Beyond Federalism: Estimating and Explaining the Territorial Structure of Government

Liesbet Hooghe* and Gary Marks†

*University of North Carolina—Chapel Hill and VU University Amsterdam; hooghe@unc.edu
†University of North Carolina—Chapel Hill and VU University Amsterdam; marks@unc.edu

This article suggests that the basic distinction between federal and unitary government has limited as well as served our understanding of government. The notion that variation in the structure of government is a difference of kind rather than degree has straight-jacketed attempts to estimate the authority of intermediate government. One result has been the claim that a country’s footprint, not its population, is decisive for government. Analyzing data for thirty-nine countries since 1950, and comparing our own findings with those of alternative measurements, we find evidence for the causal effect of population. This can be theorized in terms of a trade-off between responsiveness to soft information and per-capita economies in public good provision.

The structure of government—the allocation of authority across general purpose jurisdictions—is a deep and puzzling phenomenon. Philosophers from Hobbes to Madison to Ostrom, and political scientists from Riker and Elazar to Lijphart have sought to explain why some countries are, or should be, decentralized while others are centralized.

Since the creation of the United States, government structure has been conceived as a basic choice between a unitary and a federal system. A unitary system is one in which decision making may be deconcentrated or even decentralized, but final authority rests with the center. A federal system, by contrast, disperses authority between “regional governments and a central government in such a way that each kind of government has some activities on which it makes final decisions” (Riker 1987, 101). Most importantly, regions or their representatives can veto constitutional reform.

Each conception is rooted in a coherent, but opposing, philosophy: “The contrast between Hobbes’ formulation of the institutions of government and that formulated in the American experiments in constitutional choice suggests that fundamentally different approaches exist for the organization of governance in human societies. In the one, a single center of authority, the sovereign, is designed...
to dominate the whole. Each unit, as a sovereign state, is fully independent of other units as sovereign states. In the other, no single center of authority dominates the rest: all assignments of authority are subject to limits” (Ostrom 1987, 22).

The unitary/federal distinction lies at the heart of our understanding of constitutional choice, and has produced a rich literature on political institutions, but it has limited as well as structured our understanding of government. It conceives variation among governments as difference of kind, rather than degree (Wibbels 2005, 67). It has directed attention to variation among federal regimes, but has much less to say about variation among unitary regimes. And while it is highly attuned to constitutional choice, it is less informative about incremental reform.

One result is a disconnect between sophisticated case studies of federal regimes and relatively crude attempts to estimate variation in government structure. Measures of government structure have struggled to impose continuous variation on a categorical distinction. One response has been to use fiscal data. However, as we explain below, the money a government raises or spends is a poor indicator of its authority.

In this article we compare measures of regional authority and apply them to intermediate government in thirty-nine democracies. We revisit the structural determinants of government structure, paying special attention to the relative influence of the size and population of a country. The notion that countries with larger footprints tend to be more decentralized is grounded in the theory of spatial externalities which has provided a foundation for fiscal federalism. An alternative line of thinking is that more populous countries tend to be more decentralized because the provision of public goods depends on soft information.

In the next section we set out some basic expectations. We then estimate government structure in order to discriminate between them, paying detailed attention to alternative measurement instruments. We conclude that population is more potent than area in accounting for regional decentralization and that more refined measurement can be decisive in assessing competing claims about the structure of government.

Theorizing Government Structure

Spatial Theory

Three lines of argument have been used to connect the size of a country with its level of decentralization. The first goes under the heading “spatial decay” which describes the increasing costs of communication imposed by distance. These costs could include “inefficiency in the provision and delivery of local public goods as well as transportation costs” (Arzaghi and Henderson 2005, 1165). If policy provision is subject to spatial decay, centralized provision of public goods becomes
more wasteful the more distant the region, and correspondingly, decentralized provision becomes relatively more efficient.

The second argument applies the same spatial logic to policy externalities. The larger the footprint of a jurisdiction, the less its policy making will affect neighboring jurisdictions, and the smaller the efficiency loss arising from policy spillover. The argument informs public goods theory that sets up a trade-off between the benefit of adapting policy to particular regions and the loss arising from failure to internalize the effects of local decisions for neighboring regions (Alesina and Spolaore 2003; Oates 1972). “Perfect mapping” of government structure requires that different government tiers provide efficient levels of outputs of public goods whose benefits are “encompassed by the geographical scope of their jurisdictions” (Oates 2005: 351; Olson 1969). The implication is that the larger a country, the smaller the costs arising from policy spillover among regional governments (Oates 2005, 357).

A third line of argument assumes that larger countries are more heterogeneous, and that as a consequence, larger countries decentralize in order to fit policy to the preferences of those living in particular regions (Alesina et al. 1995, 754; Breuss 2004, 40; Färber 2001, 112).

Empirical research has confirmed the idea that larger countries are more decentralized. Arzaghi and Henderson (2005, 1179) conclude that “[L]and effects are enormous and significant, where greater land area and hence spatial dispersion increase the likelihood of being federal. From the base probability of 0.18…a one-standard deviation in land raises the probability of being federal to 0.60.” Panizza (1999, 113) finds that “When more than one measure of size is included in the regression, only Area shows a robust correlation with fiscal centralization.” Garrett and Rodden (2003, 97) find that “As expected, countries with larger area are significantly more decentralized…. Population and urbanization have no effect on decentralization in any of the estimations, so we drop them from subsequent analysis.”

