. Let us, for simplicity, assume that these are the only schools in Silkeborg. We need to compute a redundancy score for the municipality. A high redundancy score must reflect that, in 2007, the school structure has a high likelihood that one or more schools will close in the future, leading to changes in the average school size. A low redundancy score must entail that there is a relatively small probability that one or more schools will close. The redundancy score for municipalities must therefore be a function of the likelihood that schools within the jurisdiction will close. This again is likely to depend on two factors, which we know, as we return to below, must be exogenous: First, the distance from the school to other schools. Schools located very close to other schools are more likely to be redundant. Second, the size of the school, compared to other schools. Smaller schools are more likely to close than larger schools, especially if they are located in close proximity. Intuitively, it should be clear that Voel Skole (located right in the center) is more redundant than, for example, Gjern Skole (located to the North-East, near the border of the municipality). Based on the map, we should expect that it is more likely that the school located in the center of the map will close than this will happen to more isolated schools. Hence, the closer it is to other schools, the more likely it is to close. The smaller the school, the more likely it is to close. And, crucially, these two factors may interact: Relatively small schools, located close to larger schools, are relatively redundant and therefore more exposed to risk of being closed down. To quantify this, we use the following equation for school level redundancy (SR), where \( size_i \) is the number of pupils in school \( i \), and \( distance_{j-i} \) is the distance from school \( i \) to \( j \):

\[
SR_i = \sum_j f(size_i) \times g(distance_{j-i})
\]
Hence, the SR-score is, to capture the interaction effect, a multiple of some function of the two factors, size and distance. We return to the question on the functional form of these functions below. Substantively, the SR score is a measure of the sum of ‘threats’ from all schools within same municipality (based on municipal belonging before/after the reform). The score increases the smaller the school, and the closer it is to other schools. Given this information, it is easy to compute a similar score at the municipal level for municipality k (a municipal level redundancy score, MR_k):

$$MR_k = \sum SR_i$$

The MR-score is a measure of the sum of school level redundancy for all schools in municipality k before and after the 2007 reform. Municipalities with high MR-scores have relatively many small schools, located close to other schools. Municipalities with low MR-scores have few such schools. This is likely to be correlated with future changes in average school size.

The Danish structural reform offers exogenous change in this score. To see this, consider panel (C) of Figure 2. Panel C shows the same geographical as in panel D, but with the municipal structure before the reform in 2007. In 2006, Resenbro Skole was located in the old municipality of Silkeborg, while the other four schools were located in the municipality of Gjern. Hence, although Voel Skole is located very close to Resenbro Skole, they were, before the reform, in different municipalities. This implies, of course, that neither of the schools was highly redundant before the reform. After the reform the schools were both much more redundant. To take an example, SR for Voel Skole increase by 3.78, while SR for Gjern Skole increased by only 0.86.3

Hence, on January 1st 2007, when the reform took effect, exogenously determined changes in SR’s, and therefore also MR’s, occurred. We use changes in MR as an instrument variable. We know that it cannot be affected by the dependent variable (school efficiency), and it is likely that it can predict

---

3 These scores can be computed based on (1) with the functions as specified below. Please note that Silkeborg Kommune has several other schools, so the scores cannot be calculated based on the information in Figure 1 alone.
the independent variable (changes in the average school size) well. We will test the latter presumption below.

Before we can turn to the analyses, two questions must be addressed. First, the functional form of (1) must be specified. Second, we must find sizes of all schools and, for each school, distances from it to all other schools within its municipality. In the following section we show that changes in MR’s ($\Delta MR$) is a valid IV by demonstrating that $\Delta MR$ can actually predict changes in average school size well.

