doi:10.1017/S000305541100013X

Jurisdiction Size and Local Democracy: Evidence on Internal Political Efficacy from Large-scale Municipal Reform

DAVID DREYER LASSEN *University of Copenhagen*
SØREN SERRITZLEW *Aarhus University*

Optimal jurisdiction size is a cornerstone of government design. A strong tradition in political thought argues that democracy thrives in smaller jurisdictions, but existing studies of the effects of jurisdiction size, mostly cross-sectional in nature, yield ambiguous results due to sorting effects and problems of endogeneity. We focus on internal political efficacy, a psychological condition that many see as necessary for high-quality participatory democracy. We identify a quasiexperiment, a large-scale municipal reform in Denmark, which allows us to estimate a causal effect of jurisdiction size on internal political efficacy. The reform, affecting some municipalities, but not all, was implemented by the central government, and resulted in exogenous, and substantial, changes in municipal population size. Based on survey data collected before and after the reform, we find, using various difference-in-difference and matching estimators, that jurisdiction size has a causal and sizeable detrimental effect on citizens' internal political efficacy.

For centuries, scholars have debated how the size of political jurisdictions affects the quality of government. This question can be traced back almost 2,500 years to Plato and Aristotle, who preferred smaller entities, large enough to be self-sufficient, but small enough to ensure that citizens can know one another's characters (Dahl and Tufte 1973, 4–5). Today, the debate over the number and size of political jurisdictions is framed in terms of decentralization, federalism, and local and urban politics. Small is often perceived as beautiful because, it is argued, citizens are closer to decision makers and feel more efficacious, politics is less abstract, and preferences more homogeneous, allowing for a better fit between what citizens want and what they get. Large is lovely because larger jurisdictions are more likely to be self-sustainable, can be responsible for broader ranges of political questions, and can benefit from economies of scale.[1] Although the question of size and democracy is old and prominent, empirical studies of the effect of size on democracy and its underpinnings are strikingly scarce.

In this article, we focus on one psychological condition that many see as necessary for high-quality participatory democracy: internal political efficacy (IPE), or individual citizens' beliefs that they are competent to understand and take part in politics. Since the 1950s, scholars have studied the determinants and consequences of IPE (Almond and Verba 1963; Campbell, Gurin, and Miller 1954; Finifter 1970; Hayes and Bean 1993; Morrell 2003; Verba and Nie 1972). Using original survey data from a unique quasiexperiment conducted in Denmark since 2001, we provide evidence that increasing the size of local jurisdictions has a causal effect on IPE. As the size of municipalities increases, citizens feel less qualified to play an active role in municipal politics.

The optimal size of a jurisdiction is a fundamental question in political science. Since Plato and Aristotle, political philosophers, political scientists, and economists have debated how size is related to democratic participation and the efficient production of collective goods. Plato argued that the optimal size of a state should be "sufficient to maintain a certain number of inhabitants in a moderate way of life" (Plato, *Book of Laws* V), and even suggested that 5,040 citizens would be appropriate. Also, Aristotle argued that a population must be of a certain magnitude to "be self-sufficient for the purpose of achieving a good way of life," but that a state should be small enough to ensure that the citizens can "distribute offices of government according to the merit of the candidates" and "know one another's characters" (Aristotle 1948, 292). Dahl and Tufte (1973, 4–12) show how the question has remained relevant to thinkers in the history of political thought from Jean-Jacques Rousseau, who argued that the average citizen's share in decisions must vary inversely with the size of the polity, to John Stuart Mill, who found that the need for the whole people to participate implies that the perfect government must be

David Dreyer Lassen is Professor of Economics, University of Copenhagen, Øster Farigmagsgade 5, Building 26, DK-1353 Copenhagen K, Denmark (david.dreyer.lassen@econ.ku.dk).

Søren Serritzlew is Professor of Political Science, Aarhus University, Bartholins Allé 7, DK-8000 Aarhus C, Denmark (soren@ps.au.dk).

This research is partly based on data collected in the project Size and Democracy (Kjær and Mouritzen 2003). We extend special thanks to Poul Erik Mouritzen for generously providing us with access to his data. We also thank KREVI (Danish Evaluation Institute for Local Government) and our respective departments for funding, and Lotte Bøgh Andersen, Jens Blom-Hansen, Camilla Dalsgaard, Rasmus Dørken, Jørgen Elklit, Peter Holdt-Olesen, Ulrik Kjær, Søren Leth-Petersen, Søren Rud Kristensen, Asbjørn Sonne Nørgaard, Palle Svensson, Jens Peter Frølund Thomsen, and Lise Togeby; *APSR* coeditor Daniel Treisman; his journal coeditors; and three referees for valuable comments and suggestions. Peter B. Mortensen gave us particularly thorough and useful comments on several drafts. Conference and seminar participants at Midwest Political Science Association 2008, Nuffield College, University of Southern Denmark, Aarhus University, and University of Copenhagen also provided very useful feedback.

[1] Treisman (2007, 11–15) provides a detailed discussion of these arguments for and against decentralization. The rest of his book expands on this and conducts a critical study of the economics and politics of decentralization.

representative. Of course, the question of size is also prevalent in more recent discussions. Robert A. Dahl, in his presidential address to the American Political Science Association's annual meeting in 1967, argued that the optimum size of a city would be between 50,000 and 200,000 citizens (Dahl 1967, 965). Dahl and Tufte (1973, 3) give a clear account of the central dilemma in the question of the effect of size: Smaller polities provide better opportunity to participate, but larger polities offer citizens the opportunity to influence a wider range of potentially more important political decisions.

Although the question of whether size affects IPE is both simple and important, the empirical literature is inconclusive. Of six studies addressing the question, three find that jurisdiction size is not, or almost not, correlated with political efficacy (Almond and Verba 1963, 234; Lyons and Lowery 1989; Muller 1970, 797), and three that jurisdiction size is negatively associated with efficacy (Finifter 1970; Finifter and Abramson 1975; Vetter 2002, 13). The fact that all six studies are cross-sectional constitutes a major methodological challenge. Cross-sectional estimates may reflect the deliberate decisions of people with differing interests to sort themselves into jurisdictions with different characteristics, as emphasized by Tiebout (1956). This makes it difficult to interpret existing empirical estimates as reflecting a causal effect of jurisdiction size. At the same time, time variation in jurisdiction structure is rarely observed, and studies that make use of such variation when analyzing local political issues, including DeHoog, Lowery, and Lyons (1990), are subject to sample selection due to the voluntary political nature of most amalgamations and mergers.

To get at estimates of causal effects in this setting, we turn to administrative reform. As noted by Campbell (1969), administrative reforms occasionally provide researchers in the social sciences with opportunities to estimate causal effects by generating quasiexperiments that exogenously determine, or at least significantly influence, the treatment assignment in a nonexperimental setting (Meyer 1995; Shadish, Cook, and Campbell 2002). We exploit a recent large-scale administrative reform of the Danish municipal structure, the so-called Structural Reform (SR). Through amalgamation with one or more neighboring municipalities, individuals residing in reform (treatment) municipalities suddenly experienced a dramatic increase in the population size of their jurisdiction, whereas residents in nonreform (control) municipalities experienced no change in jurisdiction size. The fact that 33 municipalities were not amalgamated enables us to control for exogenous national shocks and national-level trends in overall survey response and local political interest. The latter could well be affected due to both the public debate on the large-scale reform and the concurrent revision of the system of intergovernmental grants. The reform was, as we argue in detail later, largely exogenous from the viewpoint of individual municipalities and definitely so for individual citizens. The empirical analysis focuses on the short-run effects of reform in an attempt to identify the pure effect of change in size.

We estimate the causal effect of jurisdiction size on IPE using individual-level survey data collected before and after the reform in both unaffected (control) and affected (treatment) municipalities. We find, using various difference-in-difference (DiD) and matching estimators on both repeated cross-sectional data and retrospective evaluations, that citizens in amalgamated municipalities report lower IPE after the reform, and, crucially, that the decline is larger, the larger the change in population size as a result of the reform. Our preferred specification suggests that the largest experienced increases in population size translate into an estimated decrease in IPE of 1.7 on a scale from 2 to 10; to put this into perspective, this corresponds approximately to a drop from the 75th percentile to the median or from the median to the 20th percentile. The result that population size has a causal, significant effect on IPE is robust across samples and estimators, and we demonstrate that local variations in the amalgamation process, as well as changes in local public finances and municipal political control following reform, do not affect this relationship.

JURISDICTION SIZE AND INTERNAL POLITICAL EFFICACY

Various scholars have argued that high-quality democracy requires that citizens participate actively in politics, above and beyond voting in elections or referenda. In a healthy democracy, in this view, individuals should attend and speak at public meetings, petition officials, and seek to influence decisions in other ways. Two conditions are necessary, although, of course, not sufficient, for citizens to participate. First, they must believe that the government authorities are responsive to citizens' demands, so that their participation is likely to achieve something. Second, they must believe that they are competent to understand and contribute effectively to political decision making. Together, these beliefs are often referred to as political efficacy. The first is known as external political efficacy, whereas the second is known as internal political efficacy.

Building on the work of Almond and Verba (1963), Dahl and Tufte (1973) argue that citizens tend to feel more competent to participate in government in small jurisdictions. Simply because any particular citizen's share in power declines when the number of citizens increases, it is reasonable to expect that "[t]he larger the citizen body, the weaker the sense of individual effectiveness (the greater the sense of powerlessness)" (43). In small jurisdictions, it is easier to "know the ropes of politics," and it is possible to focus on the local community (Verba and Nie 1972, 231). In addition, size is likely to affect the nature of political discussions. In small jurisdictions, discussions often concern very specific questions. In large jurisdictions, it is impossible to discuss every single detail; political discussions must be more general to be tractable. Consider local political discussions on primary schools, which in many countries is one of the most important tasks of local government. Small municipalities with, say, 5,000

citizens, corresponding to the smallest included in this study, have perhaps two or three schools, and debates on school policy often concern a specific school. To be effective, arguments must be specific and related to particular problems in one of the schools. Larger municipalities with perhaps 500,000 citizens, corresponding to the largest municipality included in this study, have more than 50 public schools, and school policy is typically more general. Arguments referring to particular problems in a particular school often drown in a flood of ideas and complaints. To be effective, arguments should apply more generally and generate a larger coalition extending beyond the boundaries of just one school district.

Dahl and Tufte (1973, 61–65) conclude, based on a review of existing empirical studies, that feelings of efficacy are lower in large jurisdictions. Although Almond and Verba (1963) find that political efficacy does not depend on city size (see also Muller 1970, 797), Finifter and Abramson (1975) argue that this conclusion is misleading; because education levels tend to be higher in larger cities, and because education is positively related to the sense of political efficacy, the true relationship between city size and political efficacy may be suppressed. When education is controlled for, it is revealed that city size is, in fact, negatively related to feelings of IPE (Finifter 1970, 403–4; Finifter and Abramson 1975, 194). More recent studies contribute to this somewhat blurry picture; Lyons and Lowery (1989) find no effects, whereas Vetter (2002, 13) finds a small effect.

In the next section, we discuss the concept of IPE, how it is defined, and how it is related to other important aspects of a well-functioning democracy. We then discuss methodological problems endemic to existing empirical studies of the effect of size on efficacy.

Concept of Internal Political Efficacy

The concept of political efficacy was originally defined as the "feeling that individual political action does have, or can have, an impact upon the political process, i.e., that it is worthwhile to perform one's civic duties" (Campbell, Gurin, and Miller 1954, 187). As noted by Lane (1961, 149), however, this definition includes two dimensions of efficacy: one related to self-image, which is closely related to the concept of subjective political competence analyzed by Almond and Verba (1963, 181–213), and one to the image of democratic government. The two dimensions are referred to as internal political efficacy, defined as "beliefs about one's own competence to understand, and to participate effectively in, politics," and external political efficacy, which is "beliefs about the responsiveness of governmental authorities and institutions to citizen demands" (Niemi, Craig, and Mattei 1991, 1407). Since the 1960s, internal political efficacy has been a central measure in the literature; Morrell (2003, 591) has thus identified more than 75 studies of its measurement, causes, and consequences.