Soft Information Versus Scale Economies

An alternative to spatial theory conceives jurisdictional design as a trade-off between responding appropriately to soft information, which favors decentralization, and exploiting scale economies in providing the public good, which favors centralization. Both sides of the trade-off are influenced by the number of people in the jurisdiction.

The argument that soft information requires decentralization was put on the table by organizational economists and political scientists who conceive “decision-making for an organization as a process of repeated messages or dialogue” (Arrow 1991, 5; Kochen and Deutsch 1969, 735). “We may regard it as
close to an impossibility for individuals in close contact with the productive processes to transmit their information in all its details to another office. This proposition, long recognized in practice, is the basis of the management literature on the questions of centralization and decentralization” (Arrow 1961, 11).

Arrow is taking issue with the Fordist notion that the job of front-line personnel is to convey information to their superiors who then make investment decisions. Decision making in firms engages persons as well as things, and such information may be soft—difficult to standardize, resistant to batching, and correspondingly expensive to pass up an organizational hierarchy. A local bank manager, for example, is better positioned to decide on a small-business loan by talking directly with the borrowers to assess their honesty and acumen than is a central banker who has access to a report written by the bank manager (Stein 2002, 1892–93).

Government itself can be conceived as a process of repeated dialogue with citizens, and the information that is required for the provision of public goods is no less soft than that for small business loans. This motivates Elinor Ostrom’s (2010, 8) summary of her decades-long contribution to the study of metropolitan government: “Advocates of the metropolitan reform approach assumed that size of governmental units would always be positive for all types of goods and services. Scholars using a political economy approach [by contrast] assumed that size of governmental units would be positive or negative depending on the type of public good or service. Those involving face-to-face delivery, such as education, policing, and social welfare, would show a negative effect of governmental unit size; those involving economies of scale, such as highways and utility systems, would show a positive effect.”

The argument applies both to the input and the output side of government. A decentralized government is better placed to respond to soft information summarizing the preferences of those who live in a region and to implement appropriate policy. This is the case even if there is no heterogeneity of preferences across localities. The contexts of human interaction may vary even if preferences do not (Jeffery 2012). “Street-level” case studies of policy making, beginning with the classic example of the Tennessee Valley Authority, generated a vocabulary to describe this—“task environment,” “local interaction,” “local stimuli” (Scholz, Twolmby, and Headrick 1991; Hodge 1938; Pritchett 1943).

The notion that policy making under soft information demands dialogue between local and central decision makers underpins social, environmental, research and educational policy making in the European Union. Zeitlin (2011, 2) theorizes this as “experimentalist” governance: “At the core of these new forms is a recursive process of provisional goal-setting and revision through feedback from experience of pursuing them in different contexts. Sub-units within and beyond the organization are given substantial responsibility for defining the best ways to achieve these goals, separately and in conjunction with one another. They are also
responsible for monitoring their own operations to find and fix gaps in the ensuing plans as they occur. Their results are then compared against one another, to identify and diffuse opportunities for performance improvement. Finally the goals themselves are periodically revised in response to the problems and possibilities revealed by such reviews” (see also Sabel and Zeitlin 2010).

To the extent that information is soft, informational costs can be expected to increase with population. The larger a group, the more difficult it is to sustain meaningful dialogue: “[T]he need for a minimum message length to achieve any use leads to the view that the number of participants in the decision making dialogue must be strongly limited.... The argument becomes even clearer if we take the cost of communication to be basically a delay in making the decision.” (Arrow 1991, 6).

One scenario is where there are no economies of scale at all—that is, a world of soft information where each message has to be individually handled (Treisman 2007, 63–69). Minimizing the time delay in communication between a government and its population requires intermediaries who send and return messages to other agents and so on down to each person in the jurisdiction. The delay-minimizing setup is a hierarchical network—a system of multilevel governance—where each agent communicates with the same number of agents. Larger populations require more agents at more intermediate levels.

The number of intermediate government levels under perfectly hard information is, by contrast, zero. If there are infinite economies of scale in sending, receiving, and processing messages, then there is no need for intervening agents. The ruler communicates directly with the entire population, sending the same message to each person, and processing all messages received in a single batch. Such standardization is more akin to dictatorship than democracy. However, if preferences can be summarized along a single-peaked dimension, the ruler can be a computer algorithm producing pareto optimal policy. Under perfectly hard information, government structure is impervious to population size.

The extent to which soft information requires authoritative decentralization, and not merely deconcentration, is debated (Hooghe and Marks 2009; Oates 2006; Treisman 2007). In principle, the central state could adjust policy to local contexts without empowering subnational actors. There are several possibilities. A central legislature could be composed of locally elected representatives who make local policy; the central government could appoint or control local agents; or the central government might use local agents to collect and report the relevant information (Besley and Coate 1997; Lockwood 2002; Treisman 2007). This view rests on the assumption that information collection does not require discretion, an assumption we can evaluate with the data we have generated.

The benefits of decentralization exist in tension with its costs. These arise because the population of a decentralized jurisdiction may be too small to reap
economies of scale. Decentralization becomes inefficient when the per-capita cost of
centralized provision of the public good is lower than the per-capita cost of
the good provided by each subnational jurisdiction. The fixed per-capita cost of
a public good is inversely related to the population of the jurisdiction that
provides it.

