

Part I
Measurement

Not for circulation

Not for circulation

1

Transparency in Measurement

This book is the first of four books theorizing the structure of governance above and below the central state. We describe the theory as postfunctionalist because it claims that governance, which we define broadly as authoritative decision making in the public sphere, is determined not just by its functionality but by its emotional resonance. Multilevel governance within the state, the topic of this book, evokes intense preferences not just for what it *does*, but for what it *is*. Jurisdictional design has intrinsic meaning for people. It expresses their national, regional, and local identities. The premise of post-functionalism is that this cannot be reduced to the extrinsic functions of governance. It is about “who are we” as well as “who gets what.”

This raises questions that can be answered only by looking within countries. Over the past two decades there has been an upsurge of research on territorial governance within countries, but measurement has lagged behind. Case studies investigate the mobilization of ethnic minorities and the efforts of central rulers to accommodate or suppress them, but the effects are only dimly perceived in national indices, and they escape fiscal measures entirely.

A measure is a disciplined summary. It attaches conceptual relevance to some phenomena and ignores others. As one begins to conceptualize variation in territorial governance, one enters a subterranean world in which there are numberless possibilities. Jurisdictional regions vary enormously in size and population. Their authority varies more than that of states. Some are merely central outposts for conveying and retrieving information. Others exert more influence over the lives of people living under their rule than the national state itself. One must leave behind the idea that territorial governance is constitutionalized, and therefore highly stable. Regional governance is governance in motion. The regional authority index detects more than 1300 changes in sixty-two countries.¹

¹ This is the number of changes of 0.1 or more on one of the ten dimensions for a region or regional tier.

Transparency in Measurement

Thirty-four new tiers of regional governance have been set up and seven have been abolished. Precise observation of territorial governance reveals a landscape that is fascinating in its flux and diversity.

This book sets out a measure of regional authority that can be used by social scientists to investigate the character, causes, and consequences of governance within the state. In this chapter we explain the key decisions that underpin our measure. How do we conceptualize regional authority? How do we summarize this abstract concept in dimensions? What indicators do we use to tap variation along these dimensions? And how do we score cases using these indicators? Each step is a theoretically motivated move from the abstract to the concrete. Subsequent chapters allow the reader to assess the validity of these steps and of the final product. Chapter Two compares our measure with other commonly used measures of decentralization. Chapter Three is a hands-on guide to the rules underpinning the measure and its indicators. The book concludes with profiles that overview change in regional authority across eighty-one countries on a common analytical frame.

We have three purposes. First, we wish to provide a reasonably valid measure of subnational government structure that is sensitive to cross-sectional and temporal variation. The measure conceives subnational governance as a multidimensional phenomenon that can take place at multiple scales. Fiscal measures provide annual data for a wide range of countries, but the amount of money that passes through a subnational government may not accurately reflect its authority to tax or spend. And there is much more to the structure of government than spending or taxation. Some regional governments can block constitutional change; some control local government, immigration, or the police; some play an important role in co-governing the country as a whole. The concept of federalism does a better job at capturing regional authority, but it is insensitive to reform short of constitutional change and does not pick up cross-sectional variation among federal or among unitary countries. The measure proposed here detects a lot of variation both within these categories and over time. The figures preceding the country profiles reveal that the territorial structure of government is much more malleable than is implied by the classics of comparative politics (e.g. Lijphart 1999; Riker 1964).

Our second purpose is to break open subnational government so that others may look inside. Comparative politics is conventionally seen as the study of politics across countries. Still, the field has a prominent and longstanding tradition of studying politics not just across, but also within, countries. Among the most celebrated examples are Tocqueville's *Democracy in America* (1838), which compares American states to assess the effects of slavery, Seymour Martin Lipset's *Agrarian Socialism* (1950), which compares wheat-belt provinces in Canada and the US, and O'Donnell's (1973) discussion of regions in Argentina and Brazil.

Measurement

The past decade has witnessed an upsurge in the number of articles and books comparing regions within and across countries.² The most obvious reason is that we live in an era in which authority has spun away from central states to subnational and supranational governments. We see this very clearly in our measure, which reveals an increase in the authority of regional governments in two-thirds (fifty-two) of the countries we observe. Another reason is that comparing regions can lead to better causal inference. Democracy, economic growth, crime, and many other things that people care about, vary within as well as among countries (Snyder 2001; Giraudy 2015; Giraudy, Moncada, and Snyder 2014).³

Subnational comparison can increase the number of relevant observations. More importantly, it can provide inferential leverage in engaging the fundamental problem with observational data: too much varies and the controls one can impose through matching and fixed effects are both demanding and incomplete. This is where subnational comparison is particularly useful. Many of the confounding factors that are difficult to control for are national, and controlling for national factors is a powerful lever for explaining variation against a background of commonality. This is precisely Robert Putnam's inferential strategy in *Making Democracy Work* (1993). Comparing regions in the north and south of Italy allows him to control for a wide array of factors—including Catholicism, parliamentarism, and the legacy of fascism—that could plausibly influence democratic performance.

This calls for measurement at the level of the individual region rather than the country—a decision that has shaped every aspect of this book. Examining territorial government inside countries brings to life phenomena that are otherwise invisible. More than half of the countries with a population greater than twenty million have not one, but two or more levels of intermediate government. An increasing number of countries are differentiated, that is, they have one or more regions that stand out from other regions. We wish to compare not just countries, but regions and regional tiers within countries. And we compare not only how regional governments exert authority over those living in its territory, but also how they co-govern the country as a whole. In short, the question we are asking is “In what ways, and to what extent, does a regional government possess authority over whom at what time?”

² This trend encompasses Western countries (e.g. Dandoy and Schakel, eds. 2013; Gerring, Plamer, Teorell, and Zarecki 2015; Kelemen and Teo 2014; Kleider 2014), Latin America (e.g. Giraudy 2015; Chapman Osterkatz 2013; Niedzwiecki 2014), Africa (Posner 2004), Russia (Robertson 2011), and China (Landry 2008; Tsai 2007).

³ For studies that are explicitly motivated by this insight, see e.g. Agnew (2014); Charron and Lapuente (2012); Gibson (2012); Harbers and Ingram (2014).

Transparency in Measurement

The product is a measure that provides information on the financial, legal, policy, representational, and constitutional competences of individual regions and regional tiers on an annual basis. Each of the ten dimensions of the measure picks up a distinct component of regional authority.⁴ We aggregate dimensional scores for regions and tiers to the country level, but researchers can re-assemble the constituent dimensions for their own purposes. They can also begin to examine the effects of variation in the way in which regional governments exert authority. Why, for example, do some regional governments exercise considerable powers within their own borders, but have almost no role in governing the country? What is the effect of tying a region into country-wide governance? How do fiscal, legal, policy, representational, and constitutional competences interact, and with what results? Why has subnational governance become more differentiated over time?

Our third purpose relates to measurement in general. How should one go about measuring a big abstract concept such as authority? In our 2010 book which introduced the regional authority index (RAI) we emphasized that it was vital to lay our method bare before the reader “so that others may replicate, amend, or refute our decisions” (Hooghe, Marks, and Schakel 2010: 3). We wanted to make it possible for others to evaluate how the measure was constructed, and we were intensely aware that our decisions were theory-driven. This is the commitment to transparency that has been set out by the American Political Science Association in a series of collectively authored statements. Beyond the well-recognized (though not always practiced) norm that researchers provide access to the data and analytical methods they use in their publications, the APSA (2012: 10) calls for production transparency: “Researchers providing access to data they themselves generated or collected, should offer a full account of the procedures used to collect or generate the data.”

Production transparency implies providing information about how the data were generated or collected, including a record of decisions the scholar made in the course of transforming their labor and capital into data points and similar recorded observations. In order for data to be understandable and effectively interpretable by other scholars, whether for replication or secondary analysis, they should be accompanied by comprehensive documentation and metadata detailing the context of data collection, and the processes employed to generate/collect the data.

⁴ The financial statistics produced by the International Monetary Fund (IMF) are a model worth emulating. The data take the form of a multidimensional matrix which breaks down financial flows by type of transaction, institutional unit, sector, and as discussed later, by jurisdictional level. “In contrast to summary measures, the detailed data of the GFS [Government Finance Statistics] system can be used to examine specific areas of government operation. For example, one might want information about particular forms of taxation, the level of expense incurred on a type of social service, or the amount of government borrowing from the banking system” (IMF 2014: 3). The RAI consists of ten dimensions and a larger number of indicators that can be individually analyzed and re-aggregated.

Production transparency should be thought of as a prerequisite for the content of one scholar's data to be truly accessible to other researchers. Analytic transparency is a separate but closely associated concept. Scholars making evidence-based knowledge claims should provide a full account of how they drew their conclusions, clearly mapping the path on the data to the claims (Lupia and Alter 2014: 57, citing a memo by Lupia and Elman 2010).

Production transparency is a public good that lies at the heart of the scientific method. Science operates by the light of day, by making the process of confirmation and disconfirmation explicit. This applies as much to measurement as to the methods used to analyze data. Estimating a political concept requires a series of theoretical, conceptual, operational, and coding decisions. Each step is a move from the general to the particular in which an abstract concept is translated into the language of numbers. Measurement, no less than theory, is "the art of discerning what we may with advantage omit" (Popper 1982: 44).

The process can be broken down into six steps.

- 1) *Defining the background concept.* How have social scientists understood the concept?
- 2) *Specifying the measurement concept.* Which of those meanings does one wish to include?
- 3) *Unfolding the concept into dimensions.* How does one break down the measurement concept into discrete pieces that can be independently assessed and aggregated to capture its meaning?
- 4) *Operationalizing the dimensions.* How does one conceptualize and specify intervals on the dimensions? What rules allow one to reliably detect variation across intervals?
- 5) *Scoring cases.* What information does one use to score cases? Where is that information, and how can others gain access to it?
- 6) *Adjudicating scores.* How does one interpret gray cases, i.e. cases for which scoring involves interpretation of a rule?

Figure 1.1 is an expanded version of Adcock and Collier's (2001) schema.⁵ The arrows are verbs to describe the steps down from the background concept

⁵ We make two additions. The first is a level of measurement, *dimensions*, in which the abstract concept is broken down into components prior to developing indicators. Virtually all concepts of major theoretical interest in the social sciences are complex in that they are comprised of more than a single dimension of variation. So an important step in operationalizing abstract concepts such as regional authority, democracy, or gross national product (GNP) is to conceive a limited set of dimensions that are amenable to operationalization and that together summarize the meaning of the overarching concept. The second addition is a final important step, *adjudicating scores*, which lays out rules for exceptional or difficult cases that arise in any coding scheme. Social science measurement is replete with gray cases, and one telling indication of the transparency of a measure is whether these are explicitly communicated.

Transparency in Measurement

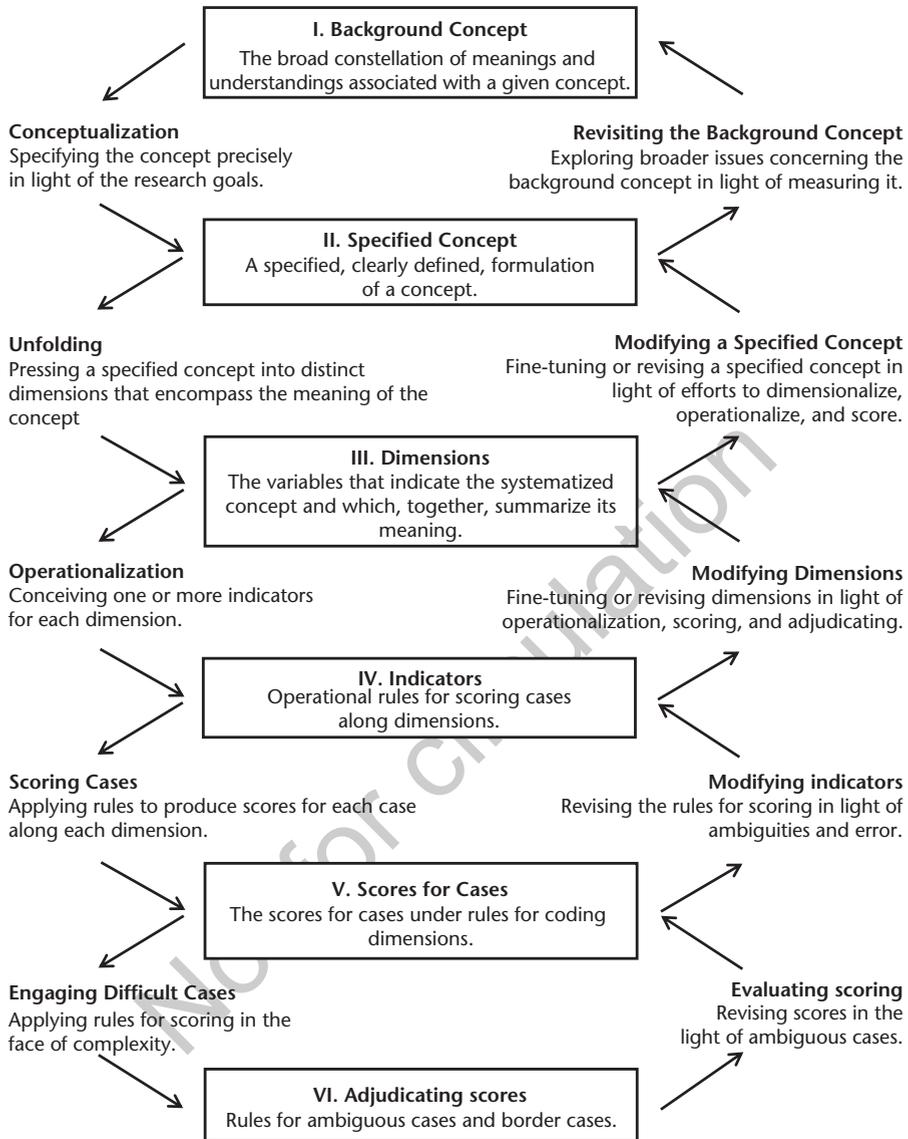


Figure 1.1. Measurement model

to individual scores or up from individual scores to the background concept. The boxes contain nouns to describe the concept, its dimensions, indicators, and scores as one presses the concept closer to phenomena that can be observed at lower levels of abstraction. The figure makes the point that these steps are interdependent. How one specifies the scope of a concept has consequences for breaking it into dimensions. How one operationalizes those

dimensions frames the choice of appropriate indicators. Even minor differences in the indicators can have serious consequences for scoring.

Making this transparent is good for several reasons. Transparency facilitates replication. It is true that we rarely replicate each other's results, but the possibility of replication has an effect on the quality of science that reaches beyond its incidence. Most findings will never be replicated, but the more influential a finding, the greater the likelihood it will be replicated. Replication is insurance for Richard Feynman's (1985: 343) first principle of science: "[Y]ou must not fool yourself, and you are the easiest person to fool."⁶

Transparency allows others to understand and probe the inner workings of a measure, and this can help in assessing its validity. A dataset is a matrix of decisions that cannot, even in principle, be inducted from the numbers that appear in the cells. One must have access to those decisions to assess the numbers. Transparency directs attention to the construction of a measure, and exposes the decisions that underpin it. I have little direct knowledge of how the gross domestic product (GDP) of the US grew in the last quarter, but I do have direct knowledge of the process by which the data were collected (Landefeld et al. 2008). I have little direct knowledge of the people who, in the week of September 8, 2015, intended to vote in favor of Scottish independence, but I do know (or should know) how a survey instrument was constructed, how the population was sampled, and how the survey was conducted.

However, transparency can do more than tell one how a measure is produced. It can allow others to evaluate the validity of the scores for individual cases. We can be reasonably sure that some experts will know more about the structure of government in their country than we will ever know. Transparency can reveal the evidence and reasoning that go into individual scores. Let others see how one arrives at particular scores for cases with which they are deeply knowledgeable. Let them have access to the judgments that produce scores for gray cases. This is why we devote considerable space to country-specific profiles that provide an overview of regional governance and explain how we score particular regional reforms in a country. Explaining the construction of a measure and investigating its reliability are not at all the same as explaining how individual cases are scored. However, it is the scores for individual cases that are of most use-value. The profiles provide a birds-eye view of regional governance across a wide range of countries on a common

⁶ Or, as Alexander Pope ([1734] 1903: 157) wrote

To observations which ourselves we make,
We grow more partial for th' observer's sake.

Transparency in Measurement

format. By making our judgments explicit we can ask experts: “Have we used the appropriate evidence?” “Do our judgments make sense?”

Measurement, Error, and Fallibility

Measurement is inherently prone to error. This is the thrust of Lakatos’ philosophy of scientific method, which rejects the demarcation of measurement and theory (Lakatos 1970; see also Bouwmans 2005). Measurement maps a property of the empirical world onto a set of numbers, a procedure that requires a series of inferential steps. In the words of a contemporary philosopher of science: “Measurement involves a host of theoretical and statistical representations of measuring systems and the data they produce” (Tal 2013: 1164). Social scientific measurement is at least as inferentially complex as measurement in the physical sciences, so it is worth taking epistemologists seriously when they point out that “physicists are forced to test the theories of physics on the basis of the theories of physics” (Chang 2004: 221). An observation is a theoretically guided experiment that produces information by making claims about what is observed and how it is observed. The philosopher–scientist Pierre Duhem ([1906] 1954: 182) stresses that “it is impossible to leave outside the laboratory door the theory we wish to test, for without theory it is impossible to regulate a single instrument or to interpret a single reading.”

The appearance of hard facts is deceptive even in the measurement of something as basic as temperature. Comparing temperature observations in different places called for some well defined fixed points. The temperature of the human body and that of the cellar in the Paris observatory provided useful (but not entirely reliable) fixed points until Anders Celsius created a universal scale using the boiling point and freezing point of water. Evidently Celsius conceived his scale as a measure of degrees of cold, not heat. Water boiled at 0° on Celsius’ original scale, and froze at 100° (Beckman 1997; Chang 2004: 159ff). Early thermometers used either alcohol or mercury. But the premise that alcohol and mercury thermometers could be made to “speak the same language” was disconfirmed when Réaumur found that recalibration from one to the other failed to produce uniform readings (Gausson 1739: 133; Réaumur 1739; Chang 2001). Mercury became the standard because the rate at which it expanded approximated the ratios of mixing ice and boiling water.⁷ However, this assumed that mercury thermometers would give uniform readings if they were made of different kinds of glass, and more fundamentally, it assumed

⁷ Or, more precisely, nearly freezing and nearly boiling water (Chang 2004: note 27).

Measurement

that the temperature is an additive function of the ratio of freezing and boiling water. When the linear theory of mixing was disconfirmed, thermometers used gas on the ground that the molecular interactions that produced non-linearity in liquids would be nearly absent in gas. Gas thermometers were accurate for most purposes, although the technology has moved on and the current International Temperature Scale has the boiling point of water at atmospheric pressure as 99.975°C rather than 100°C.

Having an accurate thermometer is just the first step in reliably measuring global temperature.⁸ Many measurement stations are located near population centers that are warmer than the surrounding areas. Irrigation has the opposite effect. The coverage of many parts of the globe, including particularly the hottest and coldest regions, is incomplete. Not only are estimates inexact, but there are numerous sources of systematic bias. Ships now measure ocean surface temperature with water flowing through engine cooling water intakes rather than with water collected in buckets (Matthews 2013). The introduction of the new method coincides with a rise in ocean temperature in the 1940s, perhaps because water collected in buckets cooled prior to measurement. Social factors come into play. Daily mean temperatures are calculated by summing the maximum and minimum over a twenty-four-hour period and dividing by two. However, volunteer weather observers have an understandable reluctance to take midnight readings, and until the 1940s most weather stations recorded the maximum and minimum temperatures for the twenty-four hours ending near sunset (Karl et al. 1986). Scientists seek to correct these and other possible sources of bias using proxies such as satellite measurement of the intensity of night light to adjust for the urban heating effect. None of these potential biases is large enough to shake the inference that global warming is taking place, but they do lead an expert inquiry to emphasize that on account of urbanization and observational irregularity, “Temperature records in the United States are especially prone to uncertainty” (Hansen et al. 2010: 103).