The feeling of political efficacy is positively correlated with participation (Clarke and Acock 1989; Pollock 1983, 404), but although empirically related, feelings of efficacy and actual participation are two distinct phenomena. First, just that people feel more competent to participate does not imply that they will participate more. Despite confidence in their own competence, they may believe that government is unresponsive (i.e., they may have low external political efficacy), or they may simply be too busy or uninterested in participating. It is also perfectly possible to feel efficacious and have plenty of time without having particular reasons to participate. We do not have data on whether individuals actually participate less in the large amalgamated jurisdictions of Denmark or whether governments in small jurisdictions are more responsive to citizen demands. Second, feelings of IPE and rates of participation have different implications for democracy. Almond and Verba (1963, 255–57) conclude that the feeling of efficacy, compared to actual participation, which may also have a destabilizing effect, has several desirable effects: It increases the legitimacy of a system; it leads to political stability, greater acceptance of participatory democratic systems, and higher valuation of the participation of fellow citizens; and it "may be taken as an index of the extent to which they consider their nation democratic." Self-confident citizens are also more likely to be active (Almond and Verba 1963, 257; Pollock 1983) and more satisfied (DeHoog, Lowery, and Lyons 1990), and a well-functioning participatory democracy increases feelings of efficacy (Clarke and Acock 1989; Madsen 1987). When it comes to the quality of a democratic order, the question of whether citizens believe that they have adequate possibilities to participate in and influence political decision making, rather than whether they believe that they have sufficiently weighty reasons to actually participate, is crucial. According to Dahl (1989, 109), the central question is whether citizens "have adequate and equal opportunities for placing questions on the agenda and for expressing reasons for endorsing one outcome rather than another."

IPE has been measured in a number of ways. Five survey questions, labeled Qualified, Understand, Public office, Informed, and Complex, have been used in different combinations. We use these questions, translated into Danish (see Lolle 2003) and adapted to local politics (see Table 2 for wordings). Niemi, Craig, and Mattei (1991) recommend using Qualified, Understand, Public office, and Informed, but not Complex. Morrell (2003, 592) identifies four frequently used ways of measuring IPE. He concludes that the four-item index developed by Niemi, Craig, and Mattei (1991) is among the most popular, and that it consistently performs as a reliable and valid measure (Morrell 2003, 600). We call this measure *IPE-Niemi*. The survey question "Complex," which is not employed in IPE-Niemi, is used much more in the American National Election Studies (ANES), where it, unlike the other questions, has appeared since 1952 (ANES 1994, 504)—and, in contrast to Qualified, Public office, and Informed, still appears. Together with "Understand," this question offers an alternative and, we believe, more robust measure because it does not involve indirect evaluations of the

respondent's organizational and managerial capacity. We call this measure *IPE-Complex* and employ both measures in the empirical analysis that appears later in this article, commenting on differences where appropriate.

Empirical studies of the causes of individual citizens' sense of IPE show the importance of several individual-level factors, including gender, age, education, and income (Almond and Verba 1963; Hayes and Bean 1993, 269–70; Titus 1981). As seen, although the results of existing studies are not unequivocal, the effect of size on internal political efficacy, if any, is negative. This also applies to other aspects of democracy. For example, Verba and Nie (1972, 237–47) conclude that participation tends to be higher in small local government units. Dahl and Tufte (1973, 63) report evidence from Swedish local government that supports the size–participation relationship. Following Dahl and Tufte, a large number of studies have investigated the effects of jurisdiction size and consolidation on democracy broadly construed; some with specific focus on size, and some as parts of studies of decentralization more broadly.[2] In a study of the longitudinal relationship between aggregate turnout and waves of municipal consolidation, Morlan (1984) finds that turnout is higher in national elections than in local elections. However, at the local level, turnout is inversely related to jurisdiction size (see also, among others, Larsen 2002 and Oliver 2000). Jurisdiction size is also negatively related to citizens' likelihood of contacting officials and politicians (Verba and Nie 1972, 231), as well as to attending meetings on local issues or in local organizations (Larsen 2002; Oliver 2000).

To sum up, many cross-sectional, and a few longitudinal, studies investigate the democratic consequences of differences in jurisdiction size, and there is some, but by no means unequivocal, evidence that larger size is associated with lower quality, however measured, of democracy and its underpinnings. However, these studies face two methodological challenges.

A Simple Idea, But a Difficult Test

The idea that jurisdiction size matters for IPE in a causal sense is simple, but due to two thorny methodological challenges, it is difficult to evaluate empirically. First, differences in observable, but unmeasured or poorly measured population characteristics can affect estimates, as shown in Finifter and Abramson (1975). The populations of small and large jurisdictions can be different on many dimensions, which causes problems of common support rarely addressed in this literature. Education is one important example, but also income levels, occupations, and employment patterns are different in small and large jurisdictions. This problem can be partly solved by statistical control for the confounding variables, but a fully satisfactory solution requires that all such variables are known, identified, and measurable.

Second, cross-sectional studies cannot address problems of selection that arise if people with unobservable preferences over local democracy self-select into jurisdictions of different size, partly with an eye on exactly this. The idea that people may sort themselves into different jurisdictions is classic (Tiebout 1956), and solid evidence for sorting based on preferences for local public goods such as school quality and environmental quality is beginning to emerge (Banzhaf and Walsh 2008). This constitutes a problem of endogeneity. When sorting effects exist, differences in political efficacy among citizens residing in jurisdictions of varying size may reflect reverse causality (i.e., that people with interest in local politics are drawn to smaller jurisdictions), rather than a causal effect of jurisdiction size on IPE.

The problem of endogeneity is difficult to handle statistically. One way to correct for endogeneity is to use an instrumental variable approach, but valid instruments can be difficult to find, and we have not seen any studies on size and democracy that attempt this. Another solution would be to study jurisdiction size and the quality of local democracy over time, but this is often infeasible because important jurisdictional boundaries rarely change. When change is observed in connection with large-scale reform, two additional complications arise. First, in large-scale reforms, several factors, such as local political institutions or systems of intergovernmental grants, often change simultaneously, making it difficult to isolate the effects of changing jurisdiction size. Second, such reforms are often encompassing, leaving no individuals or jurisdictions unaffected.

In addition, when jurisdictional change occurs, or is refrained from, it typically does so endogenously. In many settings, jurisdictional change, whether at the city (Lassen 2005), county (Alesina, Baqir, and Hoxby 2004), or country level (Alesina and Spolaore 1997), is often decided by voting. Alesina, Baqir, and Hoxby (2004) find that jurisdictions in the United States forego consolidation by voting against it, if it means increasing population heterogeneity. Such majority consent is necessary for consolidation in most countries and introduces the issue of selection bias into estimates of the effect of jurisdiction size on local democracy indicators.[3]

To identify a causal effect of the size of the populace on IPE, it is therefore necessary to identify a quasiexperiment that provides an exogenous source of variation in jurisdiction size. In the next section, we introduce the SR of Danish municipalities and argue that this is an excellent opportunity to test how jurisdiction size affects IPE.

[2] Treisman (2007) reports results on the relationship between country-level measures of (fiscal) decentralization and broad, expert-based measures of democracy.

[3] For a general criticism of using endogenously decided changes, see Besley and Case (2000). A large literature on urban economics and politics in the United States considers both case and large *N* studies of various consolidation efforts (e.g., Benton and Gamble 1983; DeHoog, Lowery, and Lyons 1990), but these efforts are prone to this type of selection bias.

DANISH STRUCTURAL REFORM AS A QUASIEXPERIMENT

The SR, implemented politically in 2005 and administratively in January 2007, was an exogenous shock to the size of Danish municipalities. The SR was planned and implemented by the Danish central government. Of the 271 original municipalities, 238 were merged into 65. The 33 municipalities that were not amalgamated allow us to control for exogenous national shocks and national-level trends in local political interest. Hence, by exposing citizens to shocks of different sizes, the SR introduced an exogenous element into the size of the local polity. We use this exogeneity for identification of causal effects.

Danish Local Government

The local level is a vital part of the Danish public sector, both culturally and politically. Culturally, a majority of Danes identify primarily with their local area, rather than with their region or country.[4] Politically, Danish local governments are responsible for about 40% of total Danish public expenditure and are relatively autonomous. They are, within certain restrictions, free to set the local income tax rate and have wide discretion in most policy areas. Most notably, they control welfare services such as the school system, child care, and elderly care. The municipalities are governed by a city council elected for 4-year terms in popular elections. Turnout is generally high, typically 70% to 80%. The mayor is elected indirectly by the city council. In addition to presiding over the council and being the formal head of the municipality, the mayor is the only full-time politician, head of the local administration, and chairman of the economic committee, which assumes a central role in the political process (Berg and Kjær 2005). The local party systems are typically dominated by local branches of the national parties, but there are some so-called local lists.

Just as in Almond and Verba's results, the sense of political efficacy in Denmark tends to be higher toward the local level than toward the national level and the European Union, although the differences between the local and the national levels are limited (Andersen 2000, 128–30). In a thorough cross-sectional study of the effect of jurisdiction size on IPE in Danish municipalities, Lolle (2003) concludes that jurisdiction size does not have a strong effect on political efficacy. The association between characteristics of citizens and their sense of political efficacy is much stronger. In particular, education is strongly correlated with IPE (Andersen 2000, 141–42).

Structural Reform of 2007

The SR has three parts. First, the number of municipalities was reduced from 271 to 98, increasing average jurisdiction size from 20,100 to 55,600 citizens. Second, 14 counties were merged into five new administrative units, so-called regions. Third, municipalities were given additional responsibilities (most important, certain social services for the disabled, health-related prevention, and environmental regulation), whereas tax collection was transferred away from municipalities. The net transfer of tasks to the municipalities was limited and amounts to DKR 25 billion compared to the total municipal expenditure of (in 2007) DKR 333 billion.

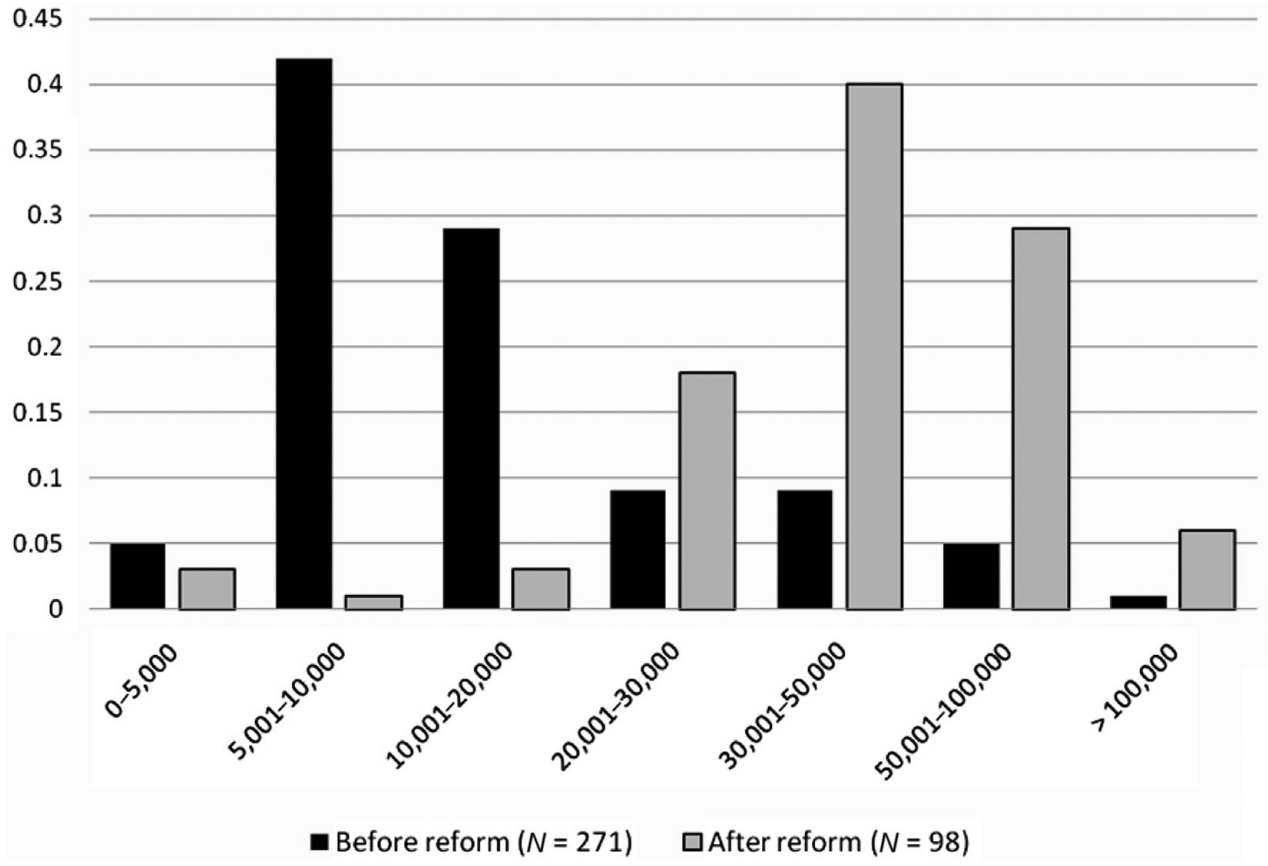
The idea of reforming Danish local government was first mentioned in 2002. The SR was enacted in 2004, and took effect starting on January 1, 2007. The reform process was initiated by the central government, with the explicit aim to generate larger municipalities to reap economies of scale in local public goods production. The process began in 2002, when the government appointed the *Commission on Administrative Structure* with a chairman appointed by the Minister of the Interior, three independent expert members, four representatives from municipalities and counties, and four civil servants from Danish ministries. In January 2004, the committee published its four-volume report, which recommended a large-scale structural reform. Based on this, the center-right minority government introduced a bill, which immediately got a majority when the Danish People's Party declared that it had only one question: Where do we sign? In June 2004, it was clear that Danish local government would experience dramatic reform. It is the fact that the SR was imposed on municipalities by the central government, combined with the fact that not all municipalities were equally affected by the SR, that makes it an attractive setting for studying the causal effect of jurisdiction size on the quality of local democracy.