According to (1), the SR for a school depends on the product of some function $f$ of its size and some function $g$ of its distance to other schools within the municipality. It can be practical to normalize these functions. Hence, we would like $f$ to be decreasing in size (since increased size should make the redundancy score lower) and $f(size) \in ]0; 1[$. Similarly, we would like $g$ to be decreasing in distance (since longer distance should also make the redundancy score lower) and $g(distance) \in ]0; 1[$. Furthermore, it seems unrealistic that $f$ and $g$ are linear in size and distance. For example, an extra distance to the nearest school of 5 km is likely to matter much more for schools very close to one another than for schools with a distance between them of, say, 50 km. We therefore choose the following function for $f$:

$$f(size) = \logit(a + b \times size), b > 0$$

This function has useful properties. It is always decreasing in distance, and it is possible by setting the parameters $a$ and $b$ to determine in what ranges of the size $f(size)$ changes the most. Figure 3(a) shows $f(size)$ for $a = -5.5; b = 1.8$. These values were chosen in such a way that $f(size)$ gets close to 0 for very large school sizes (reflecting that very large schools are very unlikely to close, even when near to other schools) and close to 1 for very small school sizes (reflecting that a very small school is very likely to close if it is very near to another school). Furthermore, $a$ and $b$ are chosen such that
\( f(\text{size}) = 0.5 \) for school sizes about 300.\(^4\) These values are of course somewhat arbitrary, and we shall later perform robustness checks with other combinations.

**Figure 3: Functions of size and distance**

![Graphs showing functions of size and distance](image)

We choose, for similar reasons the following functional form for \( g \):

\[
(4) \quad g(\text{distance}) = \text{logit}(c + d \times \text{distance}), \quad d > 0
\]

This function has, of course, properties identical to (3). We choose \( c = -5.5; \ d = 45 \) to ensure that \( g(\text{distance}) \) approaches 1 for very small distances, approaches 0 for very large distances (to reflect that the likelihood of closing is very high for a small school very near to another one, but very small for a school located far away from other schools, even if it is small), and such that \( g(\text{distance}) = 0.5 \) for distances about 8 km.\(^5\) Figure 3(b) shows \( g \) as a function of distance. We perform robustness checks for various combinations of \( c \) and \( d \).

---

\(^4\) A school size of 300 is chosen because smaller schools will not be able to fill classes completely (typically 11 grades, each with one class of 28 pupils).

\(^5\) We chose 8 km because municipalities are required to provide free school transportation for children with more approximately 5 km (on average, depending on age) from their residence to school. This gives municipalities some
Hence, to compute \( \Delta \text{MR} \), we need data for size of all schools at the time of reform and for distance \( d_{ij} \) (the distance from each school to all other schools within the same municipality, immediately before and immediately after the reform). Data for size is available from the Ministry of Education (Databanken: EAK). Data for distance has been obtained in the following way. Information on street addresses and home municipality (before and after the reform) for all schools was collected from the Ministry of Education (the Institution Register) or, alternately Kommunalhåndbogen (2005). Distances for a total number of combinations of 54,514 from each school to each other co-municipality school was computed using Google Maps. We then, using (1) and (2) computed MRs before and after the reform. \( \Delta \text{MR} \) can then easily be found by subtraction. Figure 4 shows a histogram for \( \Delta \text{MR} \).

---

6 Addresses from the Institution Register was validated and corrected by the authors based on information from Kommunalhåndbogen (2007), or, in a few cases, Kommunalhåndbogen (2005).

7 Distance is normalized by subtracting the minimal distance between two schools within a municipality and dividing by the difference between the maximum and minimum distance within the municipality. Google Maps offer distances measured by car, walk or public transportation. We choose walking distance.
It appears from the figure that there is large variation in the changes in the municipal level redundancy scores. The large number of municipalities with zero change in MR (which signifies that changes in average school size, due to exogenous change, is unlikely) reflects that 32 municipalities were left unchanged by the reform.

In the following section we regress municipality level changes in expenditures on changes in school population (the number of schools) using ΔMR as an instrument for changes in school population. We instrument school population rather than average school size, because our instrument is designed to predict changes in school population, not the student population. We therefore control for pupil population (as well as other potential confounders), which we believe are exogenous to changes in expenditures. Before proceeding to this, we recapitulate the argument that the IV is exogenous and demonstrate its relevance.
First, the instrument is exogenous because it is based on two variables, distance and size, measured just before the reform (December 31st 2006) and just after the reform (January 1st 2007). Changes in distance can only be a result of municipal amalgamations. The amalgamations were not, as we have argued above, affected by changes in costs (which is the dependent variable in this study). It must therefore be exogenous. Size does not change at all at the instance of reform. Since it does change, it must be exogenous to changes in costs.