Both the benefits and costs of decentralization can be expected to vary across a
government’s policy portfolio. The information necessary for efficient provision of
defense or utilities, for example, is chiefly hard, whereas that for job retraining,
kindergartens, or home care, involves soft information and on-the-job appraisal.
Similarly, least-cost output in defense, transportation, utilities, and networks will
usually require large amounts of capital, while that for school systems, hospitals,
refuse disposal, welfare provision, and policing will require smaller amounts of
capital (Ostrom and Ostrom 1971; Ostrom and Parks 1999).

The literatures on the provision of public goods produce plausible, but
contrasting, implications for government structure. Spatial theory implies that
government structure is shaped by the territorial size of a country on the grounds
that larger countries tend to be more heterogeneous, face greater spatial decay in
the delivery of public goods, or suffer less from spillovers among subnational
governments. An alternative, but not mutually exclusive, line of thinking highlights
population. If government structure is determined by a scissors whose blades are
soft information and per-capita cost, one would expect that, irrespective of their
footprint or preference heterogeneity, countries with larger populations will be
more decentralized.

Two inferential challenges lie in wait as we evaluate the validity of these claims.
First, we need to estimate government structure in a reasonably unbiased way.
Second, we must control for contending influences on government structure. We
tackle these in turn.

**Estimating Government Structure**

In order to assess the claims set out above, we use a measurement instrument that
(i) evaluates the scope and depth of subnational authority; (ii) encompasses
multiple subnational levels of government where they exist; (iii) estimates variation
among thirty-nine unitary countries and among federal countries, and (iv) tracks
annual change from 1950 to 2006 (Hooghe, Marks, and Schakel 2010).

The dependent variable is the authority exercised within a country by regional
government. Operationalization of regional authority can draw on well-established
concepts. A region is defined as a general-purpose government at any tier between
the local and national government having an average population of 150,000 or
more. Authority is conceived as legitimate power, that is, power recognized as
binding because it is derived from accepted principles of governance (Dahl 1968).
A regional government may exercise authority in its own jurisdiction or in the country as a whole. This is the distinction between *self-rule* and *shared rule* (Elazar 1987; Amoretti and Bermeo 2004; Keating 1998; Lane and Ersson 1999; Watts 1998, 1999; Ziblatt 2006). This distinction provides a frame for disaggregating regional authority in eight dimensions (listed in Table 1). The Cronbach’s alpha across these dimensions in 2006 is 0.93. This suggests that they can be interpreted as indicators of a single construct. Principal components analysis reveals that the eight dimensions are closely associated with the domains of self-rule and shared rule, but that 68 percent of the variance across the dimensions is shared.

Table 1 contrasts the Regional Authority Index (RAI) with five alternative measurement instruments. Table 2 lists the observations where the RAI diverges more than two standard deviations from these measures.

The most cited measure is Lijphart’s (1999) “Federal/Unitary dimension” with which he tests hypotheses about consensual versus Westminster democracy. The greatest differences with Lijphart are for Italy and France, which shift markedly in the RAI between 1965 and 1995, but little in the Lijphart dataset. Italy created a new regional tier, *regioni*, in 1972 with directly elected councils and competencies in urban planning, health, and education. In France, Napoleonic départements gained authority in 1982 when the powers of centrally appointed prefects were transferred to the presidents of directly elected départemen
tc councils.

There is also divergence at the top end of the scale. All but one of the seven federal regimes identified by Lijphart in 1995 score the maximum, while estimates range between 18 (Austria) and 29.3 (Germany) on the RAI. Austria and Germany are both federal polities, but there are some sharp differences. In contrast to German Länder, Austrian Länder have little authority over national legislation or over the base or rate of regional taxes. Furthermore, decentralization in Germany does not stop at the Länder, but encompasses a second tier of Kreise and, in the larger Länder, an intermediate tier of Regierungsbezirke.

Arzaghi–Henderson estimate considerably more decentralization in Poland for 1990 and 1995 than does the RAI. At issue is the difference between decentralization and deconcentration. Regional governments in Poland were downgraded in 1990 to central outposts and direct elections were reinstated only in 1999 (Schakel 2008, 156; Council of Europe 2000, 47–8). Brancati registers no change in Belgium from 1985 to 2000, whereas the RAI spikes up in 1989 when Belgian regions and communities obtained broader policy competencies, taxation powers, and shared rule.