No less than in the physical sciences, measurements in the social sciences are based on a series of inferences, each of which can be questioned. The general lesson is that no observation can sit in judgment of a theory without being cross-examined. And there is no reason why the interrogation of an observation should be less searching than the interrogation of a theory. The implication that Lakatos draws from this is that “clashes between theories and factual propositions are not ‘falsifications’ but merely inconsistencies. Our imagination may play a greater role in the formulation of ‘theories’ than in the formulation of ‘factual propositions’, but both are fallible” (Lakatos 1970: 99–100).

⁸ We thank Michaël Tatham for drawing our attention to this.

Transparency in Measurement

All observation is fallible, but some observations are more fallible than others. Social scientists are concerned with highly abstract concepts, many of which have normative connotations. The chain of inference that links the observation of a particular behavior to the concept of democracy, inequality, or decentralization is both long and complex. In this endeavor the assumption that measurement error is random rather than systematic is false comfort, for it suggests that issues of validity can be reduced to issues of reliability. One of the purposes of observation is precisely to discipline our theories or “guesses” (Feynman 1965: 156). However, this takes the form of a conversation rather than a judgment, for the observations that one brings to bear are themselves built on a scaffold of theoretically motivated short-cuts.

Perhaps in no other field of political or economic science is this more apparent than in the study of the structure of government, and decentralization in particular. Theoretical expectations often line up on both sides of the street, but the information that is used to test them can be slippery. Weak theory and poor measurement are complementary because almost any set of observations appears consistent with one or another theory. Summarizing the effects of decentralization for economic performance and the quality of government, Treisman (2007: 5) writes that “as one would expect given the uncertain and conditional results of theory, almost no robust empirical findings have been reported about the consequences of decentralization.”

An extensive literature takes up the question of the effect of decentralization on the size of the public sector. This is the “Leviathan” question introduced by Brennan and Buchanan (1980): Is government intrusion in the economy smaller when the public sector is decentralized? Brennan and Buchanan argue that it is, but others have developed plausible models that claim exactly the opposite (e.g. Oates 1985; Stein 1999). Intervening variables can change the sign of the effect. Oates (2005) argues that “it is not fiscal decentralization *per se* that matters, but what form it takes” (Oates 2005; Rodden 2003a; Jin and Zou 2002).

The standard measure of decentralization in this literature is World Bank data derived from the IMF’s Government Finance Statistics (GFS) measuring subnational expenditures or subnational revenues as a proportion of total government expenditures or revenues.⁹ Data are rarely reported for the two tiers of subnational government in the GFS framework, and the criteria for intermediate and local government vary across countries. Several countries, including France, Italy, Japan, the Philippines, and Thailand, have no

⁹ The World Bank is explicit about the limitations of these data: “Shared taxes appear as sub-national revenue, although the sub-national government has no autonomy in determining the revenue base or rate, since the GFS reports revenues based on which level of government ultimately receives the revenues.” <<http://www1.worldbank.org/publicsector/decentralization/fiscalindicators.htm#Strengths>>.

Measurement

intermediate tier of government in the dataset because their regions are reported as local government. Belgian communities, which form one the strongest intermediate levels of government anywhere, are classed as part of central government with the result that Belgium comes out as the most centralized country in the Organisation for Economic Co-operation and Development (OECD).¹⁰

But the more fundamental issue is conceptual. Is the amount of money that a subnational government raises or spends a valid measure of decentralization (Rodden 2003a)? If decentralization involves the authority to make decisions, the answer must be “not necessarily.” Sweden, Norway, Finland, and Denmark—big spending governments with a history of social democratic rule—are considered to be highly decentralized because they channel considerable funds through their local governments. However, local governments in these countries spend and tax according to national laws (see Chapter Two). The IMF data consider these countries, on average, to be as decentralized as the US or Germany, and more than twice as decentralized as Spain, Italy, or France. Perhaps not surprisingly, a recent paper using these data concludes that “fiscal decentralization leads to larger public sectors when the federal government is controlled by a left-wing party, and to smaller public sectors when it is controlled by a right-wing party” (Baskaran 2011: 500).

The most commonly used alternative measure in the Leviathan literature is a dichotomous variable that distinguishes federal from non-federal countries. This has surface validity, but it is useful only in cross-sectional analysis because few countries cross the federal divide. This variable also censors variation within each category. Non-federal countries include both highly centralized countries, such as El Salvador and Luxembourg, and countries, such as Indonesia and Spain, which in our data are more decentralized than several federal countries. Knowing whether a study uses this federalism variable or IMF fiscal data helps one predict whether that study confirms or disconfirms the hypothesis that decentralization reduces public spending. A meta-analysis (Yeung 2009: 22) concludes that “Despite over 36 years of research, little consensus has emerged on the effect of fiscal decentralization on the size of government” and that the reasons for disagreement have to do with theoretical and conceptual choices that are implicit in “a study’s unit of analysis and measure of decentralization.”

Every measure produces information by making theoretical and conceptual claims about the world. A measure of regional authority can no more be insulated from theory than a measure of temperature. Neither theory nor data can sit in judgment on the other. Rather they need to be brought into a

¹⁰ Similarly, Scotland and Wales are assessed as part of the UK central government (IMF 2008: 546).

Transparency in Measurement

dialogue in which each is regarded as fallible. Charles Darwin, who spent much of his life making careful observations, remarked that “a good observer really means a good theorist” (Darwin 1903: 82).

Nuts and Bolts

We seek to measure the authority exercised by regional governments in eighty-one countries on an annual basis from 1950, or from the time a country becomes independent, to 2010.¹¹ The sample consists of all European Union (EU) member states, all member states of the OECD, all Latin American countries, ten countries in Europe beyond the EU, and eleven in the Pacific and South-East Asia.¹²

Table 1.1 lists four prior measures of regional authority by year of publication. Measurement has become more comprehensive over time, providing more information for more years. The measure set out here continues this development and has some unique features.

Most importantly, the unit of analysis is the individual region, which we define as a jurisdiction between national government and local government.

Table 1.1. Measures of regional authority

	Lijphart (1999)	Woldendorp, Keman & Budge (2000)	Arzaghi & Henderson (2005)	Brancati (2008)	Regional Authority Index (2016)
Country coverage	36 pre-1990 Western democracies	37 Balkan, OECD, EU democracies	48 countries with population > 10 million	37 countries with regional ethnic groups	81 Western, post- communist, Latin American, Southeast Asian & Pacific countries
Time coverage	1945–1996	1945–1988	1960–1995	1985–2000	1950–2010
Time points	1	1	8	16	61
Individual regions	no	no	no	no	yes
Multiple tiers	no	no	no	no	yes
Observations per country/year	5	4	8	5	10–130

¹¹ On average a country in the dataset is coded for forty-seven years. Forty-eight countries are coded for the entire 1950–2010 period.

¹² The case selection reflects a trade-off between an effort to cover the largest possible number of countries and the team’s resources—chiefly their time—and the availability of sources and country expertise.

Measurement

We draw the boundary between local and regional government at an average population level of 150,000. This excludes the lowest tier of government in all eighty-one countries, but allows us to capture intermediate governments, often arrayed at two nested levels between the local and national. We relax the population criteria for individual jurisdictions, such as Greenland or the Galapagos islands, that stick out from a tier of government that meets the regional threshold.

A focus on regional or intermediate government has some theoretical and practical virtues. It encompasses virtually all subnational governments that exert self-rule within distinct homelands. Such governments tend to form part of a regional tier of government with an average population greater than 150,000 or they have special authoritative competences alongside a regional tier. Where subnational governments play an important role in co-governing a country, these are almost always intermediate governments. To the extent that subnational governments play a formally recognized role in shaping constitutional reform, one needs, again, to look to the intermediate level. Yet many countries lack any form of intermediate governance or have regional governments that are merely deconcentrated. Regional jurisdictions are the most variable elements of territorial governance within the state and are generally the most contested.

The decision to conceptualize the individual region as the unit of analysis has several consequences. It raises the possibility that regions may be nested within each other at different scales. Altogether, there are 103 levels of regional government in the sixty-five countries that have at least one tier of regional government. So researchers can begin to compare regional tiers within countries. The measure picks up reform even when it is limited to a single region in a country. A reform in a single region may not seem much, but if it undermines the norm that all be treated equally, it may be hotly contested by other regions as well as the central government. Moreover, such a reform may threaten the break-up of the state.

The measure comprises ten dimensions that tap the diverse ways in which a region may exert authority. These dimensions are quite strongly associated with each other and can be thought of as indicators of a latent variable. Yet those who are interested in examining the pathways to regional authority can disaggregate regional authority into its components. Some dimensions, including those that tap regional representation, policy scope, and borrowing autonomy, exhibit more reform than others.

Combining a regional approach with fine grained attention to the ways in which a region can exert authority produces a measure that is considerably more sensitive to change than any previous one. Twenty-one percent of the variation occurs over time. The territorial structure of governance is much less fixed than one would assume when reading the classics of comparative

Transparency in Measurement

politics such as Arend Lijphart's *Patterns of Government* (1999) or Daniel Elazar's *Exploring Federalism* (1987).

However, the RAI is limited in some important respects. Three stand out. We do not encompass tiers of subnational government containing jurisdictions with an average population less than 150,000. Hence, we omit local government entirely. This is a topic that calls for systematic measurement, perhaps adapting the measure proposed here to variation in the policy responsibilities of local authorities (Campbell 2003; Loughlin, Hendriks, and Lidström 2011; Nickson 2011; Norton 1994; Page and Goldsmith 1987, 2010).¹³

The regional authority index excludes informal arrangements. It is concerned exclusively with authority, which we define as formal power expressed in legal rules. Hence it omits contextual factors, such as leadership, political parties, or corruption, which may affect government performance. Finally, the country coverage of the present measure is incomplete. In particular, it does not cover China or India, two continental sized countries with correspondingly complex and differentiated systems of regional government.

I. *The Background Concept: Political Authority*

Political authority is a core concern of political science, some would argue *the* core concern (Eckstein 1973; Lake 2010; Parsons 1963; Weber 1968). Political authority—the capacity to make legitimate and binding decisions for a collectivity—underpins human cooperation among large groups of individuals. Human beings cooperate in order to produce goods that they could not produce individually. These goods include law, knowledge, and security. These goods are social in that they benefit all who live in the collectivity, and they are inclusive in that their benefits cannot practically be limited to those who contribute for them.¹⁴ Whereas small communities can impose social sanctions to produce public goods, large groups are far more vulnerable. The exercise of political authority diminishes the temptation to defect from collective decisions, and reassures those who do cooperate that they are not being exploited: “For although men [in a well ordered society] know that they share a common sense of justice and that each wants to adhere to the existing arrangements, they may nevertheless lack full confidence in one another. They suspect that some are not doing their part, and so they may be tempted not to do theirs” (Rawls 1971: 211).

Authority is relational: *A* has authority over *B* with respect to some set of actions, *C*. This parallels Robert Dahl's (1957: 202–3, 1968) conceptualization

¹³ A team led by Andreas Ladner and Nicholas Keuffe is adapting the RAI to estimate local decentralization in thirty-eight countries (personal communication, March 2015).

¹⁴ The negative formulation is that public goods are non-excludable and non-rivalrous.

of power as the ability of *A* to get *B* to do something that *B* would not otherwise do. A short-hand definition of authority is *legitimate* power. One speaks of authority if *B* regards *A*'s command as legitimate and correspondingly has an obligation to obey. Authority implies power, but power does not imply authority. Whereas power is evidenced in its effects irrespective of their cause, authority exists only to the extent that *B* recognizes an obligation resting on the legitimacy of *A*'s command. Such recognition may have diverse sources, including charisma, tradition, and religion (Weber 1958). This book is concerned with the modern variant of authority—legal-rational domination based in a codified legal order.

Two conceptions have predominated in our understanding of the structure of authority. The first conceives a polity as grounded in human sociality. Families, villages, towns, provinces, and other small or medium scale communities are the ingredients of larger political formations. This idea is as close to a universal principle in the study of politics as one is likely to find. Ancient states and tribes were composed of demes, wards, or villages. Aristotle conceived the polis as a double composite: households within villages; villages within the polis. Each had a collective purpose and a sphere of autonomy. The Romans built a composite empire by attaching a vanquished tribe or polis by a *foedus*—a treaty providing self-rule and protection and demanding payment of a tax, usually in the form of manpower (Marks 2012). The Qin dynasty that united China in 221BC had a four-tiered structure extending from the family through wards and provinces to the empire (Chang 2007: 64). The Incas conceived of five hierarchically nested tiers reaching from the family to an empire encompassing much of contemporary Peru, Bolivia, Ecuador, and northern Chile (Rowe 1982). Medieval scholars conceived the state as a composite (*consociandi*) of men already combined in social groups (*sympiototes*). Johannes Althusius (1997 [1603]) conceived the state as a contract among such associations, a *consociatio consociationum* consisting of families within collegia within local communities within provinces.

The modern variant of this idea is federalism, which describes a polity “compounded of equal confederates who come together freely and retain their respective integrities even as they are bound in a common whole” (Elazar 1987: 4). Federalism highlights the basic constitutional choice between a unitary and federal system. A unitary system has a central sovereign that exercises authority, whereas a federal system 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; Dahl 1986: 114). Most importantly, regions or their representatives can veto constitutional reform. The unitary/federal distinction informs a literature on the political consequences of basic constitutional decisions, including particularly ethnic conflict (Amoretti and Bermeo 2004;

Transparency in Measurement

Lijphart 1999). Federalist scholars have told us a lot about why independent units would wish to merge and how some polities arrive at federalism in order to avoid falling apart (Rector 2009; Roeder 2007; Stepan, Linz, and Yadav 2011). And there is a rich literature comparing federal polities (Watts 1998, 1999a, 2008).¹⁵

The federal/unitary distinction draws attention to the tension between self-rule and shared rule that is inherent in a composite polity. The constituent communities wish to retain their independence, their distinct way of life, their language, religion, dress, customs, their norms of social interaction. Yet they wish also to gain the benefits of scale in security, trade, and governance by forming a state in which they share rule with the center. As we discuss later, the concepts of self-rule and shared rule motivate our measurement scheme, and they are taken directly from the federalism literature.

However, the unitary/federal distinction has some fundamental limitations for the measure we propose. It is a blunt instrument for assessing incremental institutional change. Shifting from a unitary to a federal regime (or the reverse) is a high hurdle that few countries meet. The number of federal countries in our dataset has hardly changed over the past sixty years, yet there is ample evidence that this has been a period of profound reform.¹⁶ Not surprisingly, the federalism literature tells one far less about variation among unitary countries than among federal countries (Hooghe and Marks 2013; Rodden 2004; Schakel 2008). Variation among unitary countries has grown a lot over the past six decades, whereas the contrast between unitary and federal countries has diminished. Finally, federalism is concerned with the topmost level of subnational governance, whereas several countries have two or three levels of government between the national and the local.

A second conception, the idea that governance can be more or less decentralized, has also been hugely influential. Centralization and decentralization are poles of a continuous variable describing the extent to which authority is handled by the central government *versus* any government below. This way of conceiving governance is elegant and thin. Both its virtues and vices arise from its very high level of abstraction. It travels well. It allows one to compare governance around the world and over time on a single scale.

¹⁵ There has been a veritable revival in the study of federalism. Recent examples include Anderson (2012); Bednar (2009); Benz and Broschek (2013); Bolleyer (2009); Burgess (2012); Chhibber and Kollman (2004); Erk (2008); Falleti (2010); Rodden (2006); Swenden (2006); Rodden and Wibbels (2010). This wave also comprises several handbooks, such as Loughlin, Kincaid, and Swenden (2013) on federalism and regionalism, and Haider-Markel (2014) on state and local relations in the US.

¹⁶ As Gary Goertz (2006: 34) observes, dichotomous concepts tend “to downplay, if not ignore, the problems—theoretical and empirical—of the gray zone. Often, to dichotomize is to introduce measurement error . . . [because it] implies that all countries with value 1 are basically equivalent.”

We seek to develop a measure that is similarly robust across time and place. If the RAI is aggregated to the country level it can be interpreted as a measure of decentralization. We follow decentralization scholars by distinguishing forms of decentralization: over policy making; over fiscal policy; over the appointment of subnational decision makers; and over the constitution. Each can be considered an independent variable that can register change in the absence of sweeping constitutional reform.

However abstractness has a price if it comes “at the expense of connotation” (Sartori 1970: 1051). Decentralization, but to which level of governance? Knowing whether a state is more or less centralized tells one nothing about which tier does what. Decentralization measures focus on the central state, lumping together all levels of subnational governance as “the other,” the non-central state. This can be a useful simplification in cross-national comparison, but it severely restricts the study of governance within the state. It has nothing to say to cases where one level of regional governance is empowered at the expense of another. “How does one compare two three-tier systems, A and B, when in A one-third of the issues are assigned to each of the tiers, while in B 90 percent of the issues are assigned to the middle tier and 5 percent each to the top and bottom tiers” (Treisman 2007: 27; Oates 1972: 196). One needs to map individual regions and regional tiers to probe variation in multilevel governance.

The measure we propose builds on the concepts of federalism and decentralization (Enderlein et al. 2010; Oates 1972, 2005, 2006; Stein and Burkowitz 2010). Both ways of thinking about authority have been influential in our work, as in the discipline of political science as a whole. From federalism, our measure takes the idea that regional authority consists of distinct forms of rule: self-rule within a region and shared rule within the country as a whole. This provides us with the conceptual frame for our measure. From decentralization, the measure takes the idea that the structure of government can be measured along continuous variables that together summarize regional authority.

II. *The Specified Concept: Validity and Minimalism*

Our focus in this book is on legal authority which is

- *institutionalized*, i.e. codified in recognized rules;
- *circumscribed*, i.e. specifying who has authority over whom for what;
- *impersonal*, i.e. designating roles, not persons;
- *territorial*, i.e. exercised in territorially defined jurisdictions.

These characteristics distinguish legal authority from its traditional, charismatic, and religious variants. Weber (1968: 215–16) observes that “In the case

Transparency in Measurement

of legal authority, obedience is owed to the legally established impersonal order. It extends to the persons exercising the authority of office under it by virtue of the formal legality of their commands and only with the scope of authority of the office." The exercise of legal authority over a large population involves a minimum level of voluntary compliance with codified rules that have a specific sphere of competence, and which are exercised through formal institutions, including a differentiated administration (Weber 1968: 212–17).

A focus on legal authority has two benefits. The first is that it distinguishes the structure of government from causally related but conceptually distinct phenomena such as the organization of political parties, the ideological beliefs of those in office, or the incidence of corruption. The second is that legal authority can be evaluated using public records: constitutions, laws, executive orders, statutes, or other written documents which are publicly available to researchers who can confirm, revise, or refute our coding decisions.

Our approach is minimalist. Minimalism is a concept used in design to expose the essence of a form by eliminating all non-essential features. In measurement this is the effort to specify the essential properties of a concept by eliminating its superfluous connotations. This avoids entangling phenomena that one wishes to explore empirically. If a measure of subnational authority were to include an indicator for party centralization it would not help one investigate how party organization shapes the structure of government.

Minimalism and validity often exist in tension. Public spending might be considered a minimalist indicator of decentralization, but the proportion of public expenditure that passes through a subnational government does not tell us whether that government can determine spending priorities (see Chapter Two).

Where the rule of law is weak, informal practices may undercut provisions codified in law. Bertrand (2010: 163) summarizes the problem: "[A]utonomy can sometimes become an empty shell. Powers may exist in law, but are subsequently undermined by the central state. For instance, the central state can enact other legislation that might contradict the autonomy law. By various bureaucratic or extra-institutional means, it might also slow or stall the autonomy law's implementation. Repressive policies might be launched after the autonomy law is passed, thereby reducing its meaning and ultimately its legitimacy" (see also Eaton et al. 2010; Varshney, Tadjoeeddin, and Panggabean 2008). In many regimes, as O'Donnell (1998: 8) observes, "Huge gaps exist, both across their territory and in relation to various social categories, in the effectiveness of whatever we may agree that the rule of law means."