Second, the IV is relevant if it can actually predict changes in school population. While there are good theoretical reasons to expect a correlation between changes in ΔMR and changes in school population, this is indeed a testable empirical question. Table 2 below reports the first stage statistics from two regressions predicting changes in school population with I) changes in MR, and II) changes in MR as well as changes in the control variables included in the IV-estimation reported below.

### Table 2: First stage regressions

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV (ΔMR)</td>
<td>-0.18 (-3.98)***</td>
<td>-0.13 (-3.13)**</td>
</tr>
<tr>
<td>ΔStudent population [in municipality schools]</td>
<td>0.01 (4.29)***</td>
<td></td>
</tr>
<tr>
<td>ΔTax base</td>
<td>-0.00 (-0.10)</td>
<td></td>
</tr>
<tr>
<td>ΔShare of single parent's households</td>
<td>0.44 (1.99)</td>
<td></td>
</tr>
<tr>
<td>ΔImmigrant population [in municipality schools]</td>
<td>0.01* (2.55)</td>
<td></td>
</tr>
<tr>
<td>ΔExpenditures on regional special schools</td>
<td>-0.00 (-0.55)</td>
<td></td>
</tr>
<tr>
<td>ΔExpenditures on municipal special schools</td>
<td>0.00 (0.73)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-1.18 (-3.20)</td>
<td>-1.06 (-0.82)</td>
</tr>
<tr>
<td>IV-Partial R2</td>
<td>0.13</td>
<td>0.10</td>
</tr>
<tr>
<td>IV-F-statistic</td>
<td>15.88***</td>
<td>9.80**</td>
</tr>
<tr>
<td>N</td>
<td>98</td>
<td>98</td>
</tr>
</tbody>
</table>

Notes: ***; **; *: p < 0.001; 0.01; 0.05. Unstandardized coefficients with t-statistics in parentheses.
From the table it is evident that the IV is significantly related to changes in school population. Positive changes in MR lead to fewer schools, according to these models. While the F-statistic falls just short of the recommended threshold of .10 (Sovey & Green, 2011), the IV seems relevant.

Because the exact definition of the IV is set somewhat arbitrarily (it depends on the four parameters mentioned) we have performed a robustness analysis to assess the sensitivity. Specifically, we have randomly sampled 4,000 sets of the four parameters randomly from four uniform distributions, using our parameters as mean and +/- 20% as range. Using the 4,000 sets of parameters, we have calculated the average F-statistic from models corresponding to Model II in Table 2. The average F-statistic is XX, with an interquartile range of XX. This suggests that the IV is indeed relevant and not sensible to our arbitrary set of parameters. [We still have to perform the robustness analysis using the precise set of parameters reported above, but previous robustness analyses using a slightly different set of parameters suggest robust instrument relevance].

The Causal Impact of School Size on Costs
The dependent variable is measured as the total costs of public schools. Since the independent variable (changes in school population) is based on a first-difference transformation, we also measure the dependent variable (and the controls) using first differences. Before a school closure can take place, a political decision and preparations concerning where to place the pupils and teachers from the closed school have to be made. Moreover, municipal amalgamations may make school closures politically costly in the short term because of geographical cleavages in the newly formed municipalities. (Baekgaard 2010). In conjunction, these factors render school closures that are brought about due to municipal amalgamations unlikely in the short term and we therefore measure the first differences as the difference between the years 2007 and 2011. In more mundane terms our regression model therefore estimates whether exogenous changes in school population bring about changes in municipal public school expenditures from 2007 to 2011.
Also, a number of controls are included in the analysis. First, both the size of schools and the total expenditures are likely to be affected by the total number of pupils enrolled in public schools.

Second, the municipalities may decide to send the weakest pupils (with physical or mental disabilities) to either of two kinds of so-called special schools (either run by the regions or municipalities). While this maneuver may increase total municipal school expenditures it decreases costs on public schools. We therefore add two controls to account for such substitution effects.