Panizza (1999) and Stegarescu (2005) use measures of fiscal decentralization based on IMF and OECD statistics. These much-used data estimate subnational receipts as a share of total government receipts. Panizza has estimates for 1975, 1980, and 1985, whereas Stegarescu’s slightly different fiscal measure provides an annual time series from 1965 to 2001. Our scores differ more than two standard
<table>
<thead>
<tr>
<th>Measurement instruments</th>
<th>Description</th>
</tr>
</thead>
</table>
| RAI (Hooghe, Marks and Schakel 2010) | Thirty-nine democracies  
Annual 1950–2006 |
<p>| Regional authority (0–24) per region/regional tier. | |
| Institutional depth (0–3): extent to which a regional government is autonomous rather than deconcentrated. | |
| Policy scope (0–4): range of policies for which a regional government is responsible. | |
| Fiscal autonomy (0–4): extent to which a regional government can independently tax its population. | |
| Representation (0–4): extent to which a region is endowed with an independent legislature and executive. | |
| Law making (0–2): extent to which regional representatives co–determine national legislation. | |
| Executive control (0–2): extent to which a regional government co–determines national policy in intergovernmental meetings. | |
| Fiscal control (0–2): extent to which regional representatives co–determine the distribution of national tax revenues. | |
| Constitutional reform (0–3): extent to which regional representatives co–determine constitutional change. | |</p>
<table>
<thead>
<tr>
<th>Source</th>
<th>Study Description</th>
<th>Federalism (1–5). Classification in five categories:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lijphart (1999)</td>
<td>Thirty-six pre-1996 democracies</td>
<td>• 1 = unitary and centralized;</td>
</tr>
<tr>
<td></td>
<td>Two time points (1965, 1995)</td>
<td>• 2 = unitary and decentralized;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 3 = semi-federal;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 4 = federal and centralized;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• 5 = federal and decentralized</td>
</tr>
<tr>
<td></td>
<td>Criteria: a) do states “have formally federal constitutions” or not; b)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>are federal or nonfederal states centralized or decentralized (i.e. is the</td>
<td></td>
</tr>
<tr>
<td></td>
<td>range of powers assigned to the regional level large or small); c) do some</td>
<td></td>
</tr>
<tr>
<td></td>
<td>states fall into an intermediate category? (Lijphart 1999, 188–91.)</td>
<td></td>
</tr>
<tr>
<td>Arzaghi–Henderson (2005)</td>
<td>Forty-eight countries with population over 10 million</td>
<td>Federalism (0–4). The fiscal, political, and administrative responsibilities of subnational governments averaged across six dimensions:</td>
</tr>
<tr>
<td></td>
<td>Five-year intervals for 1960–1995</td>
<td>• official federal versus unitary government structure (0 or 4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• election of a regional executive (0 or 4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• election of a local executive (0 or 4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• ability of the center to suspend lower levels of government or to override their decisions (0 or 4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• no, limited, full revenue raising authority of lower levels of governments (0, 2, or 4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• revenue sharing (0, 2 or 4)</td>
</tr>
<tr>
<td>Study</td>
<td>Countries/Time Period</td>
<td>Political decentralization (0–5). Vertical division of power among multiple levels of government that have independent decision-making power over at least one issue area:</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Brancati (2006, 2008)</td>
<td>Thirty countries with regional ethnic groupsAnnual data for 1980–2000</td>
<td>• democratically elected regional legislatures (0 or 1) • regional legislatures can raise or levy their own taxes (0 or 1) • regional legislatures have joint or exclusive control over education (0 or 1) • regional legislatures have joint or exclusive control over public order or police (0 or 1) • regions approve whether constitutional amendments become law (0 or 1)</td>
</tr>
<tr>
<td>Panizza (1999)</td>
<td>Seventy-six countriesThree time points (1975, 1980, 1985)</td>
<td>Decentralization ratio (0–100). Calculated as 100 minus fiscal centralization ratio which is central government revenues as percentage of total public sector revenues (IMF data).</td>
</tr>
<tr>
<td>Stegarescu (2005)</td>
<td>Twenty-three OECD democraciesAnnual data for 1965-2001</td>
<td>Fiscal autonomy (0–100). Autonomous sub-central tax revenues as percentage of total government tax revenues, excluding social security revenues and taxes paid to the EU (OECD data).</td>
</tr>
</tbody>
</table>
deviations from those of Panizza for Belgium, and in the case of Stegarescu, also for Germany. In both countries, regional governments exert considerable authority, but this is not indicated by their fiscal take. Belgian regions and communities came to exercise broad-ranging legal and constitutional powers after 1989, but fiscal decentralization was blocked until 1995. German Länder exchanged tax autonomy for shared rule in 1966, and are among the most authoritative subnational governments in any democracy.

Figure 1 maps the Stegarescu estimates against those of the RAI for 1999, the latest year with estimates for the maximum number of countries shared between the indices. The further north a country in Figure 1, the greater is its share of government revenue. Sweden and Japan are placed above the United States; Denmark and Finland above Belgium and Australia; and every one of these countries above Germany. None of these comparisons accord with the literature on subnational government in these countries. Fiscal measures have been used to good effect in studies of fiscal federalism (Boadway and Shah 2009; Rodden 2006), but they appear to be inappropriate as measures of political decentralization (Rodden 2004; Schakel 2008; Sorens 2010).

Table 2 Regressing the RAI on alternative instruments

<table>
<thead>
<tr>
<th>Measure</th>
<th>Residuals larger than 2 St. Dev.</th>
<th>Cases</th>
<th>Disagreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lijphart</td>
<td>–2.41</td>
<td>France 1995</td>
<td>Extent of authority for new and directly elected tier</td>
</tr>
<tr>
<td></td>
<td>–2.57</td>
<td>Italy 1995</td>
<td>Extent of authority for new and directly elected tier</td>
</tr>
<tr>
<td>Arzaghi-Henderson</td>
<td>+2.10</td>
<td>Poland 1990, 1995</td>
<td>Extent of central control over regional tier</td>
</tr>
<tr>
<td>Brancati</td>
<td>–2.37 to –3.08</td>
<td>Belgium 1989–2000</td>
<td>Reform of policy, tax, and constitutional veto powers for regions &amp; communities</td>
</tr>
<tr>
<td>Panizza</td>
<td>–2.40 to –2.48</td>
<td>Belgium 1980, 1985</td>
<td>Fiscal revenues versus political authority</td>
</tr>
<tr>
<td>Stegarescu</td>
<td>–2.36 to –2.44</td>
<td>Belgium 1989–1994</td>
<td>Autonomous taxes versus political authority</td>
</tr>
<tr>
<td></td>
<td>–2.29 to –2.31</td>
<td>Germany 1973–2001</td>
<td>Autonomous taxes versus political authority</td>
</tr>
</tbody>
</table>

Note. Cases listed are those for which the residuals of regressing the RAI on an alternative measure are equal or greater than two standard deviations. A negative sign indicates that the estimate of the alternative measure is smaller than the estimate of the RAI.
Explaining Government Structure

We measure area and population of a country on a logarithmic scale because the expected effect of an additional person or square kilometer declines as absolute population or territory increases. In order to estimate their causal effects we exert the following controls (see Appendix B at *Publius* online).