The measure we propose taps authority codified in law, but we do not interpret this mechanistically. Some written rules never make it into practice. If the constitution states that subnational governments may tax their own populations, yet enabling legislation is not enacted (as in *departamentos* and

provincias in Peru), then we do not consider the regions to have fiscal authority.¹⁷ Similarly, we code the date when a reform takes place, not when it is prescribed in legislation.¹⁸

We estimate reforms that are not enacted in law if they are codified in executive orders, decrees, or edicts that are considered legally binding. For example, we take into account the capacity of a central state to sack regional governors, as in Argentina under military rule, even though it had a flimsy legal basis. Article six of the Argentine constitution allows federal intervention only in a handful of circumstances such as civil war and violation of the constitution, but when a military junta came to power in 1966, it drafted a military decree, the *Acta de la Revolución*, which sanctioned centralization and the abrogation of civilian rule (Potash 1980: 195–6).

Eaton, Kaiser, and Smoke (2010: 24) point out that “complete institutional analysis must consider informal social norms that govern individual behavior and structure interaction between social actors.” This is true, but no measure should try to cover the entire field. To what extent should one include informal social norms in a measure of regional authority? This depends on the purpose of the measure. On the one hand, we wish to evaluate the concept of regional authority broadly to capture its reality, not just its appearance. On the other hand, we want to make it possible for researchers to investigate the causal links between the structure of government and its causes and consequences. If we included indicators for regime type, corruption, or clientelism in a measure of regional authority this would complicate causal inference.

For the same reason we leave partisanship and party politics aside. Regional governments may be more assertive if they have a different partisan complexion from that of the central government, but our focus is on the rules of the game rather than how they affect behavior. In Malaysia, for example, we code the capacity of Sabah and Sarawak to levy an additional sales tax without prior central state approval, even though this authority was used only from 2008 when opponents of the ruling *Barisan Nasional* coalition won regional

¹⁷ The 1933 and 1979 constitutions gave *departamentos* extensive fiscal authority with the capacity to set rate and base of certain taxes. However these provisions were not translated in enabling legislation, and a 1988 law mandating that national government would transfer property and income tax to the regions within three years was not implemented (Dickovick 2004: 7). The 1979 constitution also appeared to give *provincias* extensive fiscal authority, including property tax, vehicle tax, and construction tax (C 1979, Art. 257), but consecutive governments have interpreted these competences narrowly and continue to set the base of all taxes while imposing narrow bands for rates (Ahmad and García-Escribano 2006: 15; von Haldenwang 2010: 651).

¹⁸ The gap between legislation and implementation can be extensive. In South Korea it took twelve years for the Local Autonomy Act of 1988 to come into force. We code only the parts of the reform at the time they are implemented by enabling legislation (Bae 2007; Choi and Wright 2004). In Argentina, the 1994 constitution introduced direct elections for senators to replace appointment by the provincial legislature. The first direct elections took place in 2001, which is when we score direct election.

Transparency in Measurement

elections. If one is interested in finding out how political parties affect the exercise of authority, it makes sense to estimate political parties independently from the structure of government (Chhibber and Kollman 2004; Harbers 2010; Hopkin and Van Houten 2009; Riker 1964).

Regime variation poses a particular challenge given the expectation that dictatorship and centralization are related (Bird and Vaillancourt 1998; Elazar 1995; IADB 1997; Leff 1999). We want to pick up the effect of a regime in constraining or facilitating regional authority, but we do not want to build regime type into a measure of regional authority. One can expect authoritarianism to bias subnational relations toward centralization, but this is not a black-and-white phenomenon (Eaton 2006; Eaton et al. 2010; Gibson 2004; Montero and Samuels 2004; O'Neill 2005; Willis, Garman, and Haggard 1999). Authoritarian regimes typically suspend or abolish subnational legislatures or executives, but the extent, form, and timing varies considerably.

Some examples suggest the need for a nuanced approach. Whereas the *Revolución Argentina* (1966–72) replaced all elected governors and put provincial legislatures under military control, the coups in 1955 and 1964 left subnational institutions more or less intact (Eaton 2004a; Falleti 2010). The military regime in Brazil (1964–82) maintained direct elections for governorships for three years before requiring regional assemblies to select governors from a central list (Samuels and Abrucio 2000). Regional assembly elections were never canceled. Cuba's Castro regime sidelined provincial and municipal institutions in favor of sectoral juntas, but reintroduced them in 1966 (Roman 2003; Malinowitz 2006; Mendez Delgado and Lloret Feijoo 2007). In Indonesia, centralization under authoritarian rule was incremental. Provincial and municipal legislatures continued to be elected even under Suharto, and subnational executives were gradually brought under central control. In 1959, regional governors became dual appointees; in 1974, they were centrally appointed; and from 1979 the central government appointed mayors and district heads as well.

We also see some exceptional cases in which authoritarian rulers create a new regional level. In Chile, Pinochet created an upper level of fifteen deconcentrated *regiones* to empower his rural constituencies. He also shifted authority over schools and hospitals to municipal governments to weaken public sector unions. Both *regiones* and municipalities became focal points for subsequent decentralization (Eaton 2004c).

Regime change can have different effects for regional governance in different parts of a country. Democratization in Spain produced a cascade of regional bargains, beginning with the historic regions of the Basque Country, Catalonia, and Galicia. The 1978 constitution laid out two routes to regional autonomy, but competitive mobilization spurred a variety of institutional arrangements (Agranoff and Gallarín 1997).

A democratic opening is often followed by the accommodation of a previously suppressed ethnic minority. One result is that a country that had a homogenous structure of government becomes territorially differentiated. Aceh and Papua became autonomous Indonesian regions after Suharto's resignation (Bertrand 2007; Reid 2010b). Mindanao became an autonomous Philippine region following the People Power Revolution (Bertrand 2010: 178). Democratization in Russia after 1989 saw a series of bilateral arrangements with the central government empowering ethnic provinces (*respubliki*) (Svendsen 2002: 68–70).

A valid measure of regional authority should be sensitive to these phenomena. Theory in this rapidly growing field often engages the timing and character of regional authority, and it often has implications for individual regions as well as countries. If one wishes to test a theory relating democratization to multilevel governance, it is necessary to have measures in which these phenomena do not contaminate each other.

III. *Dimensions of Self-rule and Shared Rule*

One of the most important tasks in measuring an abstract concept is to decompose it into dimensions which a) can be re-aggregated to cover the meaning of the specified concept, b) are concrete in the sense that they are a step closer to observed reality, and c) are simple in that they are unidimensional and substantively interpretable (De Leeuw 2005). This can take more than one step. Measurement of the nominal GDP of the US begins by decomposing the concept into five categories—consumption, services, investment, exports, and imports—each of which is further disaggregated. Consumption, for example, consists of rental income, profits and proprietors' income, taxes on production and imports less subsidies, interest, miscellaneous payments, and depreciation. The purpose is to break down an abstract concept, in this case nominal GDP, into pieces that capture its content and can be empirically estimated (Landefeld et al. 2008). Similarly, measures of democracy disaggregate the concept into domains that can be broken down into dimensions (Coppedge et al. 2008, 2011).

Our first move is to distinguish two domains that encompass the concept of regional authority. *Self-rule* is the authority that a subnational government exercises in its own territory. *Shared rule* is the authority that a subnational government co-exercises in the country as a whole. The domains of self-rule and shared rule provide an elegant frame for our measure and they are widely familiar in the study of federalism (Elazar 1987; Keating 1998, 2001; Lane and Errson 1999; Riker 1964). The distinction appears to have empirical as well as theoretical bite. Research using our prior measure for OECD countries finds that self-rule and shared rule have distinct effects on corruption (Neudorfer

Transparency in Measurement

and Neudorfer 2015), spatial disparities (Ezcurra and Rodriguez-Pose 2013), regional representation (Donas and Beyers 2013; Tatham and Thau 2013), regional party vote share in national elections (Kyriacou and Morral-Palacin 2015), subnational coalition formation (Bäck et al. 2013), protest (Quaranta 2013), and voting (Niedzwiecki and Stoyan 2015).¹⁹

Self-rule and shared rule are distinct domains of regional governance. But we need to decompose them into dimensions to estimate variation.

The tripartite distinction between fiscal, administrative, and political decentralization is a useful point of departure. Fiscal decentralization is control over subnational revenue generation and spending; administrative decentralization is the authority of subnational governments to set goals and implement policies; and political decentralization refers to direct elections for subnational offices (Montero and Samuels 2004; Falleti 2005).²⁰ The four types of political decentralization identified by Treisman (2007: 23–7) overlap with this three-fold schema, with the important addition of a dimension for constitutional decentralization (“subnational governments or their representative have an explicit right to participate in central policy making”).

The revenue generating side of fiscal decentralization can be broken down into the authority of a regional government to control the base and rate of major and minor taxes and its latitude to borrow on financial markets without central government approval. On administrative decentralization it would be useful to know the extent to which the central government can veto subnational government and the kinds of policies over which subnational governments exert authority. And on political decentralization, one might distinguish between indirect and direct election of offices, and further, between the election of regional assemblies and regional executives.

Fiscal, administrative, and political decentralization are concerned with the authority of a regional government in its own jurisdiction. However, a regional government may also co-determine national policies. Is the regional government represented in a national legislature (normally the second chamber), and if so, to what effect? Can the regional government co-determine the proportion of national tax revenue that goes into its pocket? Does it have routinized access to extra-legislative channels to influence the national government? And, most importantly, does the regional government have authority over the rules of the game?

¹⁹ An incipient literature examines the diverse causes of self-rule and shared rule (see e.g. Amat and Falcó-Gimeno 2014). Joan-Josep Vallbe (2014) extends the self-rule/shared rule distinction to judicial regional authority.

²⁰ Falleti (2010: 329) takes a step toward a more specific conceptualization of administrative decentralization as “the set of policies that transfer the administration and delivery of social services such as education, health, social welfare, or housing to subnational governments.”

Measurement

These distinctions provide a basis for further specification. Each responds to a basic question that one can ask about regional authority. In the domain of *self-rule* we formulate five questions:

- How independent is a regional government from central state control? *Institutional depth* tracks the extent to which a regional government can make autonomous policy decisions. A deconcentrated regional administration has the apparatus of government—a physical address, a bureaucracy, an executive, a budget—but is subordinate to the center. A decentralized regional government, by contrast, can make independent policy decisions, which, at the upper end of this scale, are not subject to central government veto.
- What is the range of a regional government's authority over policy within its jurisdiction? *Policy scope* taps the breadth of regional self-rule over policing, over its own institutional set-up, over local governments within its jurisdiction, whether a regional government has residual powers, and whether its competences extend to economic policy, cultural-educational policy, welfare policy, immigration, or citizenship.
- What authority does a regional government have over taxation within its jurisdiction? *Fiscal autonomy* is evaluated in terms of a regional government's authority to set the base and rate of minor and major taxes in its jurisdiction. This dimension is concerned with the authority of a government to set the rules for taxation rather than the level of regional spending.
- Does a regional government have authority to borrow on financial markets? *Borrowing autonomy* evaluates the centrally imposed restrictions on the capacity of a regional government to independently contract loans on domestic or international financial markets.²¹
- Is a regional government endowed with representative institutions? *Representation* assesses whether a regional government has a regionally elected legislature; whether that legislature is directly or indirectly elected; and whether the region's executive is appointed by the central government, dual (i.e. co-appointed by the central government), or autonomously elected (either by the citizens or by the regional assembly).

²¹ Our prior measure overlooked borrowing (Hooghe, Marks, and Schakel 2008, 2010). Extending the sample to Latin America and South-East Asia brings regional borrowing into focus both in self-rule and shared rule. Subnational borrowing became particularly salient from the 1980s and 1990s when several Latin American countries were hit by debt crises. The financial crisis in the Eurozone has also put the spotlight on regional borrowing.

Transparency in Measurement

In the domain of *shared rule* we pose the following questions:

- To what extent can a regional government co-determine national policy making? *Law making* assesses the role of regions in structuring representation at the national level (i.e. in a second legislative chamber); whether regions have majority or minority representation there; and the legislative scope of the second chamber.
- Can a regional government co-determine national executive policy in intergovernmental fora? *Executive control* taps whether regional governments have routine meetings with the central government and whether these are advisory or have veto power.
- Can a regional government co-determine how national tax revenues are distributed? *Fiscal control* taps the role of regions in negotiating or exerting a veto over the territorial allocation of national tax revenues.
- Can a regional government co-determine the restrictions placed on borrowing? *Borrowing control* distinguishes whether regional governments have no role, an advisory role, or a veto over the rules that permit borrowing.
- Can a regional government initiate or constrain constitutional reform? *Constitutional reform* assesses the authority of a regional government to propose, postpone, or block changes in the rules of the game. Does constitutional reform have to gain the assent of regional governments or their constituencies? Does it require majority support in a regionally dominated second chamber?

A region may exercise shared rule multilaterally with other regions or it may exercise shared rule bilaterally with the center. Multilateral shared rule is contingent on coordination with other regions in the same tier; bilateral shared rule can be exercised by a region acting alone (Chapter Three).

IV. Indicators for Dimensions of Self-rule and Shared Rule

An indicator consists of rules for inferring variation along a dimension (Tal 2013: 1162; King, Keohane, and Verba 1994: 75). Chang (2004: 216) asks, “In the process of operationalizing the abstract concept, what exactly do we aim for, and what exactly do we get? The hoped-for outcome is an agreement between the concrete image of the abstract concept and the actual operations that we adopt for an empirical engagement with the concept (including its measurement).”

Our purpose is to devise indicators that encompass the meaning of the concept and can be reliably scored. All observations, even simple ones like the number of votes received by a candidate in an election, are contestable,

but some observations are more contestable than others (Lakatos 1970). For example, an indicator that asks a coder to score “the ability of the center to suspend lower levels of government or to override their decisions” (Arzaghi and Henderson 2005) is abstract and ambiguous.²² What if there are several lower levels of government and they differ? What if the central government can suspend a lower level government only under exceptional circumstances? What if some lower level decisions may be overridden and others not?

Tables 1.2 and 1.3 detail indicators for self-rule and shared rule.²³ The indicators specify institutional outcomes for an individual region or regional tier that can be reliably assessed against information in constitutions, laws, executive orders, government documents. In addition, the intervals are designed to have the following desiderata (Gerring and Skaaning 2013; Goertz 2006):

- Each interval is comprised of a set of necessary and sufficient institutional conditions for a particular score.
- The attributes for each interval encompass the prior interval with some additional unique attribute.
- The attributes are binary in order to minimize the gray zone between existence and non-existence.
- Collectively, the intervals seek to capture the relevant variation in the population that is assessed.
- The spacing of the intervals is conceived as equidistant so that a unit shift along any dimension is equivalent.

V. Scoring Cases

Scoring cases consists of obtaining and processing information in order to place numerical values on objects (Bollen and Paxton 2000). Our scoring strategy involves “interpretation through dialogue.”

Interpretation is the act of explaining meaning among contexts or persons. When measuring regional authority we are interpreting the concept of regional authority in the context of particular regions at particular points in time. As one moves down the ladder of measurement in Figure 1.1, the

²² “This dimension measures whether or not the central government has the legal right to override the decisions and policies of lower levels of government. If the central government has such a right, the country scores zero; if not, the score is four. To ‘override’ in this context means to be able to veto without due process. Many countries have legal mechanisms for the appeal and review by higher authorities of lower-level government decisions. As a rule, these do not constitute override authority, unless they are extremely lax. Instead, override authority exists when the central government can legally deny regional and local authority with an ease that calls that very authority in to question.” <<http://www.econ.brown.edu/faculty/henderson/decentralization.pdf>>.

²³ Law making consists of four sub-dimensions.

Table 1.2. Self-rule

Self-rule	The authority exercised by a regional government over those who live in the region
Institutional depth	<p>The extent to which a regional government is autonomous rather than deconcentrated.</p> <p>0–3</p> <p>0 No functioning general purpose administration at regional level. 1 Deconcentrated, general purpose, administration. 2 Non-deconcentrated, general purpose, administration subject to central government veto. 3 Non-deconcentrated, general purpose, administration not subject to central government veto.</p>
Policy scope	<p>The range of policies for which a regional government is responsible.</p> <p>0–4</p> <p>0 Very weak authoritative competencies in a), b), c), or d) whereby a) economic policy; b) cultural–educational policy; c) welfare policy; d) one of the following: residual powers, police, own institutional set-up, local government. 1 Authoritative competencies in one of a), b), c), or d). 2 Authoritative competencies in at least two of a), b), c), or d). 3 Authoritative competencies in d) and at least two of a), b), or c). 4 Criteria for 3 plus authority over immigration, citizenship, right of domicile.</p>
Fiscal autonomy	<p>The extent to which a regional government can independently tax its population.</p> <p>0–4</p> <p>0 Central government sets the base and rate of all regional taxes. 1 Regional government sets the rate of minor taxes. 2 Regional government sets the base and rate of minor taxes. 3 Regional government sets the rate of at least one major tax: personal income, corporate, value added, or sales tax. 4 Regional government sets base and rate of at least one major tax.</p>
Borrowing autonomy	<p>The extent to which a regional government can borrow.</p> <p>0–3</p> <p>0 The regional government does not borrow (e.g. centrally imposed rules prohibit borrowing). 1 The regional government may borrow under prior authorization (<i>ex ante</i>) by the central government and with one or more of the following centrally imposed restrictions a. golden rule (e.g. no borrowing to cover current account deficits) b. no foreign borrowing or borrowing from the central bank c. no borrowing above a ceiling d. borrowing is limited to specific purposes. 2 The regional government may borrow without prior authorization and under one or more of a), b), c), or d). 3 The regional government may borrow without centrally imposed restrictions.</p>
Representation	<p>The extent to which a region has an independent legislature and executive.</p> <p>0–4</p> <p>Assembly: 0 No regional assembly. 1 Indirectly elected regional assembly. 2 Directly elected assembly. Executive: 0 Regional executive appointed by central government. 1 Dual executive appointed by central government and regional assembly. 2 Regional executive is appointed by a regional assembly or directly elected.</p>

Table 1.3. Shared rule

Shared rule		The authority exercised by a regional government or its representatives in the country as a whole	
Law making	The extent to which regional representatives co-determine national legislation.	0–2	0.5
		0.5	0.5
Executive control	The extent to which a regional government co-determines national policy in intergovernmental meetings.	0–2	0
		1	1
Fiscal control	The extent to which regional representatives co-determine the distribution of national tax revenues.	0–2	0
		1	1
Borrowing control	The extent to which a regional government co-determines subnational and national borrowing constraints.	0–2	0
		1	1
Constitutional reform	The extent to which regional representatives co-determine constitutional change.	0–4	0
		1	1
		2	2
		3	3
		4	4

Regions are the unit of representation in a national legislature. Regional governments designate representatives in a national legislature. Regions have majority representation in a national legislature based on regional representation.

The legislature based on regional representation has extensive legislative authority.

No routine meetings between central and regional governments to negotiate policy. Routine meetings between central and regional governments *without* legally binding authority.

Routine meetings between central and regional governments *with* legally binding authority.

Neither the regional governments nor their representatives in a national legislature are consulted over the distribution of national tax revenues.

Regional governments or their representatives in a national legislature negotiate over the distribution of tax revenues, but do not have a veto.

Regional governments or their representatives in a national legislature have a veto over the distribution of tax revenues.

Regional governments are not routinely consulted over borrowing constraints.

Regional governments negotiate routinely over borrowing constraints but do not have a veto.

Regional governments negotiate routinely over borrowing constraints and have a veto.

The central government or national electorate can unilaterally reform the constitution. A national legislature based on regional representation can propose or postpone constitutional reform, raise the decision hurdle in the other chamber, require a second vote in the other chamber, or require a popular referendum.

Regional governments or their representatives in a national legislature propose or postpone constitutional reform, raise the decision hurdle in the other chamber, require a second vote in the other chamber, or require a popular referendum.

A legislature based on regional representation can veto constitutional change; or constitutional change requires a referendum based on the principle of equal regional representation.

Regional governments or their representatives in a national legislature can veto constitutional change.

Transparency in Measurement

concept of regional authority becomes less abstract, but even concrete concepts, such as a dual executive, a routine meeting, or a formal veto, are not directly observable. “The bridge we build through acts of measurement between concepts and observations may be longer or shorter, more or less solid. Yet a bridge it remains” (Schedler 2012: 22). Our intent is to make the link between indicators and scores both plausible and transparent.