According to the agreement between the government and the Danish People's Party, municipalities would be merged into larger units with at least 30,000 citizens. In practice, all municipalities with less than 20,000 citizens (205 of the 271 municipalities) were asked to arrange for mergers before January 1, 2005. Only neighboring municipalities would be allowed to merge; however, apart from this general restriction, the central government did not choose the merging partners.[5] Already before this agreement, 11 of the new 98 municipalities had announced how they wanted to merge, and most of the remaining municipalities were very close to agreeing on whom to merge with (Blom-Hansen, Elklit, and Serritzlew 2006, 17–18). A few municipalities, especially islands, would be allowed to continue with less than 20,000 citizens, but would have to cooperate with larger neighbors on

[4] World Value Survey, 1999 wave. The majority of respondents (55.8%) answered "Locality" to a question regarding "Geographical groups belonging to first."

[5] See Christiansen and Klitgaard (2008, 152–54), who document the reform process in detail in the book *The Unimaginable Reform*. An exception to this rule is the municipalities of *Farum* and *Hvorslev*, which had difficulty finding merging partners, and *Holmsland* and *Værløse*, which wanted to continue without merging, but were not allowed to, due to their size. They are not part of this study. When the reform debate began in 2002, five small municipalities on the island of Bornholm voluntarily merged into one islandwide municipality, reducing the total number of municipalities from 275 to 271. These municipalities are not part of our sample either.

FIGURE 1. Danish Municipalities before and after 2007 Reform; Relative Shares by Population



service provision. The end result is seen in Figure 1: many small municipalities before the reform, whereas larger municipalities dominate after the reform.

A Quasiexperiment

The SR changed the municipal structure, notably the size of jurisdictions, in a way that would not have occurred otherwise. In this sense, the SR constitutes an exogenous shock to the size of more than 200 municipalities, and we use this exogeneity to break the correspondence between preferences and choice, which leads to selection problems in observational data. We classify this type of intervention as a quasiexperiment, following Campbell (1969), but note, as Barabas and Jerit (2010, 227), that the literature does not agree on experimental terminology. For example, some see natural experiments as synonymous with quasiexperiments, defining as "[g]ood natural experiments ... studies in which there is a transparent source of exogenous variation in the explanatory variables that determine the treatment assignment" (Meyer 1995, 151), whereas others, for example, Shadish, Cook, and Campbell (2002), distinguish between natural experiments, reserved for a "naturally occurring contrast between a treatment and a comparison condition" (17) and quasiexperiments, with the latter term now being reserved for "experiment[s] in which units are not assigned to conditions randomly" (12). Angrist and Krueger (2001) share, without mentioning quasiexperiments explicitly, Meyer's sentiment, writing that natural experiments are "situations where the forces of nature or government policy have conspired to produce an environment somewhat akin to a randomized experiment" (73). In sum, our opinion is that because the SR provides exogenous variation in the explanatory variable, but is neither a naturally occurring event nor a source of random shocks, it constitutes a quasiexperiment rather that a natural experiment.

This exogenous variation is a crucial component in our research design. However, for us to be able to interpret any postreform differences in IPE as being caused by larger jurisdiction size, we need to address three additional issues. First, differences unrelated to jurisdiction size between small and large municipalities could imply a problem of comparability. Second, citizens relocating as a result of the reform could be a source of endogeneity. Third, change itself could affect feelings of IPE. We discuss these issues in turn.

The problem of comparability will arise if, for example, larger jurisdictions are subject to different institutional constraints, tasks of larger units are different or more complex in nature, or harder to understand because these factors alone could affect feelings of IPE. However, due to regulation from the central government, Danish municipalities have the same set of

tasks, are organized in similar ways, and follow identical rules for the democratic process. The differences between large and small municipalities have to do with the process citizens need to engage in if they aspire to influence political decisions. Although the formal democratic rights of citizens are identical, strategies that are effective in small jurisdictions are likely to fail in large jurisdictions. For example, parallel to the argument presented previously, organizing protests among citizens affiliated with a certain municipal institution would be fairly easy in a small municipality and could quickly involve quite a large proportion of the electorate. Simple messages on specific problems concerning this particular institution are likely to be heard. However, such strategies would often be inadequate in a large municipality. Protests from users of one particular institution do not carry much weight, and, to be effective, citizens need to organize much broader, principled protests. This requires the citizen to be capable of formulating an abstract political problem. In a large municipality, a promising strategy would be to mobilize support from interest groups. Although the formal democratic rights are identical, and the tasks of large and small municipalities are of the same inherent complexity, the political process may well be more complex and difficult for many citizens to understand in larger jurisdictions. Hence, if changes in jurisdiction size affect IPE, this is not the result of differences between the tasks of small and large municipalities.

The second issue is the possibility that individual citizens respond to the new municipal structure by relocating, which would be a source of endogeneity. Individuals with strong preferences for political participation at the local level who suddenly find themselves in a large municipality may move to a small municipality. In practice, we do not consider this a major problem because the postreform survey was conducted within a year after the reform, and we exclude recent movers from the analysis. In addition, although individuals were not allowed to choose whether and how their municipality should be merged, the municipalities and their citizens did have some influence on the merger. For example, municipalities of citizens with a strong preference for local democracy may have aimed for a merger with a small municipality. In our analysis, this would make citizens of such municipalities relatively happier with the amalgamation, which would bias the analysis against finding a detrimental effect of the reform on IPE. Finally, although the assignment of the treatment is not random, the municipalities were severely constrained by the size requirements and the condition that merging municipalities should be geographically connected. Bhatti and Hansen (2011) show that, apart from the obvious geographic constraints, social connectedness, measured by shared commuters, is the most important predictor of mergers such that socially connected municipalities are more likely to merge. Hence, the nonrandomness of the treatment may imply that merged municipalities are more connected and coherent. If this causes bias, then the bias will again be against finding a negative effect of size.

Some insight into the process of amalgamation can be obtained from the fact that referenda were held in 63 of the original 271 municipalities (Jørgensen 2006, 168). Most of the referenda were concerned with adjustments of boundaries, but some involved the choice of which municipality to merge with. In our sample of 32 municipalities, referenda took place in six cases, marked with asterisks in Table A.1 in the Appendix. In these cases, the group of citizens did have a direct influence on what would happen to their municipality, but individual citizens, of course, only had a minuscule chance of affecting the result. We observe no differences in results.

The final issue is that change in itself may alter feelings of IPE. The major components of the SR that did not involve changing jurisdictional boundaries are, as already discussed, accounted for by having a control group in the sample. However, for the amalgamated municipalities, we can decompose the jurisdictional change into two components: the fact that jurisdictional boundaries change at all and the fact that population size increases. If citizens react in a negative way to change per se, then any change in jurisdiction size, both positive and negative, should be associated with lower political efficacy. If population size is more important and has a systematic effect on political efficacy, then increasing and decreasing population size should have differing effects, and the effect should be larger, the larger the change in population. Although we distinguish amalgamated municipalities by the change in size when conducting the analysis, we do not observe jurisdictions decreasing in size as a consequence of the SR.

Evidence against the hypothesis that change in itself matters can be drawn from an experiment involving decentralization of the municipality of Copenhagen, the capital and by far the largest municipality in Denmark, into 15 administrative and political districts. Four city districts were exogenously selected as pilot districts, providing a natural experiment for the entire city to learn about the consequences of decentralization; Lassen (2005) documents the exogenous selection process, although with focus on information. In surveys carried out after a citywide referendum on implementation of the city district model, survey respondents overwhelmingly favored smaller units with respect to political efficacy, both internal and external, even if the proposal eventually failed due to concerns about economic efficiency (Lassen 2005). The fact that the selection process was exogenous, allowing citizens to learn about the consequences of smaller jurisdictions in an unbiased way, as well as the finding that citizens overwhelmingly favored change rather than the status quo when considering IPE, in combination support the hypothesis that changes in population size rather than change per se cause lower IPE in large jurisdictions.

DATA

We use two individual-level telephone surveys, conducted by professional polling firms, one in 2001, before the SR, and the other in late 2007 and early 2008, a year after the SR was implemented. A core set of identical

TABLE 1. Strata in the Post-SR Survey

Stratum	Merged?	Pre-SR Size	Post-SR Share	No. of Units	Share of Sample (%)	Share Complete Answers (%)
1: Not merged	No	N/A	N/A	8	25.0	25.1
2: Small-Small part	Yes	< Median	< Median	6	18.7	18.7
3: Small-Large part	Yes	< Median	> Median	6	18.8	18.7
4: Large-Small part	Yes	> Median	< Median	6	18.8	18.7
5: Large-Large part	Yes	> Median	> Median	6	18.8	18.8
Total				**32**	**100.0**	**100.0**

Note: SR, Structural Reform.

questions were asked in both surveys, making it possible to compare pre- and post-SR response patterns.

In the 2001 survey, 2,763 interviews were obtained. The response rate was 44%. An equal number of respondents were randomly selected in 60 of the then 275 Danish municipalities (Houlberg and Pedersen 2003, 209). The 60 municipalities were randomly selected from six strata based on size, with 10 municipalities from each. The 2007 survey resulted in 1,255 interviews. The response rate was 48%.[6]

Because our aim is to compare the two surveys, only residents aged 24 (who were 18 when the first survey was carried out) or older were included in the postreform sample. Younger respondents were excluded to rule out any cohort effects. The 60 municipalities in the 2001 survey were divided into five strata, one with continuing municipalities and four with reform municipalities. This latter group was constructed by combining below and above median size before the reform with below and above median share of the population size of the new municipality after the reform. We randomly selected eight municipalities from the control group stratum and six from each of the other four strata. The 2007 survey was stratified by pre-SR municipalities, making it possible to compare municipal-level responses in pre-SR municipalities within the same area, now part of a new, larger postreform municipality. Table 1 shows the strata; the 32 selected municipalities are listed in Table A.1 in the Appendix.