**Ethnicity.** Subnational communities—bounded groups of densely interacting humans sharing distinctive norms—are considered a magnet for regional authority. Ethnic communities often have distinct policy preferences. Rokkan (1983) suggests that culturally peripheral communities are predisposed to resist centralization and demand self-rule (Gellner 1983). Additionally, government is hypothesized to work best when it encompasses communities that “can draw on a reservoir of common cultural materials – language, experience, understandings about modes of interactions – that makes it easier for community members to communicate and work together” (Habyarimana et al. 2007, 711; Deutsch 1966).

**Inequality.** One expectation is that relatively rich regions will demand greater fiscal autonomy and poor regions will demand centralized redistribution (Bolton and Roland 1997; Dahl Fitjar 2008; Van Houten 2003). Alternatively, poor regions may prefer decentralization if their optimal fiscal policy differs substantially from that of the country as a whole (Alesina and Spolaore 2003, 63–67). Beramendi (2010) argues that poor regions with distinctive labor markets may want decentralization to implement appropriate labor market policies. The available data
do not allow us to model the interaction of regional inequality with regional fiscal or labor market variables over time. Our measure is the standard deviation of mean per-capita income across regions in thirty-nine countries.

**Affluence.** Laitin (1998) argues that citizens in richer countries may be willing to pay the costs of regional government to preserve cultural and linguistic diversity. Also, the policy portfolios of governments in richer countries may include a larger share of public goods that are efficiently delivered at the regional level (Osterkamp and Eller 2003; Peterson 1995).

**Democracy.** Democracies are considered to be more responsive than autocracies to demands for regionalization on account of their openness to societal pressures for self-rule (Hooghe, Marks, and Schakel 2010; Meguid 2009).

**Regional political parties.** Regional political parties press for greater regional authority. While regional political parties rarely form governments, a regional party may precipitate regional reform as a pivot or by inducing a party in government to steal its thunder (Hopkin and Van Houten 2009; Sorens 2009; Swenden and Maddens 2009).

### Results

The first column of Table 3 models Population and Area under controls. Whereas Area does not reach significance, Population has a large and significant effect on Regional Authority. Subsequent columns probe robustness. The second column reports a measure of multicollinearity, the variance inflation factor (VIF). Multicollinearity inflates the standard errors, making it harder to assess the independent effect of a variable. A VIF $<3$ is regarded as low, but it is still the case that because Area is associated with Population ($R = 0.69$) and Inequality ($R = 0.24$), its standard error is larger than it would be if it were uncorrelated with these variables. The inflation of the standard error for Area is $\sqrt{1.98} = 1.41$ which produces a VIF-adjusted standard error of $1.22/1.41 = 0.87$. If Area were to share none of its variance with the other independent variables, the estimate for its effect on Regional Authority still fails to reach significance ($t$-value $= 1.43$). The VIF-adjusted $t$-value for Population is 4.18.

The final columns of Table 3 estimate a lagged endogenous variable model and a jackknife test. Both tests indicate that Population is significant in explaining Regional Authority and that Area fails to reach significance. Controlling for past levels of Regional Authority minimizes concern that the association between Population and Regional Authority is spurious. The estimates for Population and Area are robust when we delete individual countries from the analysis. The analysis also confirms the significance of regional parties, democracy, ethnicity, and affluence. All trail population except for regional parties, but population comes out ahead when comparing VIF-adjusted $t$-values.
Table 3 Modeling the impact of population and area on regional authority

<table>
<thead>
<tr>
<th></th>
<th>RAI$^a$</th>
<th>VIF</th>
<th>Lagged dependent variable</th>
<th>Jackknife</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (clusters)</td>
<td>1603 (39)</td>
<td>1603 (39)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>5.68*** (2.07)</td>
<td>2.30</td>
<td>0.10** (0.05)</td>
<td>5.68** (2.41)</td>
</tr>
<tr>
<td>Area</td>
<td>1.24 (1.22)</td>
<td>1.98</td>
<td>0.01 (0.02)</td>
<td>1.24 (1.51)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>11.48** (4.61)</td>
<td>1.26</td>
<td>0.10 (0.11)</td>
<td>11.48* (6.11)</td>
</tr>
<tr>
<td>Inequality</td>
<td>−3.92 (5.79)</td>
<td>1.75</td>
<td>−0.02 (0.16)</td>
<td>−3.92 (7.11)</td>
</tr>
<tr>
<td>Affluence</td>
<td>1.79** (0.86)</td>
<td>1.11</td>
<td>0.06** (0.03)</td>
<td>1.79* (0.97)</td>
</tr>
<tr>
<td>Democracy</td>
<td>1.46*** (0.37)</td>
<td>1.23</td>
<td>0.00 (0.02)</td>
<td>1.46*** (0.45)</td>
</tr>
<tr>
<td>Regional parties</td>
<td>15.26*** (4.23)</td>
<td>1.33</td>
<td>0.16* (0.09)</td>
<td>15.26** (5.84)</td>
</tr>
<tr>
<td>Regional authority$_{t−1}$</td>
<td>−57.79 (12.43)</td>
<td>−</td>
<td>−0.78 (0.38)</td>
<td>−</td>
</tr>
<tr>
<td>Constant</td>
<td>−57.79 (12.43)</td>
<td>−</td>
<td>0.99*** (0.00)</td>
<td>−</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.62</td>
<td>0.995</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(F) test</td>
<td>31.12***</td>
<td>−</td>
<td></td>
<td>−</td>
</tr>
<tr>
<td>Mean</td>
<td>1.57</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Nonstandardized beta coefficients; robust standard errors clustered on country are listed in parentheses. ***\(p<.01\), **\(p<.05\), *\(p<.10\).