Dialogue—sustained, open-ended discussion—is intended to increase the validity of our judgments. While time intensive, dialogue among coders is vital for consistent interpretations across countries. Bowman, Lehoucq, and Mahoney (2005: 957) describe the process which underpins their democracy index as iterative consensus building: “Disagreements arose regarding the codes for several particular measures, and these differences generally reflected either a limitation in the measure or a limitation in an author’s knowledge of the facts. If the problem was with the resolving power of a measure, we sought to better define the measure until a consensus could be reached. If the problem arose not because of the measure but rather because of divergent understandings of the empirical facts, we reviewed all evidence and argued about the facts.” Our approach is similar (see also Saylor 2013).

Dialogue among coders makes it impossible to assess inter-coder reliability, but this is a sacrifice worth making. The principal challenge in estimating an abstract concept such as regional authority is validity rather than reliability. Validity concerns whether a score measures what it is intended to measure. Do the dimensions really capture the meaning of the concept? Do the indicators meaningfully pick up the variation on each dimension? Do the scores accurately translate the characteristics of individual cases into numbers that express the underlying concept? Reliability concerns the random error that arises in any measurement. How consistent are scores across repeated measurements? Would a second, third, or *n*th expert produce the same scores? If the error one is most worried about is systematic rather than random, then it may be more effective to structure dialogue among coders to reach consensus on a score than to combine the scores of independent coders.

Using expert evaluations is inappropriate for the data we seek. Expert surveys are useful for topics that are “in the head” of respondents. The information required to assess the authority of individual regions in a country on ten dimensions annually from 1950 goes far beyond this. It is not a matter of providing proper instructions to experts. The limitations of expert surveys are more fundamental (Steenbergen and Marks 2007; Marks et al. 2007).²⁴

²⁴ Expert surveys are an economical and flexible research tool when the information necessary for valid scoring is directly accessible to the experts (Wiesehomeier and Benoit 2009). The number of experts need not be large—a rule of thumb would be six or more for each observation (Steenbergen and Marks 2007; Marks et al. 2007). Expert surveys eliminate the need to have specific sources of information (e.g. laws, government documents) available for all cases. And

Measurement

An evaluation based on a series of expert surveys over fifteen years concludes that “Unambiguous question wording is necessary but not sufficient for reliable expert judgments. Perhaps the most important source of error lies neither in poor question-wording, nor in the selection of experts, but in asking questions that lie beyond the expertise of respondents” (Hooghe et al. 2010: 692). This limitation, along with our overriding concern with validity, suggests that dialogue among researchers is both more feasible and more appropriate than an expert survey for the task at hand.

The practical steps involved in interpretation through dialogue are as follows:

- *Gathering and interpreting public documents.* An initial step is to collect publicly available information related to the indicators. These are first and foremost constitutions, laws, executive decrees, budgets, government reports, and websites.²⁵ This is usually not so difficult for the most recent one or two decades, but can be challenging for the 1950s and 1960s.
- *Engaging the secondary literature.* Numerous books, articles, and non-governmental studies cover the larger and richer countries. The coverage of Latin America, Eastern Europe, and Asia has increased markedly in recent years. However, secondary sources thin as one goes back in time.²⁶ In most cases, the secondary literature is less useful as a source of “facts” than it is as a conceptual/theoretical basis for probing our measurement decisions, including particularly the contextual appropriateness of the indicators.
- *Subjecting interpretations to expert commentary.* Although it is unreasonable to expect country experts to provide strictly comparable scores for individual regions across ten dimensions on an annual basis going back to 1950, they can provide valuable feedback on the validity of scoring judgments. For countries that we regard as the most complex or least sourced, we commissioned researchers who have published extensively on

expert surveys are flexible tools for experiments designed to evaluate and improve the reliability of the measure. It is possible to introduce vignettes into the survey that tell us how individual experts evaluate benchmark scenarios (Bakker et al. 2014). However, the virtues of expert surveys are null if experts are asked to evaluate topics to which they do not have direct cognitive access. In the Chapel Hill Expert Survey, we have found that items tapping expert judgments on the contemporary positioning of political parties on major issues produce reliable scores, while items that ask experts for more specific information on the extent of division within political parties on those same issues fail to do so. The information that we seek on regional authority is much more specific than that required for evaluating divisions within political parties.

²⁵ Wikipedia lists territorial subdivisions for most countries, and <<http://www.statoids.com>>, a website run by Gwillim Law, a Chapel Hillian, is a fount of information.

²⁶ Country reports from the OECD’s multilevel governance unit are valuable sources. Also useful are studies commissioned by the Inter-American Development Bank, the Asian Development Bank, and the World Bank.

Transparency in Measurement

regional authority in their country to write commentaries on our interpretations and scores. These commentaries led us back to the primary sources, and in some cases to revisit our conceptualization of the indicators and the dimensions.

- *Discussing contending interpretations in extended dialogue.* All scoring decisions were discussed by three or more members of the research team, often at length. Difficult cases were usually discussed on more than two occasions. Divergence of interpretation led us to soak and poke by going back to the sources or finding additional sources. It was also instrumental in refining the indicators, and led us to distinguish between bilateral and multilateral shared rule.²⁷ Interpretation through dialogue made it possible to revisit our decisions on indicators and dimensions as we sought to place institutional alternatives in diverse countries on a single theoretical–conceptual frame.
- *Paying sustained attention to ambiguous and gray cases.* No matter how well designed a measure, there will always be ambiguities in applying rules to particular cases. There will also be gray cases that lie between the intervals. Our approach is to clarify the basis of judgment and, where necessary, devise additional rules for adjudicating such cases that are consistent with the conceptual underpinnings of the measure. Chapter Three sets out our rules for coding ambiguous and gray cases and is, not coincidentally, the longest chapter in this book.
- *Explicating judgments in extended profiles.* The lynchpin of our measure is the endeavor to explain coding decisions. This involves disciplined comparison across time and space. The country profiles in this volume make our scoring evaluations explicit so that researchers familiar with individual cases may revise or reject our decisions. At the same time, the profiles are intended to remove the curtain that protects the cells in a dataset from cross-examination.

VI. Adjudicating Scores

Gray cases are endemic in measurement. They come into play at every step in a measure and arise in the fundamental tension, noted by Weber, between an idea and an empirical phenomenon. Gray cases are not indicators of scientific failure. Rather they are calls for re-assessing a measurement, for ascending the arrows on the right side of Figure 1.1. One can seek to resolve a gray case by refining observation, by revising an indicator, dimension or, *in*

²⁷ See the appendix for the coding schema for multilateral and bilateral shared rule.

Measurement

extremis, by redefining the specified concept. Is this case gray because we lack good information or does it raise conceptual issues? Is the case an isolated instance of ambiguity or does it suggest a more general problem? If the latter, can one rejig the indicator for that dimension? Or does the problem go back to the specification of the concept?

Gray cases contain valuable information for users and for those who might wish to improve a measure. They flag areas for improving a measure. We notate three common sources of “grayness” in the extensive country profiles in Part II.

- *Insufficient or ambiguous information.* Outside the laboratory, observation can be plagued by poor light or deficient information. We indicate scores for which we have thin information by using the symbol α in superscript in the profile.
- *Observations that fall in-between intervals.* No matter how sharp a distinction, some observations sit between intervals. We indicate these borderline cases with the symbol β in superscript.
- *Disagreement among sources, coders, experts.* Applying a concept to an empirical phenomenon is an inferential process that is subject to error and hence to disagreement. Even simple concepts that refer to physical objects have fuzzy boundaries (Quine 1960: 114ff). We note disagreements among sources, coders, and/or experts with the superscript γ .

Conclusion

Measuring the authority of individual regions in a wide range of countries over several decades is always going to be a theoretical as well as practical challenge. Our approach, in short, is to a) disaggregate the concept into coherent dimensions that encompass its meaning; b) operationalize these dimensions as institutional alternatives that are abstract enough to travel across cases but specific enough to be reliably evaluated; c) assess the widest possible range of documentary information in the light of the secondary literature and expert feedback; and d) discuss coding decisions and ambiguities in comprehensive country profiles.

The measure can be used to estimate regional authority at the level of the individual region, regional tier, or country by combining the dimensions. Alternatively, researchers may wish to re-aggregate these to their needs. The intervals on the dimensions are conceptualized along equal increments, so one can sum dimension scores to produce a scale ranging between 1 and 30 for each region or regional tier. Country scores are zero for countries that have no regional government, but there is no a priori maximum because countries may

Transparency in Measurement

Table 1.4. Polychoric factor analysis

Components	Single-factor solution	Two-factor solution:	
		Self-rule	Shared rule
Institutional depth	.86	.87	.08
Policy scope	.91	.88	.13
Fiscal autonomy	.84	.59	.34
Borrowing autonomy	.85	.86	.08
Representation	.81	.99	–.12
Law making	.74	.08	.76
Executive control	.82	.12	.80
Fiscal control	.75	.04	.81
Borrowing control	.62	–.08	.77
Constitutional reform	.78	.05	.83
Eigenvalue	6.43	5.51	5.29
Chi-squared	859.38	859.38	
Explained variance (%)	81.9		
Factor correlation		0.61	

Note: Principal components factor analysis, promax non-orthogonal rotation, listwise deletion. $n = 80$ (country scores in 2010). For the two-factor solution, the highest score for each dimension is in bold.

have more than one tier. Chapter Three explains how we aggregate regional scores into country scores. We use this additive scale in the maps, tables, and figures in this book.

An alternative approach is to interpret the dimensions as indicators of a latent variable. The Cronbach's alpha across the ten dimensions for 2010 is 0.94, which suggests that the dimensions can be interpreted as indicators of a single latent concept. Table 1.4 presents a factor analysis for country scores in 2010. We use polychoric correlations on the conservative assumption that the indicators are ordinal. A single-factor solution accounts for 82 percent of the variance. When we impose a two-factor solution, each indicator loads strongly on one latent factor and weakly on the other factor. The solution confirms the theoretical distinction between self-rule and shared rule.²⁸

It does not make much difference which method one uses to aggregate the data. The scores derived from factor analysis and from additive scaling are very similar. The correlation is 0.98 for 2010 for the single dimension. Figure 1.2 plots correlations using interval data and shows that the index is robust across alternative weights for self-rule and shared rule. The RAI weighs shared rule to self-rule in the ratio of 2:3. When we reverse these weights, the rank order among countries in 2010 yields a Spearman's *rho* of 0.99 (Pearson's $r=0.97$).

The decision to estimate authority at the level of individual regions rather than countries is the single most important decision in this book because it affects how one thinks about the structure of governance. Governance

²⁸ The correlation between the two dimensions is reasonably strong ($r=0.61$).

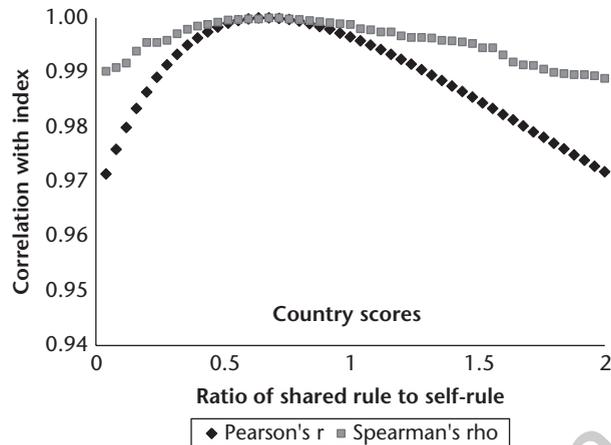


Figure 1.2. Robustness of the regional authority index across alternative weights for self-rule and shared rule

Note: Calculations are for 2010; $n = 80$. Spearman's *rho* is calculated on ordinal scores, and Pearson's *r* is calculated on the interval scores. The RAI weights shared rule to self-rule in the ratio of 2:3 (0.66). Here we vary the ratio between 0 and 2.

exhibits great variation within as well as among countries, and one cannot begin to fathom the reasons for this or understand its consequences if one conceives the state as the unit of analysis. Some regional governments have wide ranging policy competences; others deal with a single problem. Some can block constitutional reform. Some have extensive taxing powers. Some exert wide ranging authority within their own territories; others play a decisive role in the governance of the country as a whole. Some regions have a special bilateral relationship with the central government, while others exist alongside other regions in uniform tiers. The variation that the RAI detects among countries is extremely wide, and now one can also systematically probe variation within countries over time.

Finally, the effort to measure a concept as complex as regional authority may have implications for measurement in general. Measurement seeks to establish a numerical relation between an observable phenomenon and a concept. This, as Max Weber emphasized, involves interpretation. What, precisely, is being measured? How is the concept specified? What are its dimensions? How are intervals along these dimensions operationalized? How are individual cases scored on those dimensions? What rules apply to gray cases? These are questions that confront social science measurement generally. Each question involves judgment, the weighing of one course of action against others. Our goal in this book is to make those judgments explicit, and hence open to disconfirmation or improvement.

2

Crossvalidating the Regional Authority Index

Our aim in this chapter is to assess the validity of the regional authority index (RAI) by comparing it to prior institutional and fiscal measures. We begin by asking whether alternative institutional measures give similar scores to the same cases. This is convergent validation, the extent to which measures of the same concept are positively associated with each other (Bollen 1989: 188; Ray 2007: 12). To assess convergent validity we evaluate the extent to which these measures are in agreement with the RAI, explore sources of disagreement in a regression analysis, and complement this with an in-depth look at particular cases.

Convergence provides confidence in the validity of our measurement whereas disagreement provides a basis for further investigation. Each measure suffers from error, and the sources of error may vary in non-random ways. We find that differences among decentralization measures have systemic causes, both with regard to the extent of difference and the direction of difference. The most important differences arise because some countries have more than one tier of regional government between the local and the national and because measures seek to estimate decentralization over a period in which there has been extensive change.

Beyond such systematic differences, institutional measures sometimes arrive at sharply contrasting scores for individual countries, and the reasons for this are worth investigating in some detail. Knowing when, where, and how error in measurement arises helps one decide whether to use one measure over another (Adcock and Collier 2001; Bollen 1989; King, Keohane, and Verba 1994; Marks et al. 2007).

We conclude by discussing the content validity of three types of fiscal indicators and comparing their scores to the RAI. Content validity “assesses the degree to which an indicator represents the universe of content entailed in the systematized concept being measured” (Adcock and Collier 2001: 537).¹

¹ Adcock and Collier (2001: 537) also identify a third type—criterion validity, which assesses “whether the scores produced by an indicator are empirically associated with scores for other

Here the task is to clarify the domain of the concept and to judge whether the measures fully represent the intended domain (Bollen 1989: 185). Are crucial elements omitted or are inappropriate elements included? Fiscal measures have the virtue of reliability, but we suggest that they do, indeed, omit important dimensions of decentralization and are correspondingly limited as a measure of decentralization.

Institutional Indicators of Decentralization

There is no shortage of measures of decentralization with which the RAI can be compared. Table 2.1 overviews the five most commonly used measures that, like the RAI, focus on the authoritative competences of subnational governments. All five measures conceive decentralization as a latent variable with fiscal, political, and administrative indicators (Falleti 2010; Schneider 2003). Each covers an array of countries on multiple dimensions of decentralization that can be summarized at the level of the country as a whole.

The chief differences between these measures and the RAI are as follows (Table 2.2):²

- *Unit of measurement.* The RAI is distinct in conceiving the individual region and the regional tier, rather than the country, as units of analysis. This increases the number of observations and makes it possible to compare regions and regional tiers within, as well as across, countries. We use aggregate RAI country scores for the purpose of comparison, but it is worth keeping in mind that country scores are just a useful fiction. The actual units of subnational authority in all decentralization measures are individual general purpose governments within territorially circumscribed jurisdictions.
- *Time period.* The RAI provides annual observations for 1950–2010. Brancati (2006, 2008) provides annual observations for 1985–2000. Arzaghi and Henderson (2005) assess eight five-year intervals between 1960 and 1995. Treisman (2002) is a cross-sectional measure,³ and Lijphart (1999) and Woldendorp, Keman, and Budge (2000) average decentralization over several decades on the assumption that decentralization is fairly stable over time (Inman 2008).

variables, called criterion variables, which are considered direct measures of the phenomenon of concern.” We do not assess criterion validity because there is no generally accepted “criterion variable” or “gold standard” for measuring regional authority.

² For descriptive statistics see Tables 2.A.1 and 2.A.2 in the appendix.

³ Treisman (2002) is available as an unpublished paper on his website.

Crossvalidating the Regional Authority Index

Table 2.1. Institutional measures of decentralization

Measure	Dimensions	Indicators
Arzaghi and Henderson (2005)	<p>Index of institutional decentralization, or effective federalism, consisting of:</p> <ul style="list-style-type: none"> • formal government structure • political responsibilities of subnational governments • fiscal responsibilities of subnational governments 	<p>Effective federalism (0–4) is the average of:</p> <ul style="list-style-type: none"> – constitutional federal versus unitary structure (0 or 4) – election of a regional executive (0 or 4) – election of a local executive (0 or 4) – ability of the center to suspend lower levels of government or to override their decisions (0 or 4) – revenue sharing (0, 2, or 4)
Brancati (2008)	<p>Level of political decentralization:</p> <ul style="list-style-type: none"> • elective dimension • policy dimension 	<p>Political decentralization (0–5) is the sum of five dichotomous indicators:</p> <ul style="list-style-type: none"> – democratically elected regional legislatures – regional legislatures can raise or levy their own taxes – regional legislatures have joint or exclusive control over education – regional legislatures have joint or exclusive control over public order or police – regions must approve constitutional amendments
Lijphart (1999)	<p>Federalism whereby countries are categorized on the basis of:</p> <ul style="list-style-type: none"> • formal character of government structure (federal or unitary) • extent of decentralization (range of powers assigned to the regional level) 	<p>Federalism (1–5) is an ordinal scale:</p> <ul style="list-style-type: none"> – unitary and centralized (=1) – unitary and decentralized (=2) – semi-federal (=3) – federal and centralized (=4) – federal and decentralized (=5)
Treisman (2002)	<p>Decision making decentralization defined as formal rules about the distribution of political authority over decision making</p>	<p>Decision making decentralization (0–3) is an additive scale:</p> <ul style="list-style-type: none"> – autonomy = the constitution reserves to subnational legislatures the exclusive right to legislate in at least one policy area – residual authority = the constitution assigns to subnational legislatures the exclusive right to legislate on issues that are not specifically assigned to one level of government; – subnational veto = a regionally elected upper chamber exists with the constitutional right to block legislation

Measurement

Woldendorp, Keman, and Budge (2000)	Autonomy index consists of:	Autonomy index (0–8) is an additive scale:
	• fiscal centralization	– 2 if a country has a degree of fiscal centralization lower than 75%; – 1 if a country has fiscal centralization between 75% and 90%; – 0 if a country has fiscal centralization equal to or more than 90%
	• regional autonomy	– 2 if regional autonomy is formally laid down (federal states); – 1 if the country is a semi-federalist system; – 0 if neither
	• local government autonomy	– 2 if local government is mentioned in the constitution, its autonomy is recognized, and it is guaranteed direct representation; – 1 if one or two of these conditions are met; – 0 in all other cases
	• centralization	– 2 if the state is not centralized; – 1 if the state is medium centralized; – 0 if the state is highly centralized

Note: The operationalization of fiscal centralization diverges somewhat from the one published in Woldendorp, Keman, and Budge (2000). The adjustments were made after communication with Hans Keman and Jaap Woldendorp.

- *Tiers*. The RAI estimates the authority of subnational governments at each level between the local (>150,000 population) and the national. Arzaghi and Henderson and Woldendorp et al. are chiefly concerned with regional government, but have some items that encompass local government.⁴ The remaining measures do not discriminate levels of government.
- *Dimensions*. All measures conceive decentralization as multidimensional. Arzaghi and Henderson, Brancati, Treisman, and the RAI estimate regional assemblies. Brancati, Woldendorp, and the RAI estimate regional tax authority. Treisman and the RAI evaluate whether residual powers rest with the region or the central state. In addition, the RAI estimates shared rule, the authority co-exercised by a region and regional tier within the country on five dimensions for law making, executive control, fiscal decision making, borrowing, and constitutional reform (Table 1.3).