Municipalities belonging to the stratum denoted "small-small part" in Table 1 were below median size before the reform and ended up being a small (below median) part of the new municipality as a consequence of the reform, and likewise for the other strata. In the following, we refer to these strata as small-small (SS), small-large (SL), large-small (LS), and large-large (LL). Table 2 shows descriptive statistics on the final samples and definitions of the variables used in the following analyses. The prereform sample includes only municipalities sampled postreform. Both samples contain only individuals with responses to all efficacy questions; recent movers are excluded in both samples, as are the very young in the postreform sample, as described previously.

[6] Response rates are calculated using RR4 in American Association for Public Opinion Research (2008). For further information on the prereform survey, see Houlberg and Pedersen (2003, 207).

RESEARCH DESIGN AND EMPIRICAL SPECIFICATION

The exogenous shock to jurisdiction size implied by the reform constitutes, together with the repeated survey data, the basis of our research design. To estimate the causal effect of jurisdiction size on IPE, we use the DiD method on repeated cross-sectional data. DiD is well known in the program evaluation literature (see, e.g., Imbens and Wooldridge 2009, for a recent treatment). Under DiD, units receiving treatment are compared to their pretreatment levels, and the same is done for a control group. This yields the first level of differences. Subsequently, the two differences are subtracted from each other, resulting in a second-level difference, leaving an estimate of the causal effect of treatment under certain assumptions, which are detailed in this section.

The Difference-in-Difference Estimator

Let the internal efficacy of an individual i be Y_i^1 in the treated case, that of municipal amalgamation, and Y_i^0 be the outcome for the same person in case of no treatment. Our object of interest is the difference $\Delta_i = Y_i^1 - Y_i^0$, which is the causal effect of municipal population size on the IPE experienced by individual i. In practice, however, an individual i cannot be treated and nontreated at the same time, which means that either Y_i^1 or Y_i^0 will be missing. The difference Δ_i results from the pair of potential outcomes, a term coined by Rubin (1974), which is to be distinguished from the realized outcome Y_i. Because a counterfactual for the realized outcome is not available at the individual level, the statistical approach is to estimate the missing variable from appropriate group means.

We study a repeated cross-section model of different individuals surveyed before and after the reform:

$$Y_i = \alpha + \gamma_M M_i + \gamma_P T_i + \tau_{DiD} M_i T_i + \beta' X_i + u_i \quad \forall i \in I,$$

where I is the set of respondents. $M_i \in \{0, 1\}$ is an indicator for reform (as opposed to continuing)

TABLE 2. Questions, Definitions, and Descriptive Statistics

Variable	Question or Definition	Pre-SR Survey					Post-SR Survey				
		N	Mean	SD	Min	Max	*N*	Mean	SD	Min	Max
Independent variables											
Vocational training	Respondent has completed vocational training	963	0.27	0.45	0	1	1,065	0.35	0.48	0	1
College degree	Respondent has completed 2- to 4-year college	963	0.20	0.40	0	1	1,065	0.25	0.44	0	1
M.Sc. or above	Respondent has received Master's degree or higher	963	0.09	0.28	0	1	1,065	0.12	0.33	0	1
Income	Pretax income of household in categories	963	2.82	1.20	1	5	1,065	2.68	1.27	1	5
Publicly employed	Employed in the public sector (0 = no, 1 = yes)	963	0.37	0.48	0	1	1,065	0.25	0.43	0	1
Gender	Gender of respondent (1 = female, 0 = male)	963	0.47	0.50	0	1	1,065	0.49	0.50	0	1
Age	Age of respondent at time of survey	963	47.77	13.69	17	91	1,065	54.25	14.69	24	95
Dependent variables											
IPE-Complex	Additive index of Understand and Complex	963	7.41	2.04	2	10	1,065	7.07	1.82	2	10
IPE-Niemi	Additive index of Qualified, Understand, Public office, and Informed	963	13.47	3.69	5	20	1,065	12.78	3.26	4	20
IPE-Complex, retrosp.				N/A			1,035	6.70	1.35	2	10
IPE-Niemi, retrosp.				N/A			1,035	13.09	2.36	5	20
Complex	Sometimes local politics seems so complicated that a person like me can't really understand what's going on	963	3.46	1.45	1	5	1,065	3.20	1.35	1	5
Understand	I feel that I have a pretty good understanding of the important political issues facing our municipality	963	3.96	1.10	1	5	1,065	3.87	1.02	1	5
Qualified	I consider myself to be well qualified to participate in municipal politics	963	3.01	1.56	1	5	1,065	2.81	1.40	1	5
Public office	I feel that I could do as good a job as a local councilor as most other people	963	3.02	1.59	1	5	1,065	2.91	1.39	1	5
Informed	How well would you say that you are informed about local politics in your municipality?	963	3.48	0.88	1	5	1,065	3.19	1.02	1	5

Note: IPE, internal political efficacy; SR, Structural Reform.

municipalities. Later, we also consider a treatment indicator of the form $M_i \in \{0, SS, SL, LS, LL\}$, where we distinguish treated municipalities both by their prereform size and by their postreform relative size, as explained previously. $T_i \in \{0, 1\}$ is a time period indicator equal to 0 before the reform and 1 after the reform. The interaction(s) $M_i T_i$ take(s) on the value 1 for reform municipalities after implementation, and zero otherwise. Finally, u_i represents unobservable characteristics and is assumed to be independent of treatment status conditional on observables and across periods; we return critically to this assumption later.

It can be useful to consider the conditional means for the four groups: reform and nonreform municipalities before and after reform:

$$\begin{aligned}
&(\mathrm{A})\overline{Y}_{00} = E(Y|M=0, T=0, X) = \alpha + \beta' X \\
&(\mathrm{B})\overline{Y}_{01} = E(Y|M=0, T=1, X) = \alpha + \beta' X + \gamma_P \\
&(\mathrm{C})\overline{Y}_{10} = E(Y|M=1, T=0, X) = \alpha + \beta' X + \gamma_M \\
&(\mathrm{D})\overline{Y}_{11} = E(Y|M=1, T=1, X) \\
&\qquad = \alpha + \beta' X + \gamma_P + \gamma_M + \tau_{DiD}. \qquad \textbf{(1)}
\end{aligned}$$

The before–after estimator for the treated (BAT), used when no control group is available, is given by (1.D) – (1.C):

$$\overline{Y}_{11} - \overline{Y}_{10} = \gamma_P + \tau_{DiD}. \qquad \textbf{(2)}$$

The BAT estimator includes both the effect of treatment and the common trend or aggregate change. In our case, γ_P includes the effects on IPE from overall trends and aspects of the reform common to all municipalities. To estimate consistently the effect of size, we need to subtract this common trend, which is equal to the before–after estimator of the control group (BAC):

$$\overline{Y}_{01} - \overline{Y}_{00} = \gamma_P. \qquad \textbf{(3)}$$

The BAC estimates the effect of reallocation of government tasks and any aggregate time trends, possibly due to changing survey participation, in IPE with respect to local government; the latter may have increased or decreased in reaction to the substantial nationwide public debate on municipal politics and the SR. Similarly, increased fiscal pressure will also be captured by this trend, as long as it is similar across all municipalities. Based on our setup in Eq. (1), we get the DiD estimator as BAT – BAC; that is,

$$\tau_{DiD} = (\overline{Y}_{11} - \overline{Y}_{10}) - (\overline{Y}_{01} - \overline{Y}_{00}). \qquad \textbf{(4)}$$

The assumption that the trend in Eq. (2) is common for treatment and control groups alike is a key identifying assumption in the DiD approach (Blundell and MaCurdy 1999). We first proceed under the assumption that this assumption is satisfied; for example, as noted previously, the reallocation of government tasks across vertical levels of government was the same for all municipalities. However, if the impact of the reform is heterogeneous with respect to observable characteristics, and the distributions of such observable characteristics differ across control and treatment groups, then additional assumptions are needed to ensure that a comparison is based on suitably similar groups; we return to this later, where we consider matching. The second key identifying assumption, denoted no composition bias (Blundell and MaCurdy 1999, 1612), is that the populations considered are the same across time. At the population level, this means that no individuals change from treatment to control group and vice versa. This is sometimes problematic in studies of individual responses to changes in taxation or job training programs; however, in our case, this would be directly relevant only if people moved with the explicit aim of living in a treatment or control municipality, which seems highly unlikely. In addition, our empirical analysis excludes respondents who recently relocated from another municipality. The assumption is more problematic when we consider the actual sampling of individuals; we return to this later in the article.

When estimating individual-level responses to aggregate-level variables, here treatment status and other municipal-level variables, it is important to account for the possible covariance in errors. We do this by allowing for clustering at the municipal level.

Nonrandom Assignment: Different Surveys and Differences in Observables across Treatment Groups

In existing DiD studies on repeated cross-sectional data, the analysis is typically carried out on different waves of a particular representative survey, such as the General Social Survey. In our case, we compare IPE measures across surveys carried out in the same municipalities, using identically worded questions, but with different sampling strategies. The existence of differential sampling probabilities across surveys raises the problem that the surveys are not stratified in the same way, which means that controlling for confounding variables may not solve the problem of composition bias, the absence of which, as noted previously, is necessary for identification of causal effects in a DiD model. To solve this problem, we want to match respondents across surveys on a number of individual-level variables.

At the same time, treatment and control municipalities may differ with respect to observable characteristics, which can lead to a violation of the common trend assumption if effects of the reform are heterogeneous across observables. As noted previously, for example, an explicit aim of the SR was to create municipalities with a minimum of 30,000 inhabitants, which in itself affects the probability of treatment. To solve this issue, we want to match survey respondents across treatment status. In sum, there are two assignments that may be nonrandom: assignment to treatment or control group and assignment to prereform or postreform survey.

To address this, we employ a variant of difference-in-difference with matching (DiDM), an estimator developed in the program evaluation literature by Heckman, Ichimura, and Todd (1997) and Heckman et al. (1998).

We proceed by propensity score matching (see, e.g., Dehejia and Wahba 2002). Although intuitively desirable, exact matching on a vector of characteristics is often infeasible. However, Rosenbaum and Rubin (1983) show that if potential outcomes are independent of treatment conditional on X, which is a precondition for estimating a causal effect from the data anyway, then this is also the case conditional on the propensity score, defined as the probability of receiving treatment conditional on X. Because we have two assignments that are potentially nonrandom, we follow Blundell et al. (2004) in defining two propensity scores. Conditional on X, $P_M = \Pr(M = 1|X)$ is the probability of being observed in a treated municipality and $P_T = \Pr(T = 1|X)$ is the probability of being surveyed in the postreform sample. Based on the propensity scores, we can now write the identifying common trend assumption as

$$\begin{aligned} &E\left(Y_i^0|P_M, P_T, M = 1, T = 1\right) \\ &\quad - E\left(Y_i^0|P_M, P_T, M = 1, T = 0\right) \\ &= E\left(Y_i^0|P_M, P_T, M = 0, T = 1\right) \\ &\quad - E\left(Y_i^0|P_M, P_T, M = 0, T = 0\right). \end{aligned}$$

This allows the time effects to differ by X and assures that the distribution of observed characteristics is the same across all four samples. In practice, we estimate the two propensity scores by running a probit regression of assignment status (treatment and time, respectively) on X. The resulting predicted values are the estimated propensity scores. Our point of departure is the group of individuals in the treated group, pretreatment ($M = 1$, $T = 0$). To match these individuals with respondents in each of the three comparison groups, we perform a variant of nearest-neighbor caliper matching with replacement by pairing each treated respondent with a respondent in each comparison group. The respondent in each comparison group is chosen to minimize the Euclidean distance from the treated individual's two estimated propensity scores, given restrictions on the maximum allowable distance (the caliper) as well as a common support restriction. Based on these matched samples, we carry out standard DiD analysis to form estimates of the average treatment effect on the treated (ATT) under the assumption of separable additivity of time and group effects.[7] Because matching is carried out with replacement, we correct the sample sizes of the comparison groups to reflect this procedure; the distribution across municipalities within each of the four groups is shown in Table A.1 in the Appendix for both the regression and the matching analyses.