\(a\)Annual scores for 1950 to 2006 across thirty-nine democracies. Scores are calculated for each dimension at each regional tier and next aggregated to the country level weighted by population. For the operationalization of these variables, please consult Appendix B at Publius online.

Figure 2 estimates the effect of population on regional authority using Clarify. Uncertainty associated with the expected values of Regional Authority is greatest at the extremes. A country with a population of five million has an expected RAI between 7.1 and 9.7 within a 95 percent confidence band, with an average 8.4 as our best guess. This would, for example, be a country in which regional governments have indirectly elected assemblies that appoint regional executives with the power to set the rate of minor taxes and exercise competences, subject to central veto, in economic and welfare policy. A country with a population of fifty million has a mean expected RAI of 14.1 (within a 95 percent confidence band of 9.8 to 18.2). Regional governments in this country might then have elected assemblies, broader policy competencies not subject to central veto, power to set the base as well as the rate of minor taxes, and some role in national decision making, for example, by designating representatives to a second legislative chamber or by meeting routinely with the central government.

Lijphart (1999, 252), quoting Dahl and Tufte (1973), anticipates that a country’s location on the unitary/federal dimension is related to its population size, but this
is difficult to detect when countries are categorized as unitary or federal. Figure 3 plots country size and regional authority using the Lijphart measure (left-hand side) and the RAI (right-hand side) for the 24 countries common to the two datasets for 1995. Countries that Lijphart codes as unitary (left: broken line ellipse) and as federal (left: solid line ellipse) vary considerably in the RAI. The variation is particularly sharp among unitary countries. Iceland, for example, has an intermediate government (*landsvæðun*) that exists only as a statistical category; in both datasets it scores the minimum on regional authority. Greece and New Zealand, by contrast, have regional tiers with substantive authority, and this is reflected in the right-hand figure but not on the left. Greece’s regional tier consists of fifty-four prefectures (*nomoi*) which, from 1994, were run by directly elected councils and a council-selected prefect. Councils have competence over primary education, hospitals, roads, and transport. Central oversight remains extensive, and prefects continue to double as central state agents in urban planning and sanitation. New Zealand has sixteen regions run by directly elected councils that can set the base and rate of property taxes and which are responsible for public transport,
environmental policy, including air, land and marine pollution, river and coastal management.

Measurement can be decisive in assessing causality. This is revealed in Table 4 which compares results for common samples using different estimates of the dependent variable. The first column runs the model for all observations in Lijphart’s dataset. The second column replaces Lijphart’s measure with the RAI for the same set of cases. As one would suspect from Figure 3, Population is insignificant as a predictor using Lijphart’s measure and significant when using the RAI.

Subsequent columns pair alternative instruments with the RAI. Estimates for Population are insignificant while those for Area are significant for both the Arzaghi–Henderson measure and for the RAI. The result appears to be driven by the sample which is limited to sixteen countries with a population greater than ten million. These include Australia, Canada, and the United States which are vast and decentralized. The regression line connects these countries and the remaining ones. There is no association between Area and Regional Authority for the three English-speaking countries, and the association for the remaining thirteen countries is weakly negative.

Brancati selects countries to maximize variation on ethnic groups. This works well for her purpose, which is to evaluate the effect of decentralization on ethnic conflict (Brancati 2006, 2008), but produces estimates that are sensitive to outlying cases in the sample shared with the RAI. A panel jackknife dropping individual countries produces insignificant estimates for all independent variables.

Fiscal federalism measures, including those used by Panizza (1999) and Stegarescu (2005), confirm Area and disconfirm Population (final columns of Table 4). When we use the RAI for the same set of cases, we find precisely the reverse. A possible explanation is that the allocation of taxation across levels of government is peculiarly sensitive to territorial spillover on the ground that a local
Table 4 Sensitive dependence on alternative measures