⁴ For this reason, in the following analyses we exclude the “election of a local executive” dimension from the Arzaghi and Henderson measure (see Table 2.1). We thank Christine Kearney for providing us with disaggregated scores. We are unable to exclude scores for local government in the Woldendorp et al. measure because disaggregated estimates are not available.

Table 2.2. Comparing decentralization measures

	Coverage			Concept specification			Concept disaggregation		Unit of measurement
	number of countries	time period	time points	multiple regional tiers	local/regional separated	local tier included	number of dimensions	number of intervals	
Regional Authority Index	81	1950–2010	61	Yes	Yes	No	10	42	Region
Arzaghi/Henderson	48	1960–1995	8	No	Partial	Yes	6	14	Country
Brancati^a	37	1985–2000	16	No	No	No	5	5	Country
Lijphart	36	1945–1996	1	No	No	Yes	2	5	Country
Treisman^b	166	1990–1999	1	No	No	Yes	3	3	Country
Woldendorp et al.	51	1945–1998	1	No	Partial	Yes	4	8	Country

^a The Brancati measure has twenty-three countries in common with the RAJ, but the analysis reported here uses fifty-eight countries; the twenty-three countries coded by Brancati and an additional thirty-five coded by the authors on the basis of information provided by Dawn Brancati in personal communication. The smaller-N analysis is available from the authors upon request.

^b It should be noted that Treisman is reluctant to aggregate the three indicators of decision making decentralization into a single index (see e.g. 2002: 9–10).

Measurement

- *Country coverage.* Each measure covers the larger Western democracies (Table 2.A.1). Treisman covers virtually every non-micro state. The RAI covers all members of the Organisation of Economic Cooperation and Development (OECD), all Latin American countries, ten countries in Europe beyond the European Union (EU), and eleven in the Pacific and South-East Asia. Woldendorp et al. cover fifty-one democracies. Arzaghi and Henderson cover forty-eight countries with a population over ten million. Brancati covers thirty-seven countries with regional ethnic groups and Lijphart covers thirty-six democracies.
- *Intervals.* All measures go beyond the classic federal/unitary dichotomy. However, the number of intervals varies from three (Treisman) and five (Brancati; Lijphart) to forty-two (RAI) (see Table 2.2). The more fine-grained a measure, the better equipped it is to differentiate levels of decentralization among federal and among unitary countries. Lijphart's measure compresses nearly all federal countries at the high end of the scale with a score of five. Treisman's measure separates federal countries from each other but compresses most unitary non-federal countries in the lowest category.

Cross-sectional Comparison

To what extent do the measures tap a common dimension? Our first step is to conduct principal factor analyses on a cross-sectional dataset containing average country scores over time produced by each measure.⁵ Since the country overlap varies across the measures, we conduct four factor analyses in the columns labeled "Country scores" in Table 2.3. We then use all the available data and conduct principal factor analyses for the same measures with annual observations for each country. These are the results displayed under "Country/year scores" in Table 2.3.⁶

The results reveal a high degree of convergence. In no comparison does the eigenvalue of the principal axis fall below 2.4, and the common variance is around 80 percent across the board. Every decentralization index loads heavily on the principal axis with factor loadings in excess of 0.74 for both cross-sectional and panel datasets. Notwithstanding the differences among the measures noted above, they appear to tap a common latent variable.

⁵ Table 2.A.3 reports Pearson correlations.

⁶ Countries without a region or regional tier are excluded in all analyses.

Crossvalidating the Regional Authority Index

Table 2.3. Factor analysis of decentralization measures

	Country scores				Country/year scores			
	I	II	III	IV	I	II	III	IV
Regional Authority Index	0.96	0.93	0.93	0.92	0.93	0.89	0.92	0.92
Arzaghi/Henderson	0.85	–	–	–	0.91	–	–	–
Brancati	0.90	0.92	0.92	0.89	0.93	0.93	0.92	0.89
Lijphart	0.92	0.92	–	–	0.88	0.91	–	–
Treisman	0.74	0.88	0.91	0.91	0.80	0.90	0.92	0.92
Woldendorp et al.	0.96	0.87	0.87	–	0.94	0.85	0.86	–
N	10	21	31	58	70	148	265	558
Eigenvalue	4.76	4.09	3.29	2.47	4.83	4.01	3.26	2.47
Explained Variance (%)	79	82	82	82	81	80	81	82

Sources of Disagreement

High factor scores can hide significant differences in scoring that may have systematic sources (Marks et al. 2007). In this section, we consider several possible sources of disagreement among decentralization measures:

Limited Country Coverage

One might expect less researched countries to generate more disagreement than the “normal suspects,” which in this field are the larger Western democracies. All six measures encompass a set of ten democracies in North America and Western Europe, but coverage declines as one moves to Eastern Europe, Southern Europe, South America, and Central America and the Caribbean. Table 2.A.1 in the appendix lists the countries covered by each measure. *Limited coverage* is the total number of times a country is excluded by the five alternative measures on the expectation that this will be positively associated with disagreement in scoring.

Distance in Time

Measurement error is likely to increase with retrospective evaluation. This is a particular problem for the RAI, which scores regions going back to 1950. The remaining measures have shorter time periods or provide single scores for multiple decades. We expect disagreement with the RAI to be higher for earlier than for later time periods both because the availability of information declines as one goes back in time and because time invariant measures may be biased toward recent years. The variable *Distance in time* is 2010 minus the year in which a country score is assessed.

Multiple Regional Tiers

The existence of multiple regional tiers in a country can produce different scores for measures that summarize all tiers or just the most authoritative tier. The RAI aggregates scores for all tiers between the local and the national, whereas the remaining measures do not explicitly distinguish different levels of subnational governance. *Tiers* is the number of regional tiers in a country. When a tier covers only part of a country, we weight each tier by the proportion of a country's population it encompasses.

Differentiation

Regions that have special authoritative competences that differentiate them from other regions in a country may give rise to scoring differences. Whereas the RAI estimates such regions individually and then aggregates regional scores to the country level using population weights, the remaining measures are national in focus. Differentiated governance is quite common: in 2010, thirty-five countries of the sixty-two countries included here had asymmetric, autonomous, or dependent regions (Hooghe and Marks forthcoming). *Differentiation* is calculated as the difference between the maximum and minimum RAI for units within the most authoritative regional tier.

Reform

The creation or abolition of regional tiers and reform in the authority of established regions may lead to scoring differences between measures that average decentralization over multiple years and those that have annual estimates. Our expectation is that disagreement will be greatest for countries where contemporary estimates provide weak guidance in estimating prior levels of decentralization.⁷ The RAI detects jurisdictional reform in sixty countries that have one or more regional tiers, but the extent of reform varies by a factor of twenty. The variable *Reform* is calculated as the cumulative absolute change in the RAI country score going back in time, so that values are maintained or increase as one moves back from the present.

Analysis of Disagreement

To what extent do these potential biases explain disagreement between the RAI and prior measures of decentralization? Our strategy is to extract residuals

⁷ The logic is that retrospective judgments may be more unreliable (Steenbergen and Marks 2007). We calculate this variable for each measure.

Crossvalidating the Regional Authority Index

Table 2.4. Explaining absolute disagreement

Source of disagreement	Arzaghi and Henderson	Brancati	Lijphart	Treisman	Woldendorp et al.
Limited coverage	0.022 (0.016)	0.015 (0.006)	0.002 (0.012)	0.049** (0.009)	0.010 (0.010)
Distance in time	0.009** (0.002)	0.015** (0.002)	0.005** (0.002)	0.015** (0.002)	0.007** (0.001)
Tiers	0.147** (0.036)	0.145** (0.030)	0.236** (0.038)	0.121** (0.026)	0.191** (0.031)
Differentiation	0.002 (0.003)	-0.008** (0.003)	0.004 (0.003)	0.001 (0.003)	0.004 (0.003)
Reform	-0.021** (0.004)	0.023** (0.006)	-0.005 (0.007)	0.041** (0.006)	-0.016* (0.006)
Rho	0.875	0.928	0.965	0.899	0.968
R ²	0.16	0.34	0.13	0.58	0.16
Wald chi ²	237	391	133	742	168
N years	1030	847	1178	606	1137
N countries	29	58	27	63	32

Note: * $p < 0.05$, ** $p < 0.01$ (two-tailed).

OLS regressions with panel corrected standard errors in brackets. The constant is dropped. The dependent variables are the *absolute* standardized residuals resulting from an OLS regression of the RAI on one of the decentralization indices.

for disagreement between the RAI and each of the five decentralization measures by regressing the RAI on the country/year scores generated by each measure.⁸ We then regress the standardized residuals onto the variables discussed above using ordinary least squares regression with panel corrected standard errors.⁹

Error comes in two forms (Marks et al. 2007). Absolute residuals capture the sheer distance between scores. This gives us a sense of how far a measure strays from other measures regardless of the direction of the difference. Raw residuals come with signs that tell one the direction of difference between scores, i.e. whether a score is in the direction of more or less decentralization. Table 2.4 presents models explaining absolute residuals and Table 2.5 does the same for raw residuals.

Distance in time, *Tiers*, and *Reform* are consistently positive causes of difference between the RAI and prior measures. The further back in time one estimates decentralization, the greater the number of levels of regional governance, and the greater the extent of jurisdictional reform over time, the larger the discrepancy between the RAI and the alternative measures. *Tiers* has the most marked effect. The absolute difference in scoring between Lijphart

⁸ Table 2.A.4 reports Pearson correlations between the residuals.

⁹ Table 2.A.5 displays descriptive statistics for the independent variables. See Achen (2000); Plümper et al. (2005); and Beck and Katz (2011), for a discussion of the conditions under which panel corrected standard errors without a lagged dependent variable and fixed effects are appropriate.

Measurement

Table 2.5. Explaining directional disagreement

Source of disagreement	Arzaghi and Henderson	Brancati	Lijphart	Treisman	Woldendorp et al.
Limited coverage	0.066** (0.020)	-0.018** (0.006)	-0.005 (0.009)	-0.051** (0.011)	-0.003 (0.007)
Distance in time	-0.013** (0.003)	-0.012** (0.002)	-0.001 (0.001)	-0.015** (0.003)	0.000 (0.001)
Tiers	0.162** (0.048)	0.281** (0.033)	0.250** (0.027)	0.333** (0.036)	0.202** (0.022)
Differentiation	0.025** (0.004)	-0.000 (0.003)	0.002 (0.002)	0.005 (0.003)	-0.001 (0.002)
Reform	-0.006 (0.006)	-0.049** (0.008)	-0.066** (0.004)	-0.028** (0.010)	-0.069** (0.004)
Rho	0.907	0.940	0.966	0.929	0.971
R ²	0.08	0.34	0.50	0.206	0.56
Wald chi ²	68	312	550	173	803
N years	1030	847	1178	606	1137
N countries	29	58	27	63	32

Note: * $p < 0.05$, ** $p < 0.01$ (two-tailed).

OLS regressions with panel corrected standard errors in brackets. The constant is dropped. The dependent variables are the *raw* standardized residuals resulting from an OLS regression of the RAI on one of the decentralization indices. A positive sign indicates that the estimate of the RAI is higher than the estimate of the alternative measure.

and the RAI for a country, such as Finland, Italy, or Portugal, which have two tiers of regional governance instead of one, would on average be just under a quarter (23.6 percent) of a standard deviation in the RAI score, which is equivalent to a score difference of 2.2. Estimates for the effect of *Distance in time* and *Reform* are greatest for the RAI and Treisman. A country/year scored twenty years in the past would, on average, generate a difference in scoring of around one-third of a standard deviation—or around 3.2 on the RAI scale.

An examination of the results for directional disagreement provides some meat on these bones. The most notable result is that the RAI detects more decentralization than alternative measures in the presence of multiple levels of regional government. This is precisely what one would expect given that the RAI estimates each level prior to aggregating them to the country level, whereas the other measures do not distinguish multiple levels in estimating decentralization. The substantive effect is quite marked. The RAI assesses between 16 percent and a third of a standard deviation more decentralization than the remaining measures in a country that has a second tier of regional government compared to just one.

Reform also has the anticipated effect. In general, jurisdictional reform has increased the level of decentralization over the past several decades. The RAI estimates lower levels of decentralization in past years where jurisdictional reform—and hence the increase in decentralization—has been large. The difference is not significant compared to the Azarghi and Henderson measure, which

Crossvalidating the Regional Authority Index

picks up change at five-year instead of annual intervals. However, it reduces the assessment of decentralization by around two-thirds of a standard deviation compared to Lijphart and Woldendorp et al. for the period prior to decentralization reforms in, say, Greece, which empowered its deconcentrated *nomoi* to self-governing units in the 1990s, when a second intermediate tier of *periphereies* was introduced. These reforms increased Greece's country score by ten points.

Cases of Disagreement

Outlying cases can be particularly revealing. So let us take a closer look at cases where the residual is more than two standard deviations above or below the estimate for five or more consecutive years. Table 2.6 lists twenty-one such cases in ten countries. Residuals with a positive sign are those where the RAI estimate is higher than the alternative measure. Disagreement is often greatest when a country has multiple tiers of regional government; when there is considerable variation in decentralization over time; or when there

Table 2.6. Cases of disagreement

Country	Years	(Range of) z-scores	Measurement
Belgium	1989–2000	+2.20/+2.74	Brancati
	1980–1994	+2.24/+3.40	Lijphart
	1990–1994	+2.01	Treisman
	1980–1998	+2.06/+3.13	Woldendorp et al.
Chile	1960–1974	–2.01	Arzaghi and Henderson
	1995–1999	–3.12	Arzaghi and Henderson
Finland	1950–1992	–2.06	Woldendorp et al.
France	1982–1996	+2.24/+2.40	Lijphart
	1990–1999	+2.25	Treisman
Germany	1977–1989	+2.05	Arzaghi and Henderson
	1985–2000	+2.52/+2.84	Brancati
	1990–1999	+2.43/+2.47	Treisman
	1977–1989	+2.01	Woldendorp et al.
Italy	1989–1996	+2.13/+2.43	Lijphart
Serbia and Montenegro	1992–2000	+2.54/+2.83	Brancati
	1992–1999	+2.15/+2.48	Treisman
Spain	1978–1999	+2.14/+3.15	Arzaghi and Henderson
	1983–1996	+2.39/+2.86	Lijphart
	1983–1998	+2.12/+2.70	Woldendorp et al.
Trinidad and Tobago	1985–1995	–2.02	Brancati
Venezuela	1950–1960	–2.70/–2.46	Lijphart

Note: A case of disagreement is defined as two standard deviations below or above the estimate for a time period of five years. A positive sign indicates that the estimate of the RAI is higher than the estimate of the alternative measure.

is differentiation. In some cases, different scores reflect more fundamental differences in conceptualization and operationalization.

Belgium and Germany stockpile the largest number of disagreements. The RAI scores diverge with four of the five alternative measures and in each case the RAI score is higher. For Belgium, the single most important factor is the cadence of reform—five major reforms between 1970 and 2005. Static measures such as Treisman or Woldendorp et al. are poorly equipped to capture this. Lijphart's measure is not entirely static since he increases Belgium's score from 3.1 to 5 in 1993 following federalization. However, Lijphart's measure does not pick up the regional empowerment that took place in the 1970s. The divergence with Treisman for 1990 to 1994 reflects a scoring disagreement: the Belgian senate does not meet Treisman's criterion for a regional chamber, while it does according to the RAI.

Brancati's measure, which is the only one to provide annual readings between 1985 and 2000, registers no change in Belgium, whereas the RAI spikes up in 1989 when Belgian regions and communities obtain broader policy competences, taxation powers, and shared rule. This alerts us to a difference in conceptualization. Brancati's measure emphasizes electoral and policy autonomy, but the central foci of the 1989 and 1993 reforms were tax autonomy, executive federalism, and a reform of the senate.

Disagreement between the RAI and the alternative measures in estimating decentralization in Germany appears to result from conceptual differences between the RAI and these measures. The RAI evaluates multiple tiers, and it pays close attention to shared rule; Germany has both multiple levels of regional governance and high levels of shared rule. The RAI picks up the authority exercised by regional governments within *Länder* (including *Regierungsbezirke* and *Kreise*) and it considers several dimensions of shared rule, including intergovernmental meetings between *Länder* and the federal government.

Disagreement with Treisman and Brancati also reflects coding judgments. Treisman's score of 1.5 out of a possible 3.0 for Germany is based on a restrictive interpretation of *Länder* authority: the absence of constitutionally entrenched exclusive powers, and the absence of an absolute veto by the *Bundesrat* on legislation (though it can raise the hurdle). The RAI, by contrast, considers concurrent powers and the role of *Länder* in implementing national framework legislation (Swenden 2006; Watts 1999a). Brancati scores Germany 3.0 out of a possible 5.0 because she estimates that constitutional amendments do not require *Länder* approval. The RAI registers that *Länder* have a veto on constitutional reform by virtue of their representation in the *Bundesrat*.

The Lijphart index disagrees with the RAI for three more countries: Venezuela, France, and Italy. Lijphart scores Venezuela significantly higher than

Crossvalidating the Regional Authority Index

the RAI for the 1950s. Venezuela receives a score of 4 out of 5 on the Lijphart index, which is consistent with the 1947 constitution for a centralized federation, but this constitution was never put into effect due to a military coup in 1948, and the new constitution of 1952 replaced elected by appointed officials at all levels (1948–57).¹⁰ The RAI also scores Italy after its 1989 reform and France after the Defferre reform of 1982 as having more decentralization than in Lijphart's measure.

Chile and Finland have higher scores in Arzaghi and Henderson and Woldendorp et al., respectively, chiefly because these authors include local government in their measure. The RAI estimates Chilean *provincias* and *regiones*, which are primarily deconcentrated, whereas Arzaghi and Henderson code municipal authorities as not being subject to central veto. The Woldendorp et al. measure scores Finland higher because it captures Finland's relatively authoritative municipal authorities while the RAI does not. From 1993, when Finland creates self-governing regional governments, which are picked up by the RAI, the two indices fall in line.

Three indices estimate Spain to have considerably less decentralization than the RAI. In contrast to Lijphart and Woldendorp et al., the RAI encompasses scores for *provincias* as well as *comunidades autónomas*. Arzaghi and Henderson consider both levels of governance, but their score is subdued because they focus on primary education, infrastructure, and policing—areas in which the central government retained substantial authority.

The RAI and Brancati differ on Serbia and Montenegro and Trinidad and Tobago on definitional grounds. Whereas Brancati scores Serbia, the RAI scores the federation and, from 2003, the "state union" of Serbia and Montenegro. Trinidad and Tobago consists of two main islands but only Tobago has an intermediate tier of government. Brancati's score for Tobago is the same as for the country as a whole, while the estimate of the RAI is lower because the score for Tobago is weighted by its population size.¹¹

Two remaining cases of disagreement with Treisman are France and Serbia and Montenegro. Treisman gives France a score of zero because his coding registers only constitutional provisions, while the authority exercised by *départements* and *régions* is laid down in special legislation. Serbia and Montenegro has a score of 1 on a scale from zero to 3, which is surprisingly low for a (con) federation. Again, Treisman's emphasis on constitutional criteria explains this. Serbia and Montenegro's upper chamber is not coded as regional, probably because it was not directly elected but made up of twenty deputies from each member republic. Instead, the RAI registers extensive shared rule through the

¹⁰ For greater detail, see the country profile of Venezuela.

¹¹ Tobago's population is 60,000 and that of the country as a whole is 1.3 million (2011 figures).

upper chamber, giving Serbia and Montenegro one of the highest scores in the RAI dataset.

Fiscal Indicators

Fiscal indicators are widely employed in studies of decentralization (see e.g. Blöchliger 2015; Blöchliger and King 2006; Braun 2000; Castles 1999; Harbers 2010; Oates 1972; Stegarescu 2005a; Willis et al. 1999). The principal sources are Government Finance Statistics (GFS) produced by the International Monetary Fund (IMF) and Historical National Accounts and Revenue Statistics produced by the OECD.¹² Authors interested in the effects of decentralization on outcomes such as economic growth, corruption, or redistribution, have used revenue and expenditure indices in combination (Akai and Sakata 2002; Enikolopov and Zhuravskaya 2007; Jin and Zou 2002). Some authors have sought to increase the validity of specific fiscal indicators (Ebel and Yilmaz 2002; Stegarescu 2005b).