As is standard in the literature, we report bootstrapped standard errors throughout for the estimated ATTs based on the DiDM procedure. One study shows that this may, in fact, result in biased estimates for the standard error, but the issue is not resolved (Abadie and Imbens 2008). In performing the bootstrap, we resample at the cluster level of new municipalities, as previously described.

Balance, Unobserved Heterogeneity, and Exact Matching

Although the DiDM procedure allows for unobserved heterogeneity under the assumption of a common trend conditional on X and, as such, represents a considerable improvement of the use of standard cross-section estimators, two potential sources of bias remain. First, remaining unobserved heterogeneity can obviously influence results; and second, to the extent that propensity score matching does not achieve perfect balance across treatment regimes on variables that are strong predictors of the outcome, observed differences in outcomes may be erroneously attributed to treatment, in which case it may be efficient to match directly, or exactly, on these key variables (Barnard et al. 2002; Rosenbaum and Rubin 1983).

We explicitly include population size as a matching criterion in the calculation of propensity scores to reflect decision making, and, as a consequence, the matching procedure allows us to compare individuals in both small and large municipalities. However, if there are systematic differences in unobserved attitudes toward and interest in local politics among citizens in small and large municipalities, which we suspect ex ante, then including population size alongside individual-level variables in the calculation of propensity scores may not be a sufficient safeguard because population size may carry little weight in the calculation of the propensity scores. To address this, we carry out additional analyses where individuals in small (large) treatment municipalities are explicitly restricted to be matched with individuals in small (large) control municipalities. As emphasized by Heckman, Ichimura, and Todd (1997) and Cook, Shadish, and Wong (2008), DiDM functions best when matching individuals who reside within a given geographic unit and, if possible, who are administered identical surveys. This is what we do here.

Finally, to address concerns arising from the less than perfect balance of covariates in the main matching specification, we carry out propensity score matching under the restriction of exact matching on public sector employment; this being an example of an important explanatory variable for which it is difficult to achieve balance in the simple propensity score specification.[8] In practice, this requires the matching procedure to look

[7] This is reminiscent of, but not identical to, the preprocessing of data suggested by Ho et al. (2007, 205). As they note, their procedure is not directly applicable in the case of multiple treatments.

[8] We owe the idea of exact matching on key variables to one of the referees.

for matches based on the two-dimensional propensity score only among individuals with the same type of employment.[9]

Retrospective Evaluations

To supplement the evidence from the repeated cross sections, we also asked postreform survey respondents to assess, in retrospect, the consequences of the SR for measures of IPE. Although there are well-known issues with retrospective questions of this kind, they allow us to address unobserved heterogeneity in a different way: By asking one person to evaluate the differences before and after the reform, we essentially control for individual-level factors and get a direct estimate of the first-level differences. Hence, the retrospective evaluations serve as a robustness check. The limitations of retrospective questions are of a different nature than the potential methodological problems of comparing answers across samples.

We can now compare these differences across treatment and control municipalities as well as within the group of treatment municipalities. We analyze these data using both a regression-based, cross-sectional framework and matching across treatment and control groups. Throughout, we take into account the clustered nature of the data by allowing for within-municipality dependence in estimated standard errors.

RESULTS

In this section, we first present the qualitative insights from the analyses and then illustrate the substantive significance of the results in a separate section.

Regression-based Difference-in-Difference Estimates

Table 3 reports results from the standard, regression-based DiD analysis for four specifications: analysis of binary, categorical, and two continuous treatments on the full sample. The second column reports results for the IPE-Complex measure, whereas the third column reports results from identical specifications using the IPE-Niemi as the dependent variable. This yields a total of eight regression analyses. Standard control variables were included, but results for these are not shown; a full table for IPE-Complex is available in Table A.2 in the Appendix.

The results for the binary treatment—shown as model 1 at the top of Table 3, where citizens in all reform municipalities are considered identically treated—suggest strong average effects of the increase in population size. This is the estimated coefficient on the combined interaction term. It measures the average effect of the change in population size on citizens' feelings of IPE in treated municipalities. We estimate the DiD effect to be −0.36 for the full sample. This estimate is significant at the 10% level.[10]

TABLE 3. Effect of Municipal Amalgamation and Size on Political Efficacy: Evidence from Difference-in-Difference Analysis

	IPE-Complex	IPE-Niemi
Model 1: Binary treatment		
Amalgamated municipality	−0.36	−0.48
	[0.23]*	[0.24]*
Model 2: Categorical treatment		
Small-Small	−0.43	−0.84
	[0.23]*	[0.28]***
Small-Large	−0.23	−0.43
	[0.27]	[0.29]
Large-Small	−0.44	−0.31
	[0.23]*	[0.38]
Large-Large	−0.35	−0.34
	[0.25]	[0.40]
Model 3: Absolute population difference	−0.08	−0.09
	[0.04]**	[0.05]*
Model 4: Relative population difference	−0.04	−0.08
	[0.02]*	[0.03]***
N	2,028	2,028

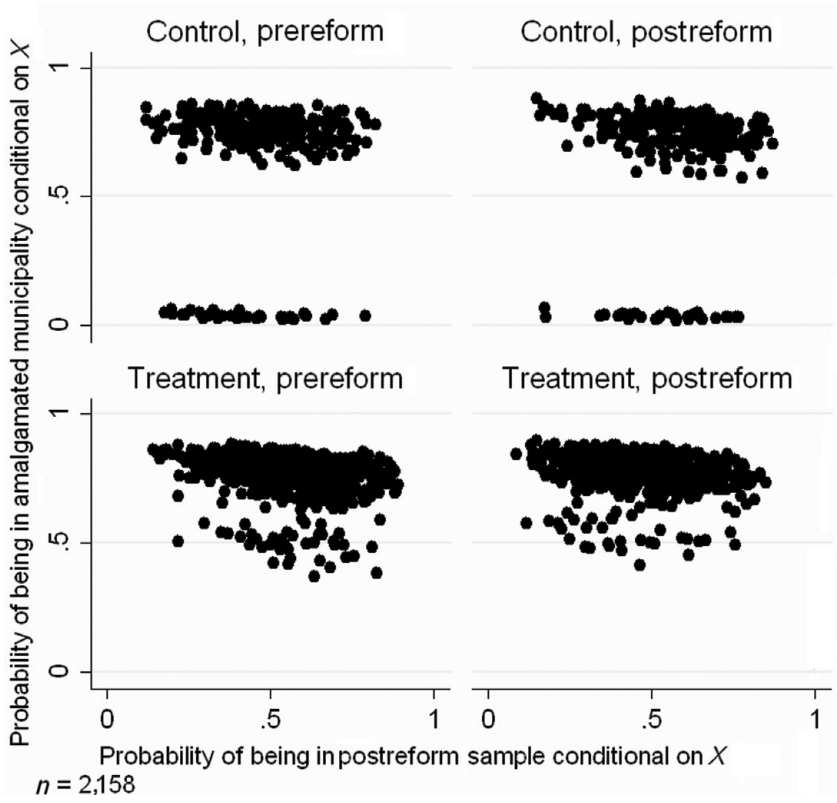
Notes: Standard errors are corrected for clustering at the municipal level. The entries show beta coefficients and associated standard errors [in brackets] for two dependent variables (in columns) in four models (rows). Model 1 compares amalgamated municipalities (=1) to continuing, unaffected municipalities (=0). Model 2 distinguishes between amalgamated municipalities by their pre- and postreform size and compares them to continuing, unaffected municipalities (cf. Table 1). Model 3 shows the effect of the absolute population difference (in ten thousands) between a postreform municipality and a prereform municipality. Model 4 shows the relative population difference, calculated as the absolute population difference divided by prereform population. Each model shows the results of a difference-in-difference linear regression specification. Each model includes the same set of control variables (not shown in the table). Full results including controls for IPE-Complex are available in the Appendix; for IPE-Niemi, they are available from the authors on request. IPE, internal political efficacy.
*Significant at 10%; ** significant at 5%; *** significant at 1%.

Model 2 presents results from the categorical treatment. In this case, we distinguish municipalities by prereform size and postreform relative size, creating the four categories Small-Small, Small-Large, Large-Small, and Large-Large presented in Table 1. Municipalities in all four categories experienced an increase in size. Municipalities in Small-Large and Small-Small used to be small prereform. These municipalities tend to experience a large increase in population size in relative terms. Municipalities in Large-Small and Small-Small constitute a small part of the newly formed municipality (postreform). These municipalities tend to

[9] Our approach is similar in spirit to the *coarsened exact matching* procedure proposed by Iacus, King, and Porro (2009), but adapted to our setting of two-dimensional treatments.

[10] In model 1 of Table A.2 in the Appendix, the coefficient of the treatment category is insignificant and close to zero, suggesting no prereform differences between control and treatment municipalities.

FIGURE 2. Distribution of Propensity Scores across Treatment Status and Scores



experience a large increase in population size in absolute terms. This way of looking at the data shows reasonably similar results for the four treatment categories: Respondents living in municipalities that have seen a large absolute population increase (Small-Small and Large-Small) experience the largest declines in IPE, significant at the 5% level, but respondents in the two other categories also experience declines.

Finally, models 3 and 4 report results from continuous treatment specifications, where we consider directly the effect of the change in population size experienced by respondents. In models 3 and 4, the absolute change and the relative change in population size, respectively, are used. Population size difference corresponds, by definition, in a monotonic way to the categorical treatment variables used in model 2, and we find similar results: the loss of IPE experienced by respondents is increasing in the size of the population increase, both absolute and relative, with the former being more precisely estimated and significant at the 5% level.[11] This is fully consistent with the results for the four treatment categories. We return to the interpretation of the substantive effects later in the article, following the results on matching.

Results from Difference-in-Difference with Matching

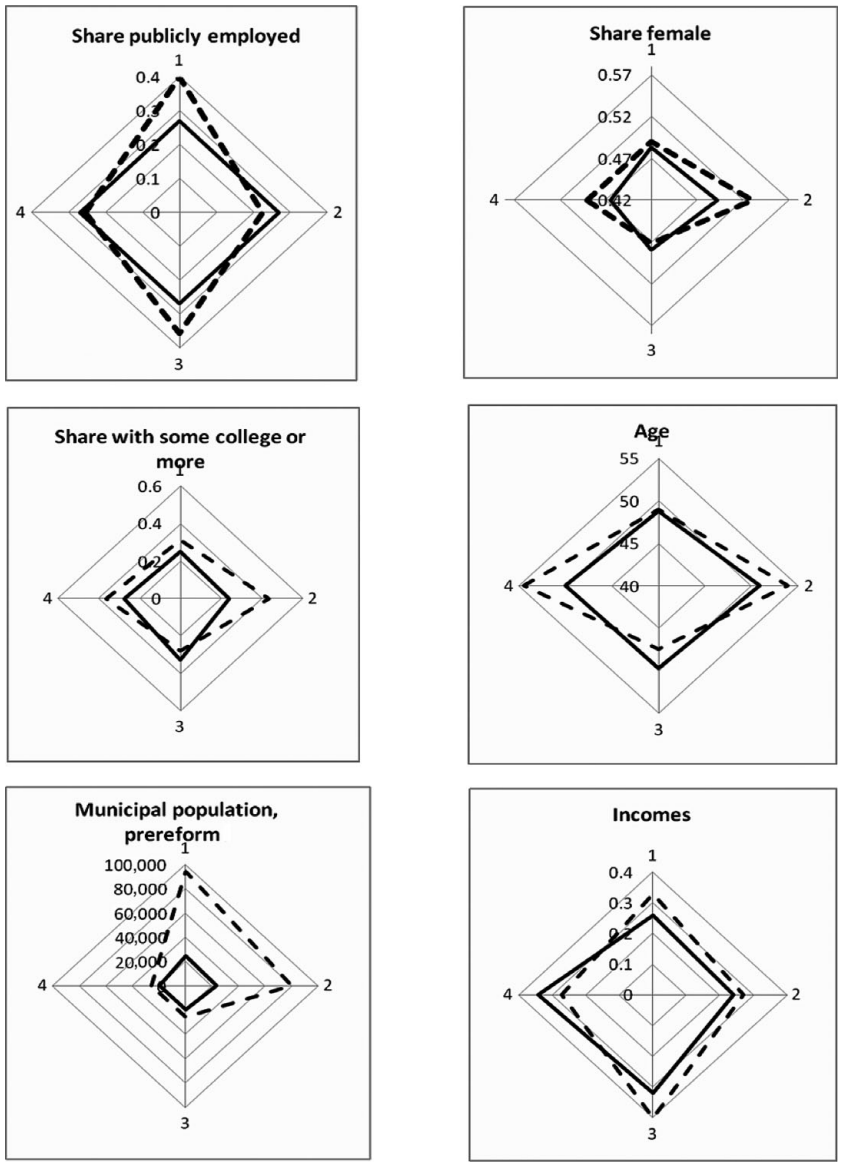
Figure 2 presents scatter plots of the two propensity scores. Each observation in Figure 2 is an individual, and we observe a wide range of observed propensity scores in both dimensions, documenting substantial differences in the probability of being observed in samples both across time and across treatment status.[12]