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Lijphart</th>
<th>Arzaghi–Henderson</th>
<th>Brancati</th>
<th>Panizza</th>
<th>Stegarescu</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (clusters)</td>
<td>46 (24)</td>
<td>100 (16)</td>
<td>209 (16)</td>
<td>61 (24)</td>
<td>742 (23)</td>
</tr>
<tr>
<td>Dependent variable</td>
<td>Lijphart RAI</td>
<td>Arzaghi–Henderson RAI</td>
<td>Brancati RAI</td>
<td>Panizza RAI</td>
<td>Stegarescu RAI</td>
</tr>
<tr>
<td>Population</td>
<td>0.50 (0.53)</td>
<td>6.39** (2.43)</td>
<td>5.01* (2.36)</td>
<td>0.77 (0.53)</td>
<td>4.13 (3.28)</td>
</tr>
<tr>
<td>Area</td>
<td>0.36 (0.34)</td>
<td>0.56 (1.22)</td>
<td>4.60*** (0.88)</td>
<td>4.61*** (1.32)</td>
<td>0.20 (0.36)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>2.08 (1.67)</td>
<td>10.51* (5.15)</td>
<td>0.39 (2.87)</td>
<td>–1.78 (4.32)</td>
<td>2.02** (0.74)</td>
</tr>
<tr>
<td>Inequality</td>
<td>0.59 (2.43)</td>
<td>–0.61 (8.39)</td>
<td>−15.39 (11.57)</td>
<td>12.49 (12.98)</td>
<td>3.36 (1.95)</td>
</tr>
<tr>
<td>Affluence</td>
<td>–0.07 (0.22)</td>
<td>1.27 (1.18)</td>
<td>1.76 (1.35)</td>
<td>2.53** (1.17)</td>
<td>0.81 (0.61)</td>
</tr>
<tr>
<td>Democracy</td>
<td>0.81*** (0.26)</td>
<td>3.67** (1.03)</td>
<td>1.47** (0.53)</td>
<td>2.03** (0.74)</td>
<td>0.28*** (0.09)</td>
</tr>
<tr>
<td>Regional parties</td>
<td>1.35 (1.20)</td>
<td>12.29** (3.48)</td>
<td>54.25* (29.29)</td>
<td>90.37* (44.42)</td>
<td>–1.24 (0.91)</td>
</tr>
<tr>
<td>Constant</td>
<td>−12.21 (4.70)</td>
<td>−87.10 (19.80)</td>
<td>−57.18 (14.82)</td>
<td>−37.20 (15.50)</td>
<td>−11.24 (5.34)</td>
</tr>
<tr>
<td>R²</td>
<td>0.45 (0.45)</td>
<td>0.67 (0.76)</td>
<td>0.76 (0.64)</td>
<td>0.64 (0.71)</td>
<td>0.82 (0.46)</td>
</tr>
<tr>
<td>F-test</td>
<td>32.92*** (4.70)</td>
<td>29.74*** (19.80)</td>
<td>122.50*** (14.82)</td>
<td>35.89*** (15.50)</td>
<td>9.53*** (5.34)</td>
</tr>
</tbody>
</table>

Note. Nonstandardized beta coefficients; robust standard errors clustered on country are listed in parentheses. ***p < .01, **p < .05, *p < .10.
tax will be suboptimal if nonresidents living in a neighboring town use city services for which they are not taxed (Blöchliger and King 2006).

The evidence presented here asks us to think again about the spatial hypothesis for the structure of subnational government. While the arguments set out above linking area to government structure are plausible, their causal power is debatable. The first, spatial decay, explains decentralization as a cost of communication which is held to depend on distance. However, the reduction of communication costs since 1950 has not had the anticipated effect of increasing political centralization. Moreover, larger countries are not much more heterogeneous than small countries. The much-cited analyses of Panizza (1999) and Alesina and Spolaore (2003) are based on models in which distance from the center (or capital) of a country is a proxy for preference heterogeneity. However, the association between country size and the Fearon (2003) measure of ethnic diversity across thirty-nine democracies and quasi-democracies is just 0.078 (sig = 0.62). It is true that vast former colonies became home to diverse groups of immigrants, but these countries have high rates of geographical mobility, and as a result, they have few territorial minorities. By contrast, many European countries, despite their small area, contain territorial minorities with distinctive languages and cultures.

Conclusion

In recent years there has been renewed concern with the effect of measurement error on valid inference about political attitudes (Ansolabehere, Rodden, and Snyder 2008), ethnic conflict (Baldwin and Huber 2010), democracy (Coppedge, Alvarez, and Maldonado 2008), and political parties (Bakker et al. 2012; Marks et al. 2007)—to list but a few topics raised in recently published articles. Forty years ago, Blalock (1970, 1105) identified one of the chief problems: “A very common practice whenever measurement is clearly recognized as being crude is to resort to a relatively small number of ordered categories. In the extreme case the analyst may use dichotomies in order to simplify his analysis. It may not be recognized that such very simple procedures produce both random and nonrandom measurement errors that become increasingly serious as the number of categories is reduced.”

This article has argued that this is precisely the case in the study of government. The contrast between unitary and federal government is fundamental, yet it appears to have straight-jacketed efforts to estimate variation across time and space. It directs our attention to the rare event of constitutional choice, but away from reforms that do not shift a country from one category to the other. By conceiving government structure in dichotomous terms, the unitary/federal distinction is insensitive to change and overly sensitive to the creation of federal regimes in former British colonies.
Researchers who wish to measure government structure have added intermediate categories or they have used fiscal data on the share of subnational tax receipts or spending as an indicator of subnational authority. This is error-prone because a government’s fiscal envelope can be a poor guide to its capacity to decide how it spends.

The measurement instrument used here is grounded on the distinction between self-rule and shared rule made by federalism scholars themselves. The instrument is sensitive to variation among federal regimes and among unitary regimes as well as between them, and we find that the population of a country is more consequential for government structure than its area.