Despite these efforts, two basic caveats remain when using fiscal indicators to tap regional authority (Blöchliger 2015; Rodden 2004; Schakel 2008; Sorens 2011). The first is that the extent of subnational expenditure or revenue does not indicate the autonomy of a subnational government from central control in spending money. *Departamentos* in Uruguay, for example, spend more than twice as much as a proportion of total government expenditure than those in Bolivia (15.4 percent versus 7.2 percent), but have less authority over taxes (Daughters and Harper 2007: 224). Subnational governments in South Korea were conduits for 34.4 percent of total government expenditure in 1978 (the latest year reported in World Bank data) at a time when the country was highly centralized under military rule. In the same year, popularly elected Malaysian subnational governments with diverse policy making powers were responsible for 17.2 percent of total government expenditure, and subnational governments in Indonesia, which were more authoritative than those in South Korea, spent just 13.4 percent of total government expenditure.

The amount a government spends does not tell us whether spending is financed by conditional or unconditional grants, whether the central government determines how the money should be spent, or whether it sets the framework legislation within which subnational governments implement (Blöchliger 2015; Akai and Sakata 2002; Breuss and Eller 2004; Ebel and Yilmaz 2002; Fisman and Gatti 2002; Martinez-Vazquez and McNab 1997; Panizza 1999). Figure 2.1 shows that subnational governments in

¹² We use the World Bank (2006) Fiscal Indicator dataset derived from the GFS (IMF) because it has the greatest overlap with the RAI: fifty-six countries with yearly scores for 1972–2000.

Measurement

direct bearing on the public budget, whereas regulatory policies, including civil and criminal law, may have considerable impact on society by virtue of the rules they impose. While the cost of expenditure programs is borne by the public budget, the cost of most regulatory policies is borne by citizens and firms (Majone 1994). To the extent that regions have control over regulatory policies, expenditure fiscal indicators reveal little about decentralization. Policy authority is captured by the RAI separately from fiscal authority.

Scholars have also produced measures for central grants to subnational governments (Akai and Sakata 2002; Oates 1972; Stegarescu 2005*b*). Vertical imbalance is the degree to which subnational governments rely on central government revenues to support their expenditures, and is measured by intergovernmental transfers as a share of subnational expenditures. This has been criticized because it does not identify whether a grant comes with a centrally imposed mandate (Shah 2007). This is a valid concern, but we do not have reliable data that distinguish between conditional and unconditional grants (Rodden 2004).

A second caveat is that fiscal indicators provide one score for all levels of subnational government. Fiscal decentralization indices do not distinguish between local and regional tiers and do not take differentiated governance into account.¹³ Further, the existence of regional governments with special powers can shape the level of decentralization in a society even if their powers are not generalized across an entire tier of subnational government. For example, the Basque *foru iurralde* (historic territories) and Navarre in Spain collect income, corporate, inheritance, and wealth taxes and can set the rate and base for these taxes autonomously, whereas in the rest of Spain the bulk of taxes are paid to the center and set amounts are transferred back to the regions (Swenden 2006). The five special *regioni* and the *provinces* of Bolzano-Bozen and Trento in Italy receive a share of taxes collected in their jurisdictions whereby the central government sets the base but the rate is negotiated bilaterally between the region and central government. In contrast, the tax autonomy of ordinary *regioni* is limited: they can set the rate within centrally determined limits for minor taxes (the vehicle tax, an annual surtax, a special tax on diesel cars, and health taxes). Failing to capture this variation can over- or underestimate fiscal autonomy.

Figure 2.2 plots subnational revenue as a percentage of total government revenue against the RAI country score. The correlation is statistically significant, but a closer look reveals that Sweden and Denmark are ranked on a par with Argentina and have higher scores than Australia, Austria, Brazil, and Mexico. Counties in Denmark and Sweden may set the rate of income tax within central government parameters but it would be wrong to conclude that

¹³ As before, we exclude countries without a regional tier.

Crossvalidating the Regional Authority Index

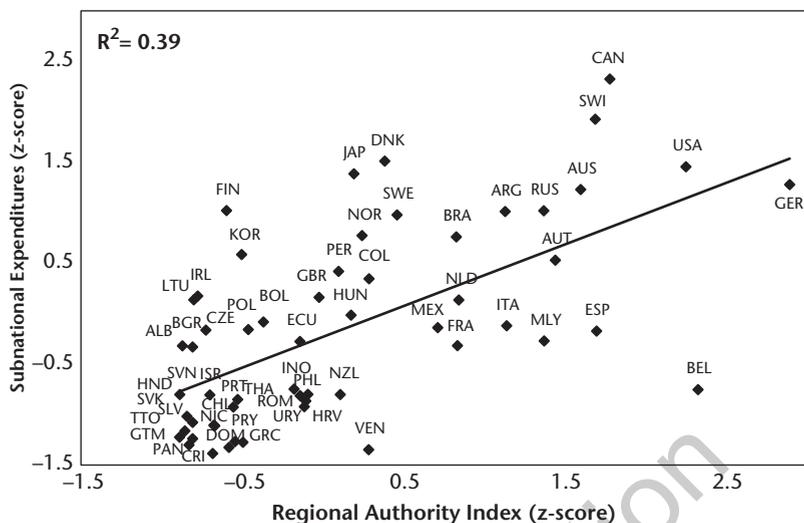


Figure 2.2. Subnational revenue and regional authority

Note: Subnational revenue as a percentage of total government revenue plotted against country scores on the RAI. World Bank (2006) Fiscal Indicator dataset. Standardized scores for averages for 1972–2001.

the subnational tiers in Sweden and Denmark enjoy the same autonomy as their peers in these federal countries.

One way to gain more insight into central involvement in subnational revenue and expenditure is by looking at the share of intergovernmental grants (Akai and Sakata 2002; Blöchliger 2015; Breuss and Eller 2004; Oates 1972; Stegarescu 2005*b*). A common measure is vertical imbalance, which is operationalized by intergovernmental transfers as a share of subnational expenditures. However, this indicator is also limited. Aside from data availability regarding unconditional and conditional grants (Rodden 2004), there is the problem that intergovernmental grants do not seem to differentiate between federal and non-federal countries.

Figure 2.3 displays vertical imbalance against RAI country scores. One would expect a negative relationship between vertical imbalance and RAI scores since high percentages of central government grants relative to total subnational revenue should be associated with low scores on the RAI. As one can observe in Figure 2.3 the vertical imbalance in decentralized federal countries such as Canada, Switzerland, and the US is comparable to that in centralized unitary countries such as Bulgaria, Lithuania, and Slovakia.¹⁴

¹⁴ An analysis of variance (ANOVA) shows that the extent of vertical imbalance does not vary significantly between unitary and federal countries ($F: 1.99; df = 52, p = 0.147$).

Measurement

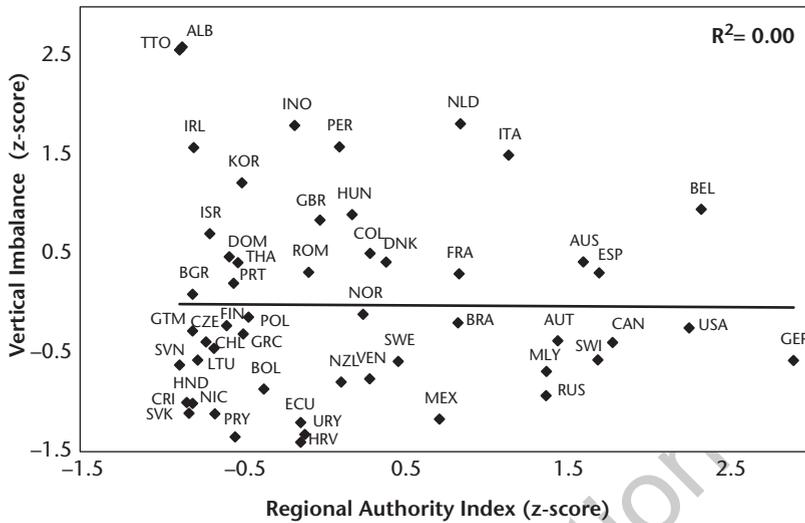


Figure 2.3. Vertical imbalance and regional authority

Note: Intergovernmental transfers as a share of subnational expenditures plotted against country scores on the RAI. World Bank (2006) Fiscal Indicator dataset. Standardized scores for averages for 1972–2001.

Fiscal measures conflate meaningful decentralization with change in public spending (Stegarescu 2005*b*). Fiscal decentralization may differ between two countries even in the case of an identical allocation of policies and functions across tiers of government (Oates 1972; Panizza 1999). A country that spends relatively more on policies that are centralized for scale efficiency reasons (such as defense) will also be more fiscally centralized.¹⁵ A similar argument applies to the question whether welfare state policies are provided by the government or by the private sector. For example, in the Scandinavian countries, a large proportion of government expenditure goes to welfare state policies and these are often provided by subnational governments. In market-liberal Anglo-Saxon countries, many welfare state functions are privatized. So a difference in political economy leads to higher expenditure (and revenue) in Scandinavian countries than in Anglo-Saxon countries, whereas the allocation of functions among levels of government may be identical.

¹⁵ The World Bank (2006) notes that particular expenditure categories can distort measures of decentralization: “For instance, the United States, despite being a much larger country, has a lower sub-national share of expenditures than Switzerland. However, when defense and interest expenses are excluded from the subnational-to-total ratio, the United States has a higher subnational share of expenditures than Switzerland.” Some scholars therefore use fiscal indicators that exclude defense, e.g. Breuss and Eller (2004) and Panizza (1999).

Crossvalidating the Regional Authority Index

The RAI avoids conflating these distinct political processes by measuring autonomy rather than expenditure.

Conclusion

The extent to which states are decentralized has been a topic of enduring interest in political science. Measures of decentralization are used in explaining a wide range of political outcomes concerning public policy, the quality of governance, and economic performance. Scholars have employed institutional indicators of decentralization or used fiscal indicators based on World Bank data on subnational expenditures and revenues.

This chapter crossvalidates the RAI with five commonly used institutional measures, and finds much agreement. Notwithstanding their marked differences in conceptualization, operationalization, and coverage, a single underlying factor accounts for more than 79 percent of the variance. Decentralization appears to have a core meaning that can be tapped by measures using very different indicators. But we also detect systematic sources of disagreement and divergent interpretation of evidence.

The most consistent sources of systematic disagreement arise from the fact that the RAI is better equipped to account for multilevel regional governance and for regionalization over time because it estimates the authority of subnational governments at multiple levels and because it produces annual observations over six decades. Beyond the systematic differences among these measures, there are numerous differences in the interpretation of particular cases. In this chapter we assess these on a case by case basis drawing on documentary evidence.

The associations between the RAI and World Bank data on subnational expenditures and revenues are relatively weak at 0.60 and 0.59, respectively. In their current form, fiscal indicators are not good at capturing whether regional governments decide autonomously over revenues and expenditures, nor do they encompass authority over regulation that does not involve much money.

The comparisons among the available measures in this chapter suggest that there is non-negligible consensus among experts in estimating decentralization at the country level. However, the purpose of the RAI is to measure authority at the regional level, a more difficult and perhaps more hazardous undertaking. Our concern in the next chapter is with accuracy rather than consensus, and this leads us to engage the substantive content of the RAI, probing the theoretical, conceptual, and operational decisions that underpin it.

Measurement

Appendix

Table 2.A.1. Country coverage across measures of decentralization

	A-H	BRA	LIJP	TRE	WKB	Total
Albania				X		1
Argentina	X	X		X		3
Australia	X		X	X	X	4
Austria		X	X	X	X	4
Belgium		X	X	X	X	4
Bolivia		X		X		2
Bosnia and Herzegovina		X		X		2
Brazil	X	X		X		3
Bulgaria		X		X	X	3
Canada	X	X	X	X	X	5
Chile	X	X		X		3
Colombia	X	X	X	X		4
Costa Rica		X	X	X		3
Croatia		X		X		2
Cuba				X		1
Czech Republic		X		X	X	3
Denmark		X	X	X	X	4
Dominican Republic		X		X		2
Ecuador	X	X		X		3
El Salvador		X		X		2
Finland		X	X	X	X	4
France	X	X	X	X	X	5
Germany	X	X	X	X	X	5
Greece	X	X	X	X	X	5
Guatemala		X		X		2
Haiti				X		1
Honduras		X		X		2
Hungary	X	X		X	X	4
Indonesia	X	X		X		3
Ireland		X	X	X	X	4
Israel		X	X	X	X	4
Italy	X	X	X	X	X	5
Japan	X	X	X	X	X	5
Lithuania		X		X	X	3
Malaysia	X	X		X		3
Mexico	X	X		X		3
Netherlands	X	X	X	X	X	5
New Zealand		X	X	X	X	4
Nicaragua		X		X		2
Norway		X	X	X	X	4
Panama		X		X		2
Paraguay		X		X		2
Peru	X			X		2
Philippines	X	X		X		3
Poland	X	X		X	X	4
Portugal		X	X	X	X	4
Romania	X	X		X	X	4
Russia	X	X		X		3
Serbia and Montenegro		X		X		2
Slovakia		X		X	X	3

(continued)

Crossvalidating the Regional Authority Index

Table 2.A.1. Continued

	A-H	BRA	LJJP	TRE	WKB	Total
Slovenia		X		X		2
South Korea	X	X		X		3
Spain	X	X	X	X	X	5
Sweden		X	X	X	X	4
Switzerland		X	X	X	X	4
Thailand	X	X		X		3
Trinidad and Tobago		X	X	X		3
Turkey	X	X		X	X	4
United Kingdom	X	X	X	X	X	5
United States	X	X	X	X	X	5
Uruguay		X		X		2
Venezuela	X	X	X	X		4
Total	29	57	26	62	30	204
Coverage	47%	92%	42%	100%	48%	66%

Note: Nineteen countries without a region or regional tier are excluded.

A-H = Arzaghi and Henderson; BRA = Brancati; LJJP = Lijphart; TRE = Treisman; WKB = Woldendorp, Keman, and Budge.

Table 2.A.2. Descriptive statistics

Measurement	Mean	St.dev.	Min	Max
Regional Authority Index	9.62	9.35	0.00	36.95
Arzaghi and Henderson	1.93	1.21	0.00	4.00
Brancati	2.09	1.13	0.00	5.00
Lijphart	2.56	1.54	1.00	5.00
Treisman	0.51	0.86	0.00	3.00
Woldendorp, Keman, and Budge	3.40	1.92	0.00	7.00
World Bank subnational expenditure	23.31	15.21	1.45	59.18
World Bank subnational revenue	17.27	13.73	0.13	54.60
World Bank vertical imbalance	36.96	22.82	0.29	96.60

Table 2.A.3. Pairwise Pearson correlations

	1	2	3	4	5	6	7	8	9
1	1.00								
2	0.84**	1.00							
3	0.73**	0.67**	1.00						
4	0.78**	0.78**	0.71**	1.00					
5	0.77**	0.63**	0.71**	0.69**	1.00				
6	0.77**	0.86**	0.73**	0.80**	0.67**	1.00			
7	0.60**	0.71**	0.50**	0.56**	0.49**	0.79**	1.00		
8	0.59**	0.67**	0.56**	0.62**	0.50**	0.82**	0.94**	1.00	
9	-0.02	-0.25**	-0.09*	-0.27**	-0.09	-0.42**	0.05	-0.27**	1.00

Note: * $p < 0.05$; ** $p < 0.01$.

1 = RAI; 2 = Arzaghi and Henderson; 3 = Brancati; 4 = Lijphart; 5 = Treisman; 6 = Woldendorp, Keman, and Budge; 7 = World Bank subnational expenditure; 8 = World Bank subnational revenue; 9 = World bank vertical imbalance.

Measurement

Table 2.A.4. Pairwise Pearson correlations between the residuals of the alternative measurements

		1	2	3	4	5
1	Arzaghi and Henderson	1.00				
2	Brancati	0.47	1.00			
3	Lijphart	0.56	0.71	1.00		
4	Treisman	0.34	0.62	0.56	1.00	
5	Woldendorp, Keman, and Budge	0.70	0.73	0.70	0.59	1.00

Note: all Pearson correlations are statistically significant at the $p < 0.0001$ level. The residuals are obtained by regressing the decentralization index on the RAI.

Table 2.A.5. Descriptive statistics of the independent variables

Variable	Mean	St.dev.	Min	Max
Limited coverage	3.17	1.32	0.00	5.00
Distance in time	35.00	14.93	11.00	61.00
Tiers	1.14	0.50	0.00	2.86
Differentiation	4.80	7.04	0.00	24.00
Reform (A-H)	6.04	8.47	0.00	58.26
Reform (BRA)	1.67	2.55	0.00	13.92
Reform (LIJP)	3.59	4.69	0.00	29.26
Reform (TRE)	1.02	1.84	0.00	9.00
Reform (WKB)	3.62	4.91	0.00	30.20

Note: A-H = Arzaghi and Henderson; BRA = Brancati; LIJP = Lijphart; TRE = Treisman; WKB = Woldendorp, Keman, and Budge.

3

How We Apply the Coding Scheme

Credible measurement hinges on the clarity and consistency of the principles that guide scoring. Chapter One (Figure 1.1) describes six steps from the abstract to the particular. The process begins by theorizing the background concept of political authority, which is then specified as a basis for disaggregating the concept into the domains of self-rule and shared rule. This provides a frame for theorizing dimensions which, in turn, nest indicators. This takes us closer to empirics, but one still has to bridge the conceptual distance between an indicator and an observation (i.e. a score for a case). This chapter seeks to make the decisions from indicator to observation transparent. This is not a story with a plot or finale, but an exercise in grappling with puzzles, each with its own tricky facets. Be warned: the elegance that one observes in the restaurant is little in evidence in the kitchen. In short, this chapter is intended for those who are not satisfied with eating the meal we have prepared, but who wish to probe through the steam and smoke to see the cooks at work.

A coding scheme—a set of items on a limited number of dimensions—should be inter-subjective so that it can produce convergent scores. However, particular cases will usually involve expert judgment no matter how carefully an item is formulated. Expert coding cannot be reduced to an algorithm, but involves disciplined conceptual problem solving as well as detailed knowledge of the cases themselves.

Disciplined conceptual problem solving is another way of saying “theory.” To get a taste of this, dip straight into the section on Constitutional Reform (p. 96ff.). In order to compare the authority of regions in the process of constitutional reform one must make a series of theoretical decisions that allow one to abstract from the particularities of individual countries and regions. The result is a conceptual framework for analysing constitutional reform. In the words of Stephen Jay Gould (1998: 155): “Theory and fact are equally strong and utterly interdependent; one has no meaning without the

other. We need theory to organize and interpret facts, even to know what we can or might observe.”

If you wish to work your way through this chapter (or parts of it), you will find it helpful to have the coding scheme within reach. Each section is self-contained so readers can consult those sections that most interest them.

Scoring raises some general challenges. The first is minimalism, the principle of specifying the essential properties of a concept by eliminating its superfluous connotations. The second is specificity, the principle that each interval should identify a unique condition on a monotonous dimension. These objectives can be broken down as follows:

- *Defining content*—precision in defining what is encompassed, and what is excluded, in a dimension.
- *Specifying intervals*—clarity in specifying what a minimum score stands for, and what one expects to find with successively higher scores.
- *Avoiding formalism*—judgment in applying formal coding rules in diverse contexts.
- *Triangulating estimates*—searching for alternative sources of evidence.
- *Avoiding contagion*—insulating the object of a measure from its causes and consequences.
- *Adjudicating ambiguity*—evaluating gray cases that can plausibly be scored in more than one way.

The *modus operandi* of this chapter is to make judgments explicit, particularly on sticky issues or where we feel that we may have erred. This is especially important because we cannot assess the reliability of our measure. Rather than use independent coders whose reliability can be evaluated through comparison, we deliberate as a team to increase the validity of our estimates.¹ We indicate three kinds of uncertainty in the text of the country profiles that follow this chapter:

- For an estimate based on thin information we use the symbol α .
- For a case that falls between intervals we use the symbol β .
- Where the sources disagree we use the symbol γ .

¹ Reliability, i.e. the extent to which estimates converge in multiple trials, is necessary but not sufficient for validity, which is the extent to which a measure accurately measures what it is supposed to.