Third, three controls are included to control for changes in municipal socio-economic conditions. Specifically we account for changes in the municipal tax base, the percentage of single parent households (a commonly used proxy for parent resources) and the number of immigrants from non-western countries. Summary statistics on the variables included in the analysis are shown in Table 3.

**Table 3: Summary statistics and data sources.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>Min.</th>
<th>Max.</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔPublic school expenditures (dependent variable)</td>
<td>-29,746</td>
<td>28,284</td>
<td>-144,614</td>
<td>27,152</td>
<td>Statistics Denmark. Sum of functions 3.22.01 and 3.22.02.</td>
</tr>
<tr>
<td>ΔNumber of schools</td>
<td>-2.1</td>
<td>3.0</td>
<td>-18</td>
<td>1</td>
<td>The Ministry of Education</td>
</tr>
<tr>
<td>ΔStudent population [in municipality schools]</td>
<td>-213.5</td>
<td>365.5</td>
<td>-1,182</td>
<td>988</td>
<td>The Ministry of Education</td>
</tr>
<tr>
<td>ΔTax base</td>
<td>16,893</td>
<td>3,654</td>
<td>8,008</td>
<td>29,079</td>
<td>Ministry of Economic Affairs and the Interior (<a href="http://www.noegletal.dk">www.noegletal.dk</a>)</td>
</tr>
<tr>
<td>ΔShare of single parent's households (%)</td>
<td>0.6</td>
<td>1.1</td>
<td>-2.5</td>
<td>5.7</td>
<td>Ministry of Economic Affairs and the Interior (<a href="http://www.noegletal.dk">www.noegletal.dk</a>)</td>
</tr>
<tr>
<td>ΔImmigrant population [in municipality schools]</td>
<td>-3.7</td>
<td>61.4</td>
<td>-328</td>
<td>135</td>
<td>Statistics Denmark. (Table Folk 2).</td>
</tr>
<tr>
<td>ΔExpenditures on regional special schools</td>
<td>-3,943</td>
<td>12,090</td>
<td>-80,963</td>
<td>15,614</td>
<td>Statistics Denmark Function 3.22.07.</td>
</tr>
<tr>
<td>ΔExpenditures on municipal special schools</td>
<td>17,747</td>
<td>23,322</td>
<td>-40,767</td>
<td>130,102</td>
<td>Statistics Denmark Function 3.22.08.</td>
</tr>
</tbody>
</table>

Notes: All cost related variables are measured in 1,000’s of Danish kroner and in 2010-prices, except Tax base, which is measured in Danish kroner (2010 prices).
It appears from the summary statistics that there is great variation on the dependent variable with changes in public school expenditures spanning from a decrease of 145 million Danish kroner to an increase of more than 27 million kroner. It also follows from the table that the number of public schools has decreased substantially from 2007-2011 with a mean of approximately two school closures. The findings from the analysis are presented in Table 4. For reasons of comparison, Model 1 presents the findings from a first difference estimation in which the change in the number of schools is not instrumented, while Model 2 contains the findings from the two stage least squares regression.\(^8\)

**Table 4: School size and expenditures**

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta ) Number of schools</td>
<td>498.65 (0.59)</td>
<td>6959.34 (2.10)*</td>
</tr>
<tr>
<td>(\Delta ) Student population [in municipality schools]</td>
<td>26.12 (3.73)***</td>
<td>-0.58 (-0.04)</td>
</tr>
<tr>
<td>(\Delta ) Tax base</td>
<td>0.37 (0.61)</td>
<td>0.25 (0.34)</td>
</tr>
<tr>
<td>(\Delta ) Share of single parent's households</td>
<td>355.21 (0.18)</td>
<td>-2440.10 (-0.89)</td>
</tr>
<tr>
<td>(\Delta ) Immigrant population [in municipality schools]</td>
<td>-24.68 (-0.68)</td>
<td>-83.87 (-1.58)</td>
</tr>
<tr>
<td>(\Delta ) Expenditures on regional special schools</td>
<td>-0.79 (-3.74)***</td>
<td>-0.75 (-2.88)**</td>
</tr>
<tr>
<td>(\Delta ) Expenditures on municipal special schools</td>
<td>-0.84 (-7.58)***</td>
<td>-0.90 (-6.46)***</td>
</tr>
<tr>
<td>Constant</td>
<td>-17889.18 (-1.66)</td>
<td>-5076.59 (-0.35)</td>
</tr>
<tr>
<td>Estimation method</td>
<td>FD - OLS</td>
<td>FD - 2SLS</td>
</tr>
<tr>
<td>R²</td>
<td>0.51</td>
<td>0.19</td>
</tr>
<tr>
<td>F-statistic</td>
<td>13.15***</td>
<td></td>
</tr>
<tr>
<td>Wald Chi(^2)</td>
<td>64.97***</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>98</td>
<td>98</td>
</tr>
</tbody>
</table>