As described previously, each observation in the treatment prereform group is matched to its nearest neighbor measured by the Euclidean distance in each of the three other groups, subject to a maximum distance requirement, set to 0.05, and the exact matching requirement on public sector employment; furthermore, the samples are trimmed to the common support. This procedure results in 1,148 observations being dropped from the main sample as used in Table 3, leaving 880 of which some will be weighted several times as sampling was done with replacement; this number is sensitive to the choice of caliper, but, as shown later in

[11] As noted, some municipalities held consultative referenda as part of the reform process. This is potentially important because the very act of a referendum may have increased the voters' feeling of legitimacy of the final amalgamation decision, which could affect their sense of IPE. Moreover, the decision to have a referendum in the first place could reflect a municipality more interested in local politics and, indeed, citizens in the six referendum municipalities in our sample did report higher IPE than citizens in nonreferendum municipalities in the prereform survey. Regression results suggest no differences across samples (available on request).

[12] Individuals with almost zero probability of being observed in the reform municipalities are living in Copenhagen, which differs in size from most of the other municipalities in the sample by a factor of 20.

FIGURE 3. Balance of Covariates, before and after Matching



Notes: Dashed line, before matching; solid line, after matching. Box shows average values of key covariates for control/treatment groups before and after reform. The closer a box is to being a square, the closer are average values of the covariates and the better is the match between treatment and control groups and before and after reform. Group 1, control group before reform; group 2, control group after reform; group 3, treatment group before reform; group 4, treatment group after reform.

the article, results are not. In standard one-dimensional propensity score analysis, it is common to focus on the postmatching densities of propensity scores and the resulting covariate balance between treatment and control groups, testing for significant differences in the latter. However, this is less straightforward in the present case, with four comparison groups and two propensity scores. To give some sense of the change in balance as a result of the matching procedure, consider Figure 3.

In Figure 3, we present covariate balance before and after matching using a radar plot. In a perfectly balanced sample, the plotted figure is a (tilted) square; the further away from a square the quadrangle, the worse the balance. Consider the bottom graph on the left: The dashed line indicates that mean populations in the control group samples before and after reform (groups 1 and 2) are 90,000 and 80,000, respectively, and approximately 25,000 in the treatment groups. Postmatching covariate means, illustrated by the solid line, show an almost perfect square, with mean population for each of the four groups around 20,000. Not surprisingly, balance on public sector employment is perfect following exact matching and improves on all important variables. Education remains less than perfectly balanced; we return to this later, when we consider robustness issues.

TABLE 4. Effect of Municipal Amalgamation and Size on Political Efficacy: Evidence from Difference-in-Difference with Matching

	IPE-Complex	IPE-Niemi
Model 1: Binary treatment		
Amalgamated municipality	−1.00	−1.24
	[0.30]***	[0.52]**
Model 2: Categorical treatment		
Small-Small	−1.38	−2.08
	[0.30]***	[0.66]***
Small-Large	−0.93	−1.06
	[0.37]**	[0.87]
Large-Small	−0.97	−1.31
	[0.33]***	[0.83]
Large-Large	−0.68	−0.34
	[0.48]	[0.40]
Model 3: Absolute population difference	−0.23	−0.27
	[0.06]***	[0.13]**
Model 4: Relative population difference	−0.14	−0.22
	[0.04]***	[0.07]***
N	1,532	1,532

Notes: Standard errors are corrected for clustering at the municipal level, and bootstrapped with 50 replications and resampling at the cluster level. The entries show beta coefficients and associated standard errors [in brackets] for two dependent variables (in columns) in four models (rows). Model 1 compares amalgamated municipalities (=1) to continuing, unaffected municipalities (=0). Model 2 distinguishes between amalgamated municipalities by their pre- and postreform size and compares them to continuing, unaffected municipalities (cf. Table 1). Model 3 shows the effect of the absolute population difference (in ten thousands) between a postreform municipality and a prereform municipality. Model 4 shows the relative population difference, calculated as the absolute population difference divided by prereform population. Each model shows the results of a difference-in-difference with matching specification. The matching is based on exact matching on public sector employment combined with two-dimensional Euclidian metric matching with caliper = 0.05. IPE, internal political efficacy.
*Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4 shows results from the matching analysis for IPE-Complex (second column) and IPE-Niemi (third column) following the template from Table 3. The average effect for IPE-Complex equals −1.0, which is significant at the 1% level and almost three times the estimate from the full-sample, regression-based result reported in Table 3. Here, the categorical treatment suggests larger differences across categories: The largest effect, which is estimated with considerable precision, is for individuals from small prereform municipalities. The pattern is generally the same as in the regression-based analysis, with smaller and less precisely estimated effects for citizens living in municipalities with little change ex post, as their prereform municipality remained the larger one.

A similar picture emerges from the continuous measures: Estimates of the effect of population difference and relative population difference both increase by approximately a factor of 3 compared to the full sample case. These results are strongly statistically significant at the 1% level. In unreported results, we find the effects on the individual questions Understand and Complex to be almost exactly identical, both in terms of size and statistical significance. In the third column, we show results for IPE-Niemi. The results are qualitatively similar to those identified for IPE-Complex. They are numerically only slightly larger, which indicates a smaller effect on these dimensions of efficacy, because the scale is now different. The relative differences between the DiDM estimates and the full-sample results are similar to those identified for IPE-Complex. The differences in size arise largely from the fact that the variable Public office does not show up significant in analyses of the individual efficacy components.

The overall conclusion is clear: The qualitative results of the regression-based DiD analysis remain valid, but taking into account nonrandom assignment across samples, in both the treatment and the time dimensions, increases the estimated effects of population size on IPE considerably. What accounts for the observed differences in results? Two things occur in the matching analysis compared to the regression analysis: The sample is different, and the estimation method, based on nearest neighbor matching, is nonparametric, dispensing with the assumptions on functional form inherent in the linear model. If we apply the linear regression model of Table 3 to the matched sample obtained previously, then we observe results essentially similar to those identified by the matching analysis, suggesting that the differences in results are a consequence of the different samples, not the different estimation methods.

Next, we attempt to gauge the importance of possible remaining unobservable heterogeneity. To do this, we match on prereform municipality size (in addition to public sector employment, as previously). This means that respondents in treatment municipalities that were small before the reform are required to be matched to respondents in municipalities of comparable size. In practice, for the nonreform group, this means the municipalities Læsø, Dragør, and Vallensbæk (see Table A.1 in the Appendix). The remaining control municipalities are used for matching respondents in large prereform municipalities. For the treatment municipalities, we use our classification of prereform small and large. The result is a considerably reduced sample of 602 observations compared to the 1,532 in the main matching analysis, down from 2,028 in the regression specification. The results are shown as models 1 and 2 in Table 5: The estimates are only slightly lower than those reported in Table 4 for both efficacy measures, suggesting that if additional unobserved heterogeneity relating to municipality size is present, it does not substantially affect results.

Finally, we consider two robustness results. First, in models 3 and 4, we carry out exact matching on both public sector employment and education because education is the only variable with less than perfect balance in Figure 3. Again, results are unaffected. Second, in models 5 and 6, we choose a stricter matching criterion, a maximum distance equal to 0.025 instead of 0.05, which tends to increase estimates even further.

TABLE 5. Effect of Municipal Amalgamation and Size on Political Efficacy: Evidence from Difference-in-difference with Exact Matching

	IPE-Complex	IPE-Niemi
Model 1: Absolute population difference	−0.18	−0.10
- Exact matching on public sector employment and size	[0.08]**	[0.14]
Model 2: Relative population difference	−0.12	−0.13
- Exact matching on public sector employment and size	[0.06]*	[0.08]*
Model 3: Absolute population difference	−0.24	−0.24
- Exact matching on public sector employment and education	[0.06]***	[0.13]*
Model 4: Relative population difference	−0.17	−0.22
- Exact matching on public sector employment and education	[0.04]***	[0.09]**
Model 5: Absolute population difference	−0.28	−0.34
- Exact matching on public sector employment	[0.06]***	[0.13]***
Model 6: Relative population difference	−0.18	−0.29
- Exact matching on public sector employment	[0.06]**	[0.08]***

Notes: Standard errors are corrected for clustering at the municipal level, and bootstrapped with 50 replications and resampling at the cluster level. $N = 602$ for models 1 and 2; $N = 1{,}056$ for models 3 and 4; $N = 864$ for models 5 and 6. The entries show beta coefficients and associated standard errors [in brackets] for two dependent variables (in columns) in six models (rows). In models 1 and 2, matching is carried out under the restriction of exact matching on public sector employment combined with size (small vs. large). Model 1 is estimated with the absolute change in population; model 2 with the relative change. In models 3 and 4, matching is carried out under the restriction of exact matching on public sector employment combined with education group (no college vs. some college or more). Model 3 is estimated with the absolute change in population; model 4 with the relative change. In models 5 and 6, matching is carried out under the restriction of exact matching on public sector employment and caliper = 0.025. Model 5 is estimated with the absolute change in population; model 6 with the relative change. Each model shows the results of a difference-in-difference with matching specification. All models are based on exact matching on public sector employment combined with two-dimensional Euclidian metric matching with caliper = 0.05, except where noted. IPE, internal political efficacy.
*Significant at 10%; ** significant at 5%; *** significant at 1%.

Substantive Significance of Results

In addition to statistical significance, the results from Tables 4 and 5 based on matching also have substantive significance, particularly for citizens experiencing large changes: A citizen experiencing a change of 72,500 in population size, which is the maximum observed in the sample, will, based on the estimated model, report a level of IPE on a 2-to-10 scale, which is lower by $7.25 \times (0.23) = -1.7$. A similar calculation based on relative population increase yields −1.5. The standard deviation of IPE-Complex is 2.0 in the matched sample. We put the estimates into perspective by comparing them with the empirical distribution of the feelings of efficacy within our sample, the effects of other variables, and the trend in levels of political efficacy over time.

Examining the distribution of IPE-Complex in the matched sample shows that a decrease of 2 in the index corresponds to a drop from the 75th percentile to the median or from the median to the 20th percentile in the matched sample, meaning that an individual experiencing the maximal observed change will find him- or herself passing on average approximately one fourth of the population on the way to the new level. This is a substantive effect.

To give an alternative interpretation of the quantitative importance of the effects brought about from changes in jurisdiction size, we also compare the estimated effect to other important explanatory variables, focusing on gender, education, and public employment because they are among the most important individual-level predictors of efficacy (see, e.g., Hayes and Bean 1993). The unconditional difference between women and men in the matched sample equals −0.4, the effect of holding a Master's degree relative to having completed lower secondary school is 1.0, and the effect of public employment is only 0.1. Hence, the maximal effect of size is much larger than the observed gender difference and larger than the difference between low and high education.