How We Apply the Coding Scheme

GROUND RULES

What is a regional government? We define a regional government as a government that is intermediate between the national and the local. We code standard regions where the units at that level had, in 2010, an average population of 150,000 or more, and we code non-standard or differentiated regions irrespective of population. We encompass metropolitan regions where these perform regional government tasks in urban areas.

Which year do we code? The dataset covers the period 1950–2010. We code institutional change from the year in which a reform comes into effect. We score a reform in representation in the year of the first election to which it applies.

How do we justify a coding decision? Our objective is to link each coding decision with the particular formal rules that regulate regional authority as laid down in executive decrees, laws, constitutions, statutes, or other documents. The profiles reference the specific articles, paragraphs, or sections that pertain to the coding decision. We triangulate with secondary sources and with expert judgments.

The remainder of this chapter is divided into sections for each of the ten dimensions in the coding scheme set out in Chapter One. Each section explicates the meaning of the dimension and how we break it into intervals. It then examines the issues and ambiguities that arise as one applies this to empirical cases.

Self-rule

Institutional Depth

We conceive institutional depth as a continuous dimension ranging from “no autonomy from the central government” to “complete autonomy.” The latter is a conceptual, but not an empirical, possibility. The variation is mostly at the lower end of the scale and the intervals are spaced accordingly.

We distinguish four categories. The first is a null category where there is no functioning general purpose regional administration. The second is described by the Napoleonic term, *déconcentration*, which refers to a regional administration that is hierarchically subordinate to central government. A deconcentrated regional administration has the paraphernalia of self-governance—buildings, personnel, budget—but is a central government outpost.² The final two categories distinguish between regional administrations that exercise meaningful authority. The more self-governing a regional government, the less its decisions are subject to central government veto.

² Hence a deconcentrated, general purpose region typically scores 1 on institutional depth, but zero on all other dimensions. A handful of deconcentrated regions add to this some representation or, in one case, shared rule.

Measurement

The box below sets out the categories. Several conceptual decisions are called for. What is a functioning government? What do we mean by general purpose? What conditions can bring us to conclude that there is *no* central government veto? How do we distinguish between formal and informal authority? And finally, how does authoritarianism affect institutional depth? We discuss these in turn.

INSTITUTIONAL DEPTH	
0:	no functioning general purpose administration at the regional level;
1:	a deconcentrated, general purpose administration;
2:	a non-deconcentrated, general purpose administration subject to central government veto;
3:	a non-deconcentrated, general purpose administration not subject to central government veto.

To score more than zero, a region must have a functioning administration. Purely statistical regions—regions created on paper for legal or statistical purposes—do not reach the bar. Several European and Latin American countries set up regions for statistical convenience in economic planning, and only a subset of these evolve into functioning administrations that score 1 or more. To distinguish these cases, we begin with the question: does the administration physically exist? Does it have an office, employees, a postal address? We then assess what the administration does.

Governments that are incapacitated—by war, disaster, or dictatorial imposition—score zero. Incapacitation, in this context, is a general and durable condition; it must affect most or all units in a regional tier for at least two years. Most subnational governments in El Salvador ceased operations during its civil war (1980–92) and score zero for this period. We do not downgrade subnational governments that are dysfunctional because they are strapped for funds. It is not uncommon for subnational governments in poor societies to vary in functionality, but we wish to estimate the authority of a region independently from the extent to which it functions well or poorly.

To score more than zero, a region must be general purpose—not task-specific. We use the term general purpose governance to describe jurisdictions that “bundle together multiple functions, including a range of policy responsibilities, and in many instances, a court system and representative institutions. . . . Type I jurisdictions express people’s identities with a particular community” (Hooghe and Marks 2010: 17, 27; 2003). A task-specific jurisdiction, by contrast, provides a specialized public good for a constituency that happens to share a problem or circumstance.

Regions are task-specific when each national ministry controls its own regional subdivision. In Thailand, centrally appointed governors who ran

How We Apply the Coding Scheme

changwat (provinces) had little control over a parallel structure of deconcentrated units set up by sectoral ministries in Bangkok. We score the *changwat* as a weak form of general purpose governance, and the ministerial sections, which took most decisions, are task-specific. Regional governments may also be task-specific if they are responsible for just a single policy. Dutch *waterschappen* (water boards) are task-specific jurisdictions in a country that lies mainly below sea level. In Peru, neither *Organismos de Desarrollo* (Development Entities, ORDE) nor *Corporaciones de Desarrollo* (Development Corporations, CORDE) meet the criterion of general purpose government. *Organismos*, established in 1975, coordinated several regional offices that specialized in regional development. In 1981, they were replaced by *Corporaciones*, which were limited to public works management. These institutions vied with *departamentos*, which coordinated central policy across a broad sweep of policies. *Departamentos* score 1 on institutional depth; *Organismos* and *Corporaciones* score zero.

Several countries have regional administrations that shift from task-specific to general purpose governance. A 1974 reform in New Zealand replaced task-specific with general purpose regions, as did a 1994 reform in England that set up deconcentrated general purpose regions. The regions in England (except for Greater London) were abolished in 2012 and were replaced by task-specific agencies (quangos). Finland's *läänit* were abolished in 2010 and their tasks allocated to deconcentrated central government outposts (*aluehallintovirastot*) and task-specific jurisdictions (*ELY-keskus*) which manage subsidies from the European Union. Regions in Costa Rica and Lithuania took the opposite path, from general purpose to task-specific governance. In 1996, Costa Rican *departamentos* were reduced to statistical categories and task-specific *mancomunidades* filled the gap. In 2010, Lithuania abolished self-governing *apskritis* and centralized their tasks in sectoral ministries, some of which set up regional outposts.

Scores at the upper end on this dimension depend on whether a regional administration is subject to central government veto. This turns on whether a region has legally enforceable protection against central government *ex ante* and *ex post* control. Such is the case when regional and central law have equal constitutional status. Federalism is the most common institutional expression of this, but it is worth noting that federalism is neither sufficient nor necessary.

Argentina is a federal country although its *provincias* have been subject to a constitutional clause that permits federal intervention “to guarantee the republican form of government or to repel foreign invasions, and upon request of its authorities created to sustain or re-establish them, if they have been deposed by sedition or by the invasion of another province” (C 1853, Art. 6; C 1994, Art. 6). Federal intervention was frequently invoked under both

Measurement

democratic and military rule, but it fell into disuse with the return to democracy in 1983 and was formally circumscribed in the 1994 amendment of the constitution making federal intervention subject to prior congressional approval. One can debate the timing of transition (and we do in the profile), but it makes sense to increase the score to 3 (“not subject to central veto”) for the recent period.³

On the other hand, the United Kingdom is not a federation, but the only ground on which a Secretary of State can refuse to submit a bill from the Scottish parliament for royal assent is if it has “an adverse effect on the operation of the law as it applies to reserved matters” or is “incompatible with any international obligations or the interests of defense or national security” (Law No. 46/1998, Art. 351). This has never happened. Scotland scores 3 from 1999. The same applies to Northern Ireland, which has had a similar provision from 2000, and Wales from 2011.

The distinction between 3 and 2 is nicely illustrated in Belgium. Since 1989 the communities and regions score 3 on institutional depth. A special law with constitutional force prohibits the central government from suspending or vetoing decrees passed by regions and communities. Conflicts between decrees and laws are adjudicated by an arbitration court with balanced national and subnational representation (Ales 1989). In contrast, the Brussels region continues to score 2 on institutional depth. Indeed, the national government can suspend and ultimately annul Brussels’ decisions on urban development, city and regional planning, public works, and transport on the ground that they detract from Brussels’ role as an international and national capital. Moreover, the legal status of Brussels’ ordinances is subordinate to that of national laws and community or regional decrees. Local courts can declare Brussels’ ordinances void if they are in breach of higher law (Ales 1989).

The region of Aceh in Indonesia walks a fine line between a score of 2 or 3. The 2006 Law on the Governing of Aceh, which is the bedrock for Aceh’s special status, does not exclude a central government veto. For example, the stipulation that “the central government sets norms, standards, and procedures and conducts the supervision over the implementation of government functions by the Government of Aceh and District/City governments” (Art. 11.1) provides openings for substantial central government authority over areas that otherwise fall under regional governance. We lean on the secondary literature and the judgment of experts such as Al Stepan and his colleagues (2011: 242–52) to come down for a score of 3.

³ We argue that 1983 is the more defensible date.

How We Apply the Coding Scheme

We code formal authority—not the exercise of power—in determining the score on institutional depth. Ireland’s regions illustrate how the two can diverge. A 1994 law establishes regions as “authorities” equipped with an executive, an indirectly elected assembly, and a small permanent staff under the mandate to coordinate EU structural funding and public service delivery among local authorities. In our 2010 book we considered them to be decentralized general purpose governments. We reconsidered our judgment after a recognized expert on local government wrote to us that “[w]hile it is true that this role of coordinating public services is expressed in Irish legislation establishing both the regional authorities and regional assemblies, in practice the extent to which regions have any role in this area has been extremely limited. The regional authorities have a mandate to prepare regional planning guidelines under spatial planning legislation (which is done only once every five years). A small minority have played a modest one-off coordination role in waste management. In both cases (spatial planning and waste management), the primary responsibility lies with local (not regional) authorities.”

While this expert confirms the legal and operational basis of intermediate government in Ireland, his comments spurred us to recode Irish regional authorities as deconcentrated. This appears to be a close call. While they have some paraphernalia of decentralized government, including an assembly and executive composed of senior management from local authorities, we conclude that the mandate to prepare regional planning guidelines under spatial planning legislation leaves little room for autonomy.

Regime type affects institutional depth, but authoritarianism rarely operates as a light switch. Our first move is to code change in formal rules relating to each of the ten dimensions of regional authority. While we are keenly aware of the character of the regime, we wish to estimate regional authority independently from regime change. An authoritarian regime may abolish national but not regional elections; it may replace a directly elected governor by a central appointee but leave the regional assembly unaffected; or it may centralize control over police but not over economic development or social policy.

If a regional tier is suspended or abolished, we code it zero on institutional depth. Few authoritarian regimes go this length. This has happened in just two countries in our dataset—Chile (departments) and Cuba (provinces)—and in both cases abolition was temporary, partial, or counter-balanced by the creation of a new tier. We find that most cases of abolition take place in democracies, including Costa Rica (1995), Denmark (2007), Finland (2010), Germany (Regierungsbezirke in some *Länder*), Greece (2011), Lithuania (2010), and the US (counties in Connecticut).

Measurement

Institutional depth drops by 1 if authoritarian rule reduces the institutional autonomy of regional governance, that is, if it tightens the overall supervision and control of central government over subnational government. Again, the incidence, timing, and severity vary.

The checkered history of Aceh in Indonesia illustrates this. Aceh, which had been a self-governing region in the dying days of Dutch colonialism, was curbed under the Sukarno and Suharto regimes. The territory lost its provincial status in 1951 and was at first run by the military (Reid 2010a, 2010b). It regained provincial status in 1957 and was declared a “special region” in 1959. But the incoming Suharto regime downgraded its special status from 1966 and, along with other *provinsi-provinsi*, it became deconcentrated in 1974. In 2001 following the transition to democracy, Aceh regained special autonomy, and in 2006 it was granted additional powers (Bertrand 2007, 2010; Stepan, Linz, and Yadav 2011). Elsewhere in Indonesia, first and second tier regional governments—*provinsi-provinsi* and *kabupaten-kabupaten/kota-kota*—retained self-government under Sukarno, but the New Order regime of Suharto gradually tightened central control, and in 1974 the regime formally revoked the self-government legislation of the 1950s (Bertrand 2007: 577).

Our coding seeks to capture these developments in the following way. We code Aceh separately from 1950 when its path already diverged from the *provinsi-provinsi*. Aceh has zero institutional depth for 1951–56; it scores 2 for 1957–73 to reflect limited institutional self-governance, and then 1 from 1974; 2 from 2001–06, and 3 thereafter. We distinguish between the Sukarno and Suharto periods for all *provinsi-provinsi*. The exact timing of the downscaling to deconcentrated government under Suharto is debatable. We opt for 1974 rather than 1966, because, while the Suharto regime moved fast to weaken provincial and district governance through executive and military orders soon after the 1966 coup, regional self-governance was not formally repealed until the law of 1974. Even after 1974, the regime continued to tolerate direct elections of provincial and district assemblies, but these were heavily regulated and the center wielded a veto over provincial governors and district mayors (Shair-Rosenfield, Marks, and Hooghe 2014). Indonesia under the New Order was highly centralized with “the lower levels of government simply implement[ing] directives” (Bertrand 2010: 175).

In contrast, the transition in Malaysia from democracy to authoritarianism after the 1969 race riots did not significantly redraw authority relations. The first postcolonial Malaysian constitution of 1957 put in place a relatively centralized federal framework that favored the central government over the *negeri* (Kok Wah Loh 2010; Stubbs 1989; Taylor 2007). *Negeri* score 2 on institutional depth, except Sabah and Sarawak which score 3 on the basis of their special constitutional status. After 1969, the “soft authoritarian”

How We Apply the Coding Scheme

government fortified its control over economic policy making but did not challenge federalism, so we score 2 on institutional depth.

In Brazil, institutional depth is decreased from 3 to 2 in 1964 following the *atos institucionais* (institutional acts) which enhanced central control over *estados*. The acts made it easier for the regime to displace opposition governors, which it proceeded to do. Central control was enhanced under the 1967 constitution, but this did not reduce *estados* to deconcentrated units (Eaton 2001*b*; see also Dickovick 2011; Falletti 2011). There is, then, no reason to drop institutional depth to a score of 1.

Policy Scope

Policy scope taps regional authority over the range of government policies, which we group in the following five categories:

- economic policy: regional development, public utilities, transport including roads, environment, and energy;
- cultural–educational policy: schools, universities, vocational training, libraries, sports, cultural centers;
- welfare policy: health, hospitals, social welfare (e.g. elderly homes, poor relief, social care), pensions, social housing;
- institutional–coercive policy: residual powers,⁴ police, own institutional set–up, control over local government;
- policy on community membership: immigration, citizenship, right of domicile.

In this section we discuss four basic scoring issues. First, we outline criteria for determining whether a regional government has authoritative competences in one or more of these policy areas. Second, we explain why we think authority regarding community membership is special. Third, we come to grips with the fact that central governments and regions often share authority. And finally, we take up the perennial challenge of deciding where formal rules end and practice begins.

The box below operationalizes regional policy scope across four intervals. These do not interpret themselves, but rest on a set of “rules about the application of rules” which are best explained using examples.

⁴ Residual powers are competences not constitutionally mandated to other jurisdictions.

POLICY SCOPE

- 0: the regional government has very weak or no authoritative competence over (a) economic policy, (b) cultural–educational policy, (c) welfare policy, or (d) institutional–coercive policy;
- 1: the regional government has authoritative competence in *one* of (a), (b), (c), or (d);
- 2: the regional government has authoritative competences in *at least two* of (a), (b), (c), or (d);
- 3: the regional government has authority in (d) plus at least two of (a), (b), or (c);
- 4: the regional government meets the criteria for 3, *and* has authority over immigration, citizenship, or right of domicile.

By “authoritative” we mean having the capacity to develop binding rules through legislation or executive orders. This capacity can be exercised solely by a regional government or, more usually, it is exercised concurrently with governments at other scales. If regional office holders have meaningful discretion—an autonomous capacity to set and pursue priorities—they need not have primary authority to warrant a positive score on this dimension.

Competence in the field of community membership is required for a maximum score. Authority over immigration, citizenship, or right of domicile are “fundamental sovereign attributes,”⁵ and regions that meet this high hurdle will already have authority in several substantive policies. Every region in the dataset that has competence in community membership also meets the criteria for a score of 3.

Many regional governments execute aspects of immigration or citizenship policy on behalf of central governments, but few have significant legislative authority over one, let alone both, areas. Just four regional tiers and six individual regions in our sample meet this criterion: the Australian states, Swiss cantons, Quebec, the Finnish Åland islands, Sabah and Sarawak in Malaysia, the two entities in Bosnia and Herzegovina, the republics in Serbia-Montenegro (until 2006), and Bashkortostan (until 2004) and Tatarstan (until 2006) in Russia.

In Switzerland, immigration and asylum is a confederal competence, but citizenship is primarily cantonal (Church and Dardanelli 2005: 173). The confederation regulates citizenship by birth, marriage, or adoption, and lays down minimum requirements for naturalization. However, the cantons can specify residence requirements and can require a language or naturalization test. In Australia, citizenship is federal (following the Australia Citizenship Act

⁵ US Supreme Court, in *Hampton v. Mow Sun Wong*, 426 U.S. 88, 101 n. 21 (1976).

How We Apply the Coding Scheme

of 1948), but regulation of immigration is a concurrent competence. A state can grant a visa to a skilled worker provided he or she passes a federal points test. The federal government has its own skills program, and also allocates family and student visas. By contrast, the states of the US score 3, not 4, on this dimension. The US constitution grants states some authority to regulate the conduct of foreigners, but immigration and naturalization are exclusive federal competences.

Canada and Quebec illustrate what it takes to move from 3 to 4. While immigration is a concurrent competence in the constitution, provincial authority remained a dead letter for decades. One might say that there was no “enabling law” until pressure from Quebec finally led to the 1978 Canada Immigration Act authorizing the federal government to conclude federal–provincial cooperation agreements on the subject. Cooperation became exclusive regional control following the Canada–Quebec Accord of 1991 which gave Quebec “sole responsibility for the selection of immigrants destined to that province” and commanded the Canadian government to “admit any immigrant destined to Quebec who meets Québec’s selection criteria” (Canada–Quebec Accord 1991, Art. 12; Simeon and Papillon 2006). After 1996, all Canadian provinces were able to “nominate” immigrants, and most do so, though, outside Quebec, the federal government still makes the final decision. Canadian provinces are, then, in a weaker position than Australian states, which can select immigrants within federal regulations. Quebec receives a score of 3 on policy scope from 1950–90 and 4 from 1991–2010, and provinces score 3 throughout the period.

The Åland islands score 4 since its government has exclusive authority to determine right of domicile in the islands which an individual needs in order to vote, stand for election, purchase, lease, or inherit property, or open a business on the islands. The Åland government grants domicile to all individuals with a parent who has the right of domicile and to others on a case-by-case basis. Similar provisions exist for Sabah and Sarawak which control immigration within their borders and issue visas to foreign visitors traveling from other countries or from other parts of Malaysia.

The Russian republics of Bashkortostan and Tatarstan had joint jurisdiction over citizenship under their bilateral treaties, but president Putin clawed back these provisions in 2005 and 2007, respectively (Chuman 2011: 135; Chebankova 2008: 1002).

Authority in systems of multilevel governance is often shared. Regional policy competences tend to be concurrent with central or, occasionally, local government. When does it make sense to say that a regional government has authority over a certain policy? To make headway, we must make some distinctions. Our primary concern is with constraints stemming from central control which can take several forms:

Measurement

- a dual structure of regional government in the form of parallel deconcentrated and decentralized administrations (e.g. military councils and *estados* in Venezuela (after 2000), or *län* and *landstinge* in Sweden);
- a mixed administration (e.g. a directly elected assembly and centrally appointed executive, as in Bolivia, France, or Thailand);
- a single administration that combines self-government and deconcentration (e.g. Dutch *provincies*).

In each of these situations, the score for policy scope reflects central constraints on a regional government's authority.

In Venezuela, Chávez' *Plan Bolívar 2000* established a parallel system to vie with *estado* and municipal governments (Hawkins 2010; Leon and Smilde 2009). The plan authorized the military to set up communal councils to arrange social services, including vaccinations, food distribution, and education, which would be implemented by "bolivarian missions" staffed by 40,000 soldiers. The dual system was constitutionalized in 2009. We acknowledge this shift in policy scope by reducing the score for *estados* from 2 to 1 in 2000.

In Sweden, responsibilities for governing the *län* (counties) are divided between *landstinge* (elected councils) and centrally appointed governors. Until 1970, *landstinge* provided health care along with occupational retraining. Centrally appointed governors had primary responsibility for law and order, local government, and implemented state legislation in health, education, and a broad range of economic policies. *Landstinge* score 1 for welfare, the core of their policy portfolio, but zero for economic development, which was heavily constrained by central regulation. In 1971, *landstinge* were given new tasks in regional development and public transport, at which point they score 2 for economic policy in addition to welfare.