Notes: ***; **; *: p < 0.001; 0.01; 0.05. Unstandardized coefficients with t-statistics in parentheses.

\(^8\) Using fixed effects instead of first differences produces a null finding with regard to the correlation between the number of schools and public school expenditures. Since this specification does not deal with the fundamental problem of simultaneity, the findings are not presented here.
Based on Model 1 it appears that changes in the number of schools do not affect changes in municipal public school costs. This finding may however, be conflated by simultaneity bias. Model 2 instead shows a positive impact of the number of schools on public school expenditures. This also means that creating larger schools by closing some schools and allocating their pupils among the remaining schools is associated with cost savings, or, put differently, that there are indeed economies of scale in the production of Danish public school services. In an even broader sense it can be inferred from this finding that larger jurisdictions are indeed favorable to smaller ones when it comes to public school costs because larger jurisdictions have the opportunity of creating a more efficient school structure than in a more fragmented local government setting.

The estimate corresponds to almost seven million Danish kroner per school which is a considerable amount bearing in mind that this amount may be saved each year if the municipality is able to close just one school. The difference in findings between Model 1 and 2 may be explained by simultaneity bias conflating the impact of changes in the number of public schools in Model 1. Additional tests of whether the number of schools is indeed endogenous to the costs lend support to this explanation as the proposition that the independent variables are exogenous is rejected in a Durbin-Wu-Hausman test (p < .02). This indeed also suggests that simultaneity bias should be taken seriously when analyzing the impact of government design on costs.

**Conclusion**

The ambitious question about the optimal design of government may be impossible to answer for a variety of reasons. According to existing research (Dahl & Tufte 1973) there is a fundamental tradeoff between democratic and economic concerns and even if focus is solely on economic concerns, scale effects may differ for different public goods (Treisman 2007) suggesting that the efficient size and structure of government may differ for different services. One way to handle this
problem is to look for simpler and more modest questions to answer; that is questions that are not subject to the problem that the efficient governmental design differs across services.

The analysis presented here illustrates that it is indeed possible to provide an answer to the simpler question about the economically efficient design of a specific important service: public school services. From a methodological point of view this is still a complex question because an answer to it critically rests on proper control for simultaneity bias. The lack of proper control may indeed be the reason that previous research on scale effects in public schools yields ambiguous results. However, when such a control is made, the analysis suggests that there are real economic gains associated with larger local governments because larger local governments are able to organize more efficiently by closing schools that are redundant in a more fragmented structure. Thus, municipal size appears to have a negative impact on public school costs. While this finding is based on an analysis of Danish municipalities, the logic behind the scale effects detected here appears to have a more universal character. Thus, there is no reason to believe that this logic should not apply to single purpose districts, for instance, and we may also expect similar negative scale effects on service areas where increasing the jurisdiction size also means increasing the opportunities of closing redundant units. Examples of such services could be day care centers or hospitals.

However, while public schools are certainly important, this finding is obviously only part of the story regardless of whether focus is on economic scale effects specifically or the impact of jurisdiction size in a broader sense. Thus, future research should focus on further development of theory about the most important outcomes of jurisdiction size and empirical analyses in which account is taken for simultaneity bias should be conducted on other service areas and with other outcome variables than economic ones in order to increase our knowledge about how jurisdiction size furthers some objectives and hinders others.
References