Finally, we compare the effect of size with the long-term trend in one of the IPE indicators. The question Complex has appeared in the ANES since 1952; in the ANES cumulative data file, answers to complex are coded as agree, disagree, and, from 1988, neither agree nor disagree (ANES 2005, 422). To facilitate a comparison with this time series, we recoded our survey question to indicate the share of respondents who disagree with the statement (i.e., do not find politics complex). We see 50.8% of respondents in the matched sample disagreeing with the statement that local politics is too complex. Carrying out the analysis with this (now binary) dependent variable, we observe an estimated coefficient for the average treatment effect, comparable to the binary treatment effect discussed previously, based on a linear probability model, equal to −0.24,

TABLE 6. Effect of Municipal Amalgamation and Size on Political Efficacy: Evidence from Postreform Retrospective Evaluations

	IPE-Complex Retrospective	IPE-Niemi Retrospective
Model 1: Binary treatment		
Amalgamated municipality	−0.59	−0.96
	[0.12]***	[0.26]***
Model 2: Categorical treatment		
Small-Small	−0.82	−1.57
	[0.17]***	[0.34]***
Small-Large	−0.87	−1.77
	[0.20]***	[0.45]***
Large-Small	−0.64	−0.73
	[0.11]***	[0.24]***
Large-Large	−0.24	−0.22
	[0.13]*	[0.21]
Model 3: Absolute population difference	−0.12	−0.24
	[0.02]***	[0.04]***
Model 4: Relative population difference	−0.07	−0.15
	[0.02]***	[0.04]***
Model 5: Binary treatment with propensity score matching		
Amalgamated municipality	−0.65	−1.11
	[0.14]***	[0.29]***
N	1,035	1,035

Notes: Standard errors are corrected for clustering at the municipal level. The entries show beta coefficients and associated standard errors [in brackets] for two dependent variables (in columns) in five models (rows). Model 1 compares amalgamated municipalities (=1) to continuing, unaffected municipalities (=0). Model 2 distinguishes between amalgamated municipalities by their pre- and postreform size and compares them to continuing, unaffected municipalities (cf. Table 1). Model 3 shows the effect of the absolute population difference (in ten thousands) between a postreform municipality and the prereform municipality. Model 4 shows the relative population difference, calculated as the absolute population difference divided by prereform population. Model 5 compares amalgamated municipalities (=1) to continuing, unaffected municipalities (=0), based on standard propensity score matching. Models 1 to 4 each show the results of one regression analysis and include the same set of control variables (not shown in the table). Full results including controls are available on the authors' Web site. IPE, internal political efficacy.
*Significant at 10%; ** significant at 5%; *** significant at 1%.

which is much larger than the span—over time—of responses to the ANES: Here, the percentage of respondents disagreeing with Complex has been between 0.74 (in 1972) and 0.58 (in 2000), equal to a span of 0.16.

Retrospective Evaluations

Table 6 shows the results from the retrospective questions, for both IPE-Complex and IPE-Niemi. Again, we show results of a number of estimated models in the form of either linear regression with a full set of controls (models 1–4) or standard propensity score kernel matching (model 5).

For all estimation methods and samples, we consistently find a statistically significant decline in IPE for individuals in reform municipalities. However, the effect is not homogenous across treatment categories: The impact of reform is larger for individuals residing in small prereform municipalities and, generally, small and less significant for inhabitants of the largest reform jurisdictions, when these jurisdictions constitute a large part of the newly formed municipality. For both measures, F tests reject the null hypothesis of equal parameter estimates for the four categories. These results are confirmed by the continuous treatment variables: The effect of amalgamation is larger, the larger the change in population size resulting from the reform; thus, individuals experiencing larger (both absolute and relative) changes in the size of their jurisdiction reported greater losses of IPE.

The retrospective evaluations also allow us to assess whether other factors associated with the amalgamation process influence individual evaluations of the change in political efficacy. If, for example, amalgamation changed local public finances dramatically, then individual evaluations of the change in political efficacy could be affected by (dis)satisfaction related to this fact. We construct two variables to account for economic and political change: Economic change is measured by the difference in municipal tax base per capita as a consequence of the reform. Political

change is measured by an indicator variable capturing whether a voter reporting to vote for the mayor's party prereform experienced a loss of political control for that party postreform. These variables do not affect the relationship between population size increases and political efficacy.

CONCLUSION

Internal political efficacy is influenced by many factors. In this article, we focus on one of the most fundamental factors that has been argued to influence the functioning of government and governing in democracies throughout the history of political thought: population size of the polity or political jurisdiction.

Our main conclusion is that jurisdiction size has a causal, detrimental effect on citizens' sense of IPE. Our estimates of the effects of jurisdiction size on IPE are based on survey data collected before and after a recent comprehensive administrative reform of Danish local government. The identification strategy is based on two key features of the reform. First, the reform was largely exogenous from the viewpoint of individual citizens and municipalities, which enables us to address problems of sorting and endogenous political change not considered in the existing literature. Second, it affected only some jurisdictions, allowing us to control, using a DiD approach, for other concurrent changes. Although DiD methods are routinely employed in political science, little attention has been given to their structural assumptions. To address potential threats to our identifying assumptions arising from a lack of common trend in IPE among amalgamated and continuing municipalities, we combine the DiD analysis with exact and propensity score matching. The estimated effects, based on the matching analysis, are substantial: For respondents experiencing large changes, the detrimental effect of the increase in size is approximately equivalent to moving a quartile down the empirical distribution of IPE. These results are robust across different estimators, samples, and measurements of the dependent variable.

We find the effects to be concentrated among dimensions of IPE related to respondents' evaluation of their understanding of local politics, whereas effects are less pronounced in dimensions tapping respondents' evaluation of their organizational or managerial capabilities; this makes sense, we believe, given that municipal responsibilities are the same regardless of size and were largely unaffected by the reform. At the same time, the short time frame since the implementation of the reform allows us to focus on the pure effects of size differences, keeping constant other potential sources of change in the possibilities for political participation in a broad sense, including the number and location of polling places, local adjustments to changes in decision-making procedures, and realignment of the local political environment following reform.

Our approach can have implications for research on the architecture of government. First, the results contribute to the debate about the optimal level of decentralization (Treisman 2007) by substantiating with empirical evidence the argument that having smaller jurisdictions increases IPE. Our analysis could be extended by considering a richer selection of outcomes, by looking at potential heterogeneity in treatment beyond that studied here, and by adding to the analysis, in the future, another round of data collection to study the long-term effects of amalgamation. In particular, a longer time frame would allow the collection of measures of participation.

Second, we believe that our approach, explicitly addressing issues of sorting and endogeneity identified in the literature on urban economic geography (Alesina, Baqir, and Hoxby 2004; Tiebout 1956), can be applied to other parts of the (de)centralization tradeoffs, specifically other dimensions of local democracy, service delivery, and economies of scale, providing a micro-level foundation for decisions on decentralization and municipal architecture.

Finally, and perhaps most important, the results have implications for jurisdictional reform. Large entities are desirable for several reasons. Large units are necessary to reap economies of scale and to influence a wider range of potentially crucial political decisions. However, larger size has real costs for a central indicator of the quality of the democratic order. According to Dahl's (1989) ideals, citizens ought to have adequate and equal access to influencing political decisions. Citizens in large jurisdictions are less likely than citizens in small jurisdictions to believe that this is possible. Although other determinants of feelings of IPE, including education levels and income structures, are difficult to change as such, the institutional framework is amenable to change and does, as demonstrated here, actually matter in a causal sense. When the size of municipalities increases, internal political efficacy drops.

APPENDIX

TABLE A.1. Municipality Characteristics

Stratum	Municipality Name	Size		Population Growth		Full Sample			Matched Sample		
		2006	2007	Growth (%)	(Abs)	Pre	Post	Total	Pre	Post	Total
NM	København	501,158	503,699	0.5	2,541	33	26	59	0	0	0
NM	Dragør[a]	13,154	13,184	0.2	30	32	36	68	97	72	169
NM	Gladsaxe	61,735	61,945	0.3	210	30	39	69	34	31	65
NM	Høje Taastrup	46,257	46,683	0.9	426	28	31	59	49	49	98
NM	Vallensbæk[a]	12,230	12,145	−0.7	−85	34	40	74	61	76	137
NM	Helsingør	61,340	61,012	−0.5	−328	30	22	52	20	13	33
NM	Ringsted	31,094	31,468	1.2	374	29	32	61	62	67	129
NM	Læsø	2,091	2,058	−1.6	−33	30	30	60	62	76	138
SS	Ramsø[a]	9,412	81,017	760.8	71,605	32	32	64	19	14	33
SS	Holmegaard[a]	7,643	80,133	948.4	72,490	30	30	60	13	8	21
SS	Glamsbjerg	5,924	41,816	605.9	35,892	23	29	52	12	13	25
SS	Augustenborg	6,525	76,825	1,077.4	70,300	30	38	68	19	15	34
SS	Gram	4,867	56,275	1,056.3	51,408	27	34	61	14	21	35
SS	Lundtoft	6,150	60,044	876.3	53,894	34	34	68	18	23	41
SL	Ledøje-Smørum	10,797	40,057	271	29,260	31	48	79	19	29	48
SL	Jægerspris	9,520	43,910	361.2	34,390	31	24	55	20	8	28
SL	Stenlille	5,634	28,956	414	23,322	23	33	56	14	11	25
SL	Trundholm	11,311	32,980	191.6	21,669	24	21	45	11	9	20
SL	Ejby	10,192	36,771	260.8	26,579	22	37	59	11	23	34
SL	Hadsten	11,969	45,037	276.3	33,068	34	30	64	16	11	27
LS	Frederikssund	19,144	43,910	129.4	24,766	33	34	67	24	20	44
LS	Korsør[a]	20,873	76,949	268.7	56,076	29	28	57	17	10	27
LS	Hedensted	17,190	44,892	161.2	27,702	26	44	70	19	28	47
LS	Grenaa	18,673	38,333	105.3	19,660	33	32	65	19	14	33
LS	Viborg	44,505	91,405	105.4	46,900	28	43	71	15	18	33
LS	Hjørring	35,296	67,118	90.2	31,822	32	30	62	24	18	42
LL	Hillerød	38,102	46,354	21.7	8,252	32	31	63	17	9	26
LL	Middelfart[a]	20,599	36,771	78.5	16,172	37	36	73	23	27	50
LL	Randers	62,524	92,984	48.7	30,460	32	27	59	12	13	25
LL	Silkeborg	55,906	86,540	54.8	30,634	33	39	72	15	18	33
LL	Skive	27,972	48,344	72.8	20,372	33	36	69	11	21	32
LL	Aalborg	163,952	194,149	18.4	30,197	28	39	67	0	0	0
Total						**963**	**1,065**	**2,028**	**767**	**765**	**1,532**

Notes: Municipality name refers to name before reform. LL, large before reform, large share after; LS, large before reform, small share after; NM, not merged; SL, small before reform, large share after; SS, small before reform, small share after.
[a] Refers to municipal referendum on amalgamation process, as described in text.