The population finding is consistent with the argument that government is shaped by a tension between centralization, which reduces the per-capita cost of a public good, and decentralization, which facilitates dialogue between government and citizens (Arrow 1961; Kochen and Deutsch 1969; Treisman 2007, 55–69).

The area hypothesis is consistent with the fact that federal countries, which include the United States, Australia, and Canada, are on average much larger than unitary countries. However, the association between area and government structure disappears when we estimate variation in decentralization among non-federal countries. Finland, France, Japan, Spain, and Sweden, for example, are similarly sized, but have contrasting levels of decentralization. Because it diminishes variation within each category, the unitary/federalism distinction is overly sensitive to the creation of federal regimes in the large spaces that were once British colonies and insensitive to variation at the other end of the scale.

Why should one care that the population of a country rather than its area is consequential for its government? A country’s footprint is an inert object fixed for decades or even centuries, but its population is alive and changeable. The causal mechanism by which area affects government structure, spillover, depends on distance. The causal mechanism by which population affects government structure, soft information, depends on the benefit of dialogue in producing public policy. While we know less about soft information than we know about distance, the line of theorizing pursued in this article may draw on—and perhaps contribute to— theories of democracy and participation.

The finding concerning the effect of population raises fundamental questions about how we should explain the structure of government. The size of a country’s population is a structural factor that lies far back, and perhaps at the very beginning, in the causal chain leading to political decisions about the creation of a new level of government or the allocation of authority across existing levels. There is much to learn about the mechanisms. How does the functional pressure of population and the demand for soft information get translated into the political pressures that shape decision making? Attempts to answer this question will need to connect theories of party competition and public policy with accurate estimation.
of government structure. By carefully bringing more information to bear on the
dependent variable, our purpose is to produce more valid estimates of causal
effects, and ultimately better theory.

**Supplementary Data**

*Supplementary data* can be found at www.publius.oxfordjournals.org.

**Notes**

Earlier versions were presented at the Conference for Europeanists, Chicago (2010),
the Annual Meetings of the American Political Science Association, Washington,
D.C. (2010), the European Union Studies Association in Boston (2011); the
Comparative Politics Working Group at the University of North Carolina, VU
University Amsterdam, Free University of Berlin, Charles University Prague, Duke
University, the European University Institute in Florence, Hertie School of
Governance in Berlin, University of Geneva, Leiden University, Mannheim
University, University of Pittsburgh, Salzburg University, and the IHS at Vienna.
The authors are grateful for comments by participants at these seminars and talks,
and would especially like to thank Michael Bauer, Pablo Beramendi, Tanja Börzel,
Jonathan Hopkin, Simon Hug, Markus Jachtenfuchs, Charlie Jeffery, Bonnie
Meguid, Jonas Pontusson, Thomas Risse, Arjan H. Schakel, Pascal Sciarini, Michael
Stein, Michael Tatham, Sonja Wälti, Carol Weissert, and four anonymous
reviewers, for helpful comments. We thank Emanuel Coman for research assistance;
Catherine De Vries, Jon Kropko and Michael Tatham for methodological advice;
and Dawn Brancati and Arjan H. Schakel for sharing unpublished data. The
authors are responsible for any remaining errors and inconsistencies. We are most
grateful to Tanja Börzel and the Kollegforschergruppe “The Transformative Power
of Europe” at the Free University of Berlin for their hospitality while we were
working on this article. This project is funded by the European Research Council
grant “Causes and Consequences of Multilevel Governance” # N.249543.

1. Brackets added.
2. The number of levels in the hierarchy is the natural logarithm of the population.
3. This is the notion of the social planner (Alesina and Spolaore 2003; Marks and Hooghe
   2000).
4. The online appendix details the measure and shows that the results reported here are
   robust when the number of tiers is treated as an independent variable or as a control in
   the model (Table B.6 at *Publius* online). The replication dataset is available on our
5. The scales are designed to have equivalent intervals. Regression models have been
   shown to be quite robust to distortions that could arise from smooth monotone
transformations, including summation of ordinal scores across rating scales (Shevlin, Miles, and Bunting, 1997).

6. Table B.5 in the appendix at Publius online lists the countries that each instrument has in common with the RAI.

7. Granger tests indicate that the causality runs from regionalist political party representation to regional authority. However, this cannot be generalized to party systems, for as Chhibber and Kollman (2004, 79) observe, “party systems become more national as governments centralize authority.”

8. The VIF for independent variable $j$ is $1/(1 – R_j^2)$, where $R_j^2$ is the squared multiple correlation from a regression of variable $j$ on all other independent variables in the model. The VIF$_j$ is proportional to the variance of variable $j$ explained by the other independent variables in the model. On VIF see O’Brien (2007).

9. The strong and significant effect of population is robust when we control for urbanization. To the extent that it is easier to coordinate public goods provision in densely populated areas, urbanization should reduce the effect of population on regional decentralization (Arzaghi and Henderson 2005). We test both the independent effect of urbanization and its interaction with population, but detect no significant effects.

10. In a paper written in the 1950s, Herbert Simon and his coauthors hypothesized that long-distance telephony makes it “unnecessary” for governments “to allow representatives much discretion” and that “improved methods of communication have made a much greater degree of centralization possible” (Simon et al. 1956, 275; 279). Of the countries dealt with here only Sweden and Germany have become more centralized since 1950, whereas twenty-nine have become more regionalized.

References