TABLE A.2. Regression-based DiD Analysis of Effect of Municipality Size on Political Efficacy

	Model 1	Model 2	Model 3	Model 4
Postreform ($P = 1$)	−0.09 [0.18]	−0.10 [0.18]	−0.14 [0.14]	−0.27 [0.11]***
Treated municipalities ($M = 1$)	0.06 [0.15]			
Treated Ms, postreform (P x $M = 1$)	−0.36 [0.20]*			
Treated municipalities ($M =$ SS)		0.05 [0.17]		
Treated municipalities ($M =$ SL)		−0.03 [0.18]		
Treated municipalities ($M =$ LS)		0.14 [0.18]		
Treated municipalities ($M =$ LL)		0.07 [0.17]		
Treated Ms, postreform (P x SS $= 1$)		−0.43 [0.23]*		
Treated Ms, postreform (P x SL $= 1$)		−0.23 [0.27]		
Treated Ms, postreform (P x LS $= 1$)		−0.44 [0.23]*		
Treated Ms, postreform (P x LL $= 1$)		−0.35 [0.25]		
Population change from reform (in 10,000s)			0.00 [0.02]	0.00 [0.01]
Population change x $P = 1$			−0.08 [0.04]**	−0.04 [0.02]*
Vocational training	−0.12 [0.07]	−0.12 [0.07]	−0.11 [0.07]	−.12 [0.07]
College degree	0.23 [0.09]**	0.23 [0.09]**	0.22 [0.09]**	0.22 [0.09]**
Master's degree or above	0.66 [0.13]***	0.66 [0.13]***	0.66 [0.14]***	0.66 [0.14]***
Income, 2nd quintile	0.24 [0.13]*	0.24 [0.13]*	0.23 [0.13]*	0.23 [0.13]*
Income, 3rd quintile	0.24 [0.14]*	0.23 [0.14]*	0.23 [0.14]	0.23 [0.14]
Income, 4th quintile	1.01 [0.18]***	1.01 [0.18]***	0.99 [0.17]***	0.99 [0.17]***
Income, 5th quintile	1.14 [0.17]***	1.14 [0.17]***	1.13 [0.17]***	1.14 [0.17]***
Publicly employed	0.29 [0.11]**	0.29 [0.11]**	0.29 [0.11]***	0.29 [0.10]***
Female	−0.32 [0.08]***	−0.32 [0.08]***	−0.33 [0.08]***	−0.33 [0.08]**
Age	0.04 [0.02]**	0.04 [0.02]**	0.04 [0.02]**	0.04 [0.02]**
Age^2	−0.02 [0.01]	−0.02 [0.01]	−0.02 [0.01]	−0.02 [0.01]
Constant	5.68 [0.43]***	5.68 [0.44]***	5.76 [0.40]***	5.76 [0.40]***
Observations	2,028	2,028	2,028	2,028
R^2	0.10	0.10	0.10	0.10

Notes: Standard errors are corrected for clustering at the municipal level. The table shows the full output of models 1 to 4 in Table 3 with IPE-Complex as the dependent variable. DiD, difference-in-difference.
*Significant at 10%; **significant at 5%; ***significant at 1%.

REFERENCES

Abadie, Alberto, and Guido W. Imbens. 2008. "On the Failure of the Bootstrap for Matching Estimators." *Econometrica* 76 (6): 1537–57.

Alesina, Alberto, Reza Baqir, and Caroline Hoxby. 2004. "Political Jurisdictions in Heterogeneous Communities." *Journal of Political Economy* 112 (2): 348–96.

Alesina, Alberto, and Enrico Spolaore. 1997. "On the Number and Size of Nations." *Quarterly Journal of Economics* 112 (4): 1027–56.

Almond, Gabriel A., and Sidney Verba. 1963. *The Civic Culture*. Princeton, NJ: Princeton University Press.

American Association for Public Opinion Research (AAPOR). 2008. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. 5th ed. Lenexa, KS: AAPOR.

American National Election Studies (ANES). 1994. *Continuity Guide to the American National Election Studies*. Ann Arbor, MI: Center for Political Studies.

American National Election Studies (ANES). 2005. ANES Time Series Cumulative Data File. Codebook variable documentation file. www.electionstudies.org/studypages/cdf/anes_cdf_var.txt (accessed May 5, 2010).

Andersen, Jørgen G. 2000. "Magt og afmagt: Nyt perspektiv på 'political efficacy.'" In *Hvad folket magter: Demokrati, magt og afmagt* [The Power of the People: Democracy, Power, and Powerlessness], eds. Jørgen Goul Andersen, Lars Torpe, and Johannes Andersen. Copenhagen: Jurist- og Økonomforbundets Forlag, 123–52.

Angrist, Joshua D., and Alan B. Krueger. 2001. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives* 15 (4): 69–85.

Aristotle. 1948. *Politics*. Trans. Ernest Barker. Oxford: Oxford University Press.

Banzhaf, H. Spencer, and Randall P. Walsh. 2008. "Do People Vote with Their Feet? An Empirical Test of Tiebout's Mechanism." *American Economic Review* 98 (3): 843–63.

Barabas, Jason, and Jennifer Jerit. 2010. "Are Survey Experiments Externally Valid?" *American Political Science Review* 104 (2): 226–42.

Barnard, John, Constantine Frangakis, Jennifer Hill, and Donald B. Rubin. 2002. "School Choice in NYC: A Bayesian Analysis of an Imperfect Randomized Experiment." In *Case Studies in Bayesian Statistics, Volume V*, Lecture Notes in Statistics Series, Vol. 162, eds. Constantine Gatsonis, Robert E. Kass, Bradley Carlin, Alicia Carriquiry, Andrew Gelman, Isabella Verdinelli, and Mike West. New York: Springer, 3–97.

Benton, J. Edwin, and Darwin Gamble. 1983. "City/County Consolidation and Economies of Scale: Evidence from a Time-series Analysis in Jacksonville, Florida." *Social Science Quarterly* 65 (1): 190–98.

Berg, Rikke, and Ulrik Kjær. 2005. *Den danske borgmester* [The Danish Mayor]. Odense: University Press of Southern Denmark.

Besley, Timothy, and Anne Case. 2000. "Unnatural Experiments? Estimating the Incidence of Endogenous Policies." *Economic Journal* 110 (467): 672–94.

Bhatti, Yosef, and Kasper M. Hansen. 2011. "Who 'Marries' Whom? The Influence of Societal Connectedness, Economic and Political Homogeneity, and Population Size on Jurisdictional Consolidations." *European Journal of Political Research* 50 (2): 212–38.

Blom-Hansen, Jens, Jørgen Elklit, and Søren Serritzlew. 2006. "Den store kommunalreform og dens konsekvenser." In *Kommunalreformens konsekvenser* [The Effects of the Structural Reform],

eds. Jens Blom-Hansen, Jørgen Elklit, and Søren Serritzlew. Århus, Denmark: Academica, 11–36.
Blundell, Richard, Monica Costa Dias, Costas Meghir, and John Van Reenen. 2004. "Evaluating the Employment Impact of a Mandatory Job Search Program." *Journal of the European Economic Association* 2 (4): 569–606.
Blundell, Richard, and Thomas MaCurdy. 1999. "Labor Supply: A Review of Alternative Approaches." In *Handbook of Labor Economics, Vol. 3A*, eds. Orley C. Ashenfelter and David Card. Amsterdam: Elsevier, 1559–695.
Campbell, Angus, Gerald Gurin, and Warren E. Miller. 1954. *The Voter Decides*. Evanston, IL: Row, Peterson.
Campbell, Donald T. 1969. "Reforms as Experiments." *American Psychologist* 24 (25): 409–29.
Christiansen, Peter Munk, and Michael B. Klitgaard. 2008. *Den utænkelige reform* [The Unthinkable Reform]. Odense: University Press of Southern Denmark.
Clarke, Harold D., and Alan C. Acock. 1989. "National Elections and Political Attitudes: The Case of Political Efficacy." *British Journal of Political Science* 19 (4): 551–62.
Cook, Thomas D., William R. Shadish, and Vivian C. Wong. 2008. "Three Conditions under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-study Comparisons." *Journal of Policy Analysis and Management* 27 (4): 724–50.
Dahl, Robert A. 1967. "The City in the Future of Democracy." *American Political Science Review* 61 (4): 953–70.
Dahl, Robert A. 1989. *Democracy and Its Critics*. New Haven, CT: Yale University Press.
Dahl, Robert A., and Edward R. Tufte. 1973. *Size and Democracy*. Palo Alto, CA: Stanford University Press.
Dehejia, Rajeev H., and Sadek Wahba. 2002. "Propensity Score Matching Methods for Nonexperimental Causal Studies." *Review of Economics and Statistics* 84 (1): 151–61.
DeHoog, Ruth H., David Lowery, and William E. Lyons. 1990. "Citizen Satisfaction with Local Governance: A Test of Individual, Jurisdictional, and City-specific Explanations." *Journal of Politics* 52 (3): 807–37.
Finifter, Ada W. 1970. "Dimensions of Political Alienation." *American Political Science Review* 64 (2): 389–410.
Finifter, Ada W., and Paul R. Abramson. 1975. "City Size and Feelings of Political Competence." *Public Opinion Quarterly* 39 (2): 189–98.
Hayes, Bernadette C., and Clive S. Bean. 1993. "Political Efficacy—A Comparative Study of the United States, West Germany, Great Britain and Australia." *European Journal of Political Research* 23 (3): 261–80.
Heckman, James J., Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66 (5): 1017–98.
Heckman, James J., Hidehiko Ichimura, and Petra Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *Review of Economic Studies* 64 (4): 605–54.
Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15 (3): 199–236.
Houlberg, Kurt, and Carsten Strømnæk Pedersen. 2003. "Appendiks 1: Dataindsamling." In *Kommunestørrelse og lokalt demokrati* [Municipal Size and Local Democracy], eds. Ulrik Kjær and Poul Erik Mouritzen. Odense: University Press of Southern Denmark, 207–10.
Iacus, Stefano M., Gary King, and Giuseppe Porro. 2009. "Causal Inference without Balance Checking: Coarsened Exact Matching." Harvard University. Unpublished manuscript.
Imbens, Guido W., and Jeffrey M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 47 (1): 5–86.
Jørgensen, Lene P. 2006. "Hvor skal vi hen? Folkeafstemninger: Eliternes instrument eller folkets stemme?" In *Stort er godt* [Large Is Good], ed. Poul Erik Mouritzen. Odense: University Press of Southern Denmark, 167–87.
Kjær, Ulrik, and Poul Erik Mouritzen, eds. 2003. *Kommunestørrelse og lokalt demokrati* [Municipal Size and Local Democracy]. Odense: University Press of Southern Denmark.
Lane, Robert E. 1961. *Political Life: Why People Get Involved in Politics*. Glencoe, IL: Free Press.
Larsen, Christian Albrekt. 2002. "Municipal Size and Democracy: A Critical Analysis of the Argument of Proximity Based on the Case of Denmark." *Scandinavian Political Studies* 25 (4): 317–32.
Lassen, David Dreyer. 2005. "The Effect of Information on Voter Turnout: Evidence from a Natural Experiment." *American Journal of Political Science* 49 (1): 103–18.
Lolle, Henrik. 2003. "Kommunalpolitisk selvtillid." In *Kommunestørrelse og lokalt demokrati* [Municipal Size and Local Democracy], eds. Ulrik Kjær and Poul Erik Mouritzen. Odense: University Press of Southern Denmark, 162–76.
Lyons, W. E., and David Lowery. 1989. "Governmental Fragmentation versus Consolidation: Five Public-choice Myths about How to Create Informed, Involved, and Happy Citizens." *Public Administration Review* 49 (6): 533–43.
Madsen, Douglas. 1987. "Political Self-efficacy Tested." *American Political Science Review* 81 (2): 571–81.
Meyer, Bruce D. 1995. "Natural and Quasi-experiments in Economics." *Journal of Business and Economic Statistics* 13 (2): 151–61.
Morlan, Robert L. 1984. "Municipal vs. National Election Voter Turnout: Europe and the United States." *Political Science Quarterly* 99 (3): 457–70.
Morrell, Michael E. 2003. "Survey and Experimental Evidence for a Reliable and Valid Measure of Internal Political Efficacy." *Public Opinion Quarterly* 67 (4): 589–602.
Muller, Edward N. 1970. "Cross-national Dimensions of Political Competence." *American Political Science Review* 64 (3): 792–809.
Niemi, Richard G., Stephen C. Craig, and Franco Mattei. 1991. "Measuring Internal Efficacy in the 1988 National Election Study." *American Political Science Review* 85 (4): 1407–13.
Oliver, J. Eric. 2000. "City Size and Civic Involvement in Metropolitan America." *American Political Science Review* 94 (2): 361–73.
Pollock, Philip H. 1983. "The Participatory Consequences of Internal and External Political Efficacy: A Research Note." *Western Political Quarterly* 36 (3): 400–9.
Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (1): 41–55.
Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688–701.
Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64 (5): 416–24.
Titus, A. Constandina. 1981. "Shaping Attitudes towards Local Government: Factors That Influence Political Trust and Efficacy." *Urban Interest* 3 (2): 37–45.
Treisman, Daniel. 2007. *The Architecture of Government*. Cambridge: Cambridge University Press.
Verba, Sidney, and Norman H. Nie. 1972. *Participation in America*. New York: Harper and Row.
Vetter, Angelika. 2002. "Local Political Competence in Europe: A Resource of Legitimacy for Higher Levels of Government?" *International Journal of Public Opinion Research* 14 (1): 3–18.