Bolivian *departamentos* are dual structures with directly elected departmental councils which could propose policy initiatives and a centrally appointed prefect who made final decisions. The World Bank describes *departamentos* as "not yet fully autonomous subnational governments" (World Bank 2006: 13). *Departamentos* acquired competences in public investment, research, tourism, and welfare from 1995, but given the dominant role of the prefect we maintain a score of zero. With the introduction of direct elections for *prefectos* in 2005 we score policy scope 2. French *départements* and *régions* have a similar dual system in which the centrally appointed *préfet* has also lost some authority in recent years.

Thai *changwat* illustrate how the balance between decentralization and deconcentration can shift. Before 2004, the authority of directly elected assemblies in culture and education, infrastructure, and hospitals was shared with a centrally appointed governor. We adjust the score for policy from 1 to 2 when a regionally selected executive with competences in education, welfare, and economic planning, was established alongside the governor.

How We Apply the Coding Scheme

Finally, what is written and what is practiced may differ. Constitutional changes often require enabling legislation which may be scrapped, delayed, or diluted. To assess South Korean *do* (provinces) and *gwangyeoksi* (metropolitan cities) one must look beyond headline legislation to detect the timing of decentralization. The 1991 Local Autonomy Act authorized devolution in four broad areas: education, general welfare, and health; environment; agriculture and industry; and local government (Choi and Wright 2004). However, only education was devolved right away (hence, a score of 1). Decentralization took another step forward in 1999, when a new law laid down a procedure for transferring central competences in a broad swath of policies. However central departments and agencies continued to have the right to veto transfers—and actively used this to slow implementation of the law—a central constraint that is reflected in a score of 2, which would otherwise have been 3. After a third major law in 2003, which deprived central departments and agencies of the discretion to block or delay decentralization, the formal transfer of competences gathered pace (Bae 2007). From 2004, *do* and *gwangyeoksi* score 3 for policy scope. In this case, the implementation of the 1991 framework law stretched over twelve years.

Fiscal Autonomy

Regions may have fiscal authority in the form of taxation autonomy, co-decision on national tax regimes, and co-decision on intergovernmental grants (Swenden 2006). Our measure of fiscal autonomy captures the first of these, while the latter two fall under fiscal shared rule. Fiscal autonomy assesses a regional government's authority over its fiscal resources independently of their extent.⁶

The box describes how variation in fiscal autonomy is estimated across four intervals which distinguish between major and minor taxes and within these, between the capacity to control base and rate, or rate only.⁷ Below we delineate more precisely what is included in taxation (and what is not), which taxes are major or minor, and how we assess partial autonomy on setting the rate or base of taxes.

⁶ A 1999 OECD study distinguishes two notions of authority (control independent from central government, and shared rule with central government), and three areas of control (tax base, tax rate, and revenue split). Subsequent OECD studies refine these distinctions with an eye to estimating them (Sutherland, Price, and Joumard 2005; Blöchliger 2015; Blöchliger and King 2006: 10).

⁷ A tax is a “pecuniary burden upon individuals or property to support the government. . . . a payment exacted by legislative authority. . . [It is] an enforced contribution. . . imposed by government whether under the name of toll, tribute, tallage, gabel, impost, duty, custom, excise, subsidy, aid, supply, or other name” (Campbell 1979: 307). Similar taxes often have different labels. For example, the income tax on profits made by companies or associations is called corporate tax in the US, corporation tax in the UK and Ireland, and tax on enterprise profits in Russia. In Japan, it goes by several names depending on the taxing authority.

FISCAL AUTONOMY

- 0: the central government sets the base and rate of all regional taxes;
- 1: the regional government sets the rate of minor taxes;
- 2: the regional government sets the base and rate of minor taxes;
- 3: the regional government sets the rate of at least one major tax: personal income, corporate, value added, or sales tax;
- 4: the regional government sets the base and rate of at least one major tax: personal income, corporate, value added, or sales tax.

Fiscal autonomy “encompasses features such as a sub-central government’s right to introduce or to abolish a tax, to set tax rates, to define the tax base, or to grant tax allowances or reliefs to individuals and firms” (Blöchliger and King 2006: 9). It does not include a region’s authority to set fees or charges in return for specific services, such as fees for the preparation or deposit of official documents, bus charges, or public utilities. Fees are always tied to particular services and typically earmarked to be spent on sustaining these services. Thus, the Greater London Authority scores 1 because it can levy a property tax for which it can set rates, not because it can determine tube or bus fares or because it imposes a congestion charge for personal vehicles in central London. Royalties on mineral or other resources are considered a resource tax, not a fee, and fall under the category of minor taxes.

The distinction between major and minor taxes is somewhat arbitrary, though it is conventional to categorize personal income, corporate, value added, and sales taxes as major (Boadway and Shah 2009). Property taxes, resource taxes, excise taxes (e.g. on alcohol or cigarettes), registration taxes, etc. are usually considered minor. There are, of course, border cases. Argentine *provincias* signed away authority to tax income and sales in the 1930s in return for a share in federal taxes, though they retain control over the rate and base of a sales turnover tax, *ingresos brutos*, on companies’ gross revenues (Bonvecchi 2010; Falletti 2010). Until 1975, *provincias* also set a general tax on gross sales, which was eliminated when a federal VAT was introduced. Are these provincial sales taxes major? We argue that they are and that the abolition of the general provincial sales tax in 1975 constituted an important reduction in provincial tax autonomy which reduces fiscal autonomy from 4 to 2. *Provincias* also control inheritance tax, vehicle registration, and a stamp tax on property transactions, which are unambiguously minor taxes.

The Argentine example raises the broader issue of tax autonomy. National law may set parameters within which regions control the tax rate or base. In such cases one must assess the extent to which a regional government has discretion. Peru is a case where we judge this to be small. The 1979 constitution

How We Apply the Coding Scheme

gives *provincias* authority to decide the base and rate of several minor taxes, but leaves it to the central government to work out the modalities. Successive governments have consistently interpreted regional competences narrowly. The central government sets the base and determines the parameters for rate variation so that “such revenues are closer in concept to shared revenues (with a 100 percent share) than own-source taxes” (Ahmad and García-Escribano 2006: 15). We conclude that the tax base and rate are set centrally.

Borrowing Autonomy

Borrowing refers to the acquisition of money (on domestic or international financial markets or from domestic or international banks) against the obligation of future payment. For regional governments this can be a major source of income in addition to own taxes and intergovernmental grants. The extent to which regional governments have the authority to take on debt varies considerably across regions and over time.

The literature on public borrowing distinguishes numerical fiscal rules from procedural and transparency rules (Crivelli and Shah 2009; Ter-Minassian and Craig 1997). Numerical fiscal rules introduce some kind of ceiling on debt (Filc and Scartascini 2007; Rodden 2002). Procedural and transparency rules enhance transparency and accountability by requiring a government to publish a fiscal policy strategy and to routinely report fiscal outcomes (Ter-Minassian 2007).

Our measure of borrowing autonomy evaluates fiscal rules that constrain a region’s authority to borrow. The box below describes how we assess the extent of central government restriction. In this section we illustrate how we tackle a) differences between formal rules and practice, b) ambiguities in the bindingness of rules, and c) situations where more than two regulatory regimes co-exist. We begin by clarifying the concept of borrowing autonomy, and explaining what falls under the rubric of borrowing by a regional government.

BORROWING AUTONOMY

- 0: The regional government does not borrow (e.g. centrally imposed rules prohibit borrowing).
- 1: The regional government may borrow *under prior authorization (ex ante)* by the central government and it borrows under one or more of the following centrally imposed restrictions:
 - golden rule (e.g. no borrowing to cover current account deficits)
 - no foreign borrowing or borrowing from the central bank
 - no borrowing above a ceiling
 - borrowing is limited to specific purposes
- 2: The regional government may borrow *without prior authorization (ex post)* under one or more of the same centrally imposed restrictions.
- 3: The regional government may borrow without centrally imposed restrictions.

Measurement

In the domain of self-rule, we consider the extent to which a region may borrow autonomously, and in the domain of shared rule, we consider whether regions may collectively constrain subnational borrowing. We designate the former as “borrowing autonomy” and the latter as “borrowing control.”

We also need to be clear about what we understand by “regional government” in this context. A regional government may borrow for its own account or it may use intermediaries such as public companies or local saving banks. We encompass intermediaries provided the regional government controls the institution that contracts to borrow or, in the case of publicly listed companies, owns at least half of the shares. Particularly in countries with a statist tradition, governments sometimes provide public goods through public companies that they control at arm’s length. In such cases, the debts incurred may not show up in the core regional government budget. Still, they are financial commitments for which the regional government is ultimately accountable. In Croatia, a *županija* (canton) can issue guarantees for bank loans to a public institution/company in which it is a majority shareholder. A national law limits borrowing to 20 percent of total annual revenues which gives *županije* a score of 1.

The extreme values in the scoring scheme for borrowing autonomy are conceptually simple, but distinguishing them empirically can be challenging because the existence of rules constraining borrowing presumes that a regional government is able to borrow. A region scores zero under one of three conditions: when borrowing is explicitly prohibited by the central government; when a region has no history of borrowing; or when the regional government has no discretion over borrowing (i.e. it is deconcentrated).

At the top end of the scale, a region scores 3 when the following two conditions are met: a) there are no formal central rules regulating borrowing, and b) there is routine evidence of regional borrowing. The first of these criteria is met when a region is free to decide how much to borrow, from whom to borrow, and on what to spend the loan. Market constraints or self-imposed constraints do not negate this condition.⁸ It is not uncommon for regional governments to tie their own hands in order to enhance their credit standing, as has happened in Argentina, Canada, Switzerland, and the US. Many US states have constitutional or statutory provisions for a balanced operating budget and that allow borrowing only for capital projects (e.g. the construction of highways or schools) (Joumard and Kongsrud 2003; Plekhanov and Singh 2007). Some *provincias* in Argentina restrict borrowing

⁸ Discipline usually comes through credit ratings on subnational debt (Liu and Song Tan 2009: 2).

How We Apply the Coding Scheme

in their constitutions (Cetrángolo and Jiménez 2003; Nicoloni et al. 2002: 10). In such cases, regions score the maximum on borrowing autonomy.

The second criterion for a maximum score is that regions (or a significant proportion of regions in a regional tier) exercise their right to borrow. Rules on regional borrowing are of relatively recent vintage. In many countries there were no formal rules until the 1970s, but there was a clear norm that borrowing was not allowed. In recent decades, subnational borrowing has become more regulated, often in response to debt crises or, in the EU, in anticipation of monetary union (European Commission 2012; Rodden 2002, 2006; Sutherland, Price, and Joumard 2005). When there are no rules, we require systematic evidence of borrowing before assessing a maximum score. Does the absence of constraint indicate regional authority or does it simply indicate the perception that regulation is unnecessary because regions are not in the game of borrowing?

Naturally, there are gray cases. When Czech *kraje* (regions) were set up in 2000 they were not subject to constraints on borrowing. However, one region—Prague—did borrow, excessively it turned out. In 2001, a national law required prior central government approval for regional borrowing, and limited it to 15 percent of a region's budget. The central government refused to pay Prague's debt and the city resorted to selling property. We score *kraje* 1 on borrowing autonomy as of 2000 even though the law came into effect a year later.

Colombian *departamentos* show how borrowing evidence, regional government status, and rules all need consideration. Until the mid-1970s, *departamentos* were primarily deconcentrated: the governors, who decided on borrowing, were centrally appointed and received instructions from Bogotá. There was no regulatory framework, but regional borrowing was prohibited by the ministry of finance (Dillinger and Webb 1999b: 17, 19). By virtue of their deconcentrated status, *departamentos* score zero in this period. From the mid-1970s, *departamentos* acquired limited self-governance (Penfold-Becerra 1999: 199). Absence of borrowing and of explicit rules means we continue to score zero.

The two middle categories on this scale apply when regional borrowing is constrained by the central government, for example, to some proportion of a region's budget or to finance capital projects only. The distinction we make here is between central authorization that is *ex ante* (score=1) or *ex post* (score=2). Our premise is that *ex ante* control is substantially more imposing than control after the fact.⁹

⁹ The distinction between *ex ante* and *post hoc* control is consistent with that between an administrative and rule-bound approach to subnational borrowing (Ter-Minassian and Craig 1997).

Measurement

Contrast Colombia's regulatory framework of 1981 with that of 1997. The 1981 regime was rule-based. *Departamentos* and Bogotá could borrow after approval by the *asambleas departamentales* and the governor, in the case of *departamentos*, and the *concejo distrital*, in the case of Bogotá. Except for the prohibition to issue foreign bonds, restrictions on subnational borrowing were light. There was, for example, no *ex ante* control of cash advances from banks (Dillinger and Webb 1999b: 17–18). *Departamentos* receive a score of 2. In 1997, the Colombian government introduced a much more restrictive regulatory framework: it set strict ceilings on debt, created a fiscal and financial monitoring system involving a green, yellow, or red light, and authorized the central government to prohibit particular *departamentos* from borrowing (Daughters and Harpers 2007: 250; Olivera, Pachón, and Perry 2010: 29). That amounts to *ex ante* control, and so from 1997, *departamentos* score 1 on borrowing.

We code formal rules—even if not all governments abide by them. For example, since 1997 borrowing by Austrian *Länder* (states) is governed by the *Vorschlags- und Rechnungsabschlussverordnung* (federal financial decree), which limits borrowing to extraordinary expenses (Thöni, Garbislander, and Haas 2002). Since there is no *ex ante* control, this meets our criterion for 2, even though *Länder* have on occasion circumvented the rule by financing public investment via extraordinary budgets (Balassone, Franco, and Zotteri 2003).

Gray cases arise when violation of formal rules becomes routinized. *Estados* in Brazil between 1950 and 1963 provide an example. Their borrowing autonomy was virtually uncontrolled even though the 1946 constitution stipulated that regional borrowing required prior approval by the senate (C 1946, Art. 62). *Estados* routinely circumvented senate approval by resorting to contractual borrowing from foreign or domestic banks (especially state-owned banks), by issuing domestic or foreign bonds, or running up arrears to suppliers and personnel. This became so rooted that we judge the lack of central control to be an institutional feature of regional authority (Rodden 2006). Things changed in 1964 when the military regime shifted control over borrowing from the senate to the executive, which proceeded to enforce the rule of prior approval. At that point *estados* score 1.

Once formal rules are in place we pay attention to them even if regions do not make use of their borrowing authority. Until 2003, *provincias* in Peru could borrow without prior central authorization as long as debt was not used for current expenditures. Except for the big cities of Lima, Arequipa, and Cusco, borrowing was almost non-existent and it continues to be low to this day. The authority of a region to borrow is our target. The conditions under which a region is induced to borrow are something else. Hence, we score *provincias* 2.

We need to assess the extent to which central rules on regional borrowing are intended to be binding. For example, in 1983 the Australian federal

How We Apply the Coding Scheme

government relaxed controls on borrowing provided that states complied with an aggregate borrowing limit determined by the Loan Council, a central body (Craig 1997; Von Hagen et al. 2000). However, compliance was voluntary, and we allocate the maximum score for borrowing autonomy to Australian states and territories until 1995, when central constraints were tightened.

Argentina illustrates how bindingness can be contractual. In 2004 all *provincias* signed a contract with the federal government saying they would adhere to the new Fiscal Responsibility Law setting limits on provincial spending, an annual ceiling on borrowing, and prohibiting borrowing for current expenditure. There is no prior central government oversight. The conditions meet the criteria for a score of 2. The commitment is in the form of a contract, which *provincias* can opt out of with the consent of their legislature. However, until they cancel the contract, they are bound by its terms.¹⁰

Finally, we assess the existence of multiple borrowing channels. Mexican *estados* have this option. The national constitution limits subnational debt to domestic borrowing for productive investment. The federal congress can add conditions, which it did in 1980 by requiring *estado* governments to ensure prior approval in their assemblies. Together these conditions amount to a score of 2. The law also gave *estados* the option to use revenue-sharing funds as collateral for new debt provided that the ministry of finance approved *ex ante*, which would be a score of 1 (Haggard and Webb 2004). Because *estados* continued to have the option of the first borrowing route, we score 2. However, this was closed off in 2000, at which point *estados* could only borrow with *ex ante* approval and score 1.

Representation

Regional authority with respect to representation is the legal capacity of regional actors to select regional office holders. For regional legislators we distinguish direct election in the region from indirect election by subnational office holders. For a regional executive we distinguish selection by the regional assembly from a mixed system of a regional/central dual executive.

The box below summarizes these categories. We need to clarify the concepts of assembly, executive and, in particular, the notion of a dual executive. Among the ten dimensions of the regional authority index (RAI), representation is most easily confounded with the character of the political regime.

¹⁰ All jurisdictions opted in when it was enacted, but one province opted out in 2012 (Córdoba) and another two (Buenos Aires and Santa Fe) had legislative initiatives to do so. Incidentally, this is also how we would code the 2012 Fiscal Compact, which commits Eurozone member states to write a structural balanced budget and debt ceiling in their constitution.

Measurement

However, regional representation is not governed by the national political regime. Authoritative regional assemblies and executives can in principle co-exist with non-democratic national regimes.

ASSEMBLY	
0:	the region has no regional assembly;
1:	the region has an indirectly elected regional assembly;
2:	the region has a directly elected assembly.
EXECUTIVE	
0:	the region has no regional executive or the regional executive is appointed by central government;
1:	the region has a dual executive appointed by central government and the regional assembly;
2:	the region has an executive appointed by a regional assembly or that is directly elected.

We define an assembly as a self-standing institution in which a fixed membership using parliamentary procedures exercises legitimate authority. A regional assembly exercises legitimate authority for a regional jurisdiction. It cannot be a committee or subsidiary body that is a subset of a national assembly. This excludes grand committees composed of Scottish, Welsh, or Northern Irish members of the House of Commons who meet as caucuses to discuss bills affecting their regions.¹¹

We code the predominant principle of representation in regional assemblies. Where some legislators are directly elected and some indirectly elected, we count voting members. Hence, Hungarian regional councils (*Tervezési-statisztikai régiók*) score zero because a majority of their members are central government appointees, while Romanian regional councils (*Regiuni de dezvoltare*) score 1 because subnational appointees predominate and, unlike central appointees, can vote on regional legislation. In Ecuador, provincial councils score 2 from 1950–63 and from 1998–2008 when directly elected members predominate and members elected by *concejos municipales* are a minority. Conversely, Peru's *regiones* (1988–92) score 1 because only a minority (40 percent) is directly elected; the rest are sent by lower tier *provincias* or selected by interest associations.

Indirectly elected assemblies score 1 when the selectors are subnational. In most cases, these selectors are local governments or local government assemblies, but in Belgium until 1995, regional and community councils consisted of national parliamentarians elected for the relevant region (Flanders/

¹¹ However, these grand committees do constitute a modest channel for shared law making, discussed below.

How We Apply the Coding Scheme

Wallonia/Brussels) or community (French/Dutch speaking). From 1972–81, regional councils in France housed nationally elected politicians from the region alongside indirectly elected representatives from subnational governments. From 2008, Ecuadorian provincial councils were comprised of representatives from the *cantones* and rotating presidents of *juntas parroquiales* (parochial boards).

We define an executive as a legitimate authority that puts rules of general applicability into effect, and we assess whether the head of a regional executive is appointed by central government, the regional government, or a dual executive consisting of both the central and regional government.

The intermediate category encompasses cases where both the central and regional appointees have executive authority. Dual executives can take several forms. Some are two headed, with a central government appointee and a regional appointee, directly elected or selected by the regional assembly. Regional and departmental councils in France elect a president who presides over the executive alongside a centrally appointed prefect with *post hoc* oversight. Thai *changwat* have a directly elected regional chair alongside a centrally appointed governor. Some dual executives vest central and regional authority in a single body. In the Netherlands, the *Commissaris van de Koning* is appointed by the central government on nomination by the provincial assembly. This person chairs the provincial council as well as the executive and formally represents central authority in the province. The remaining members of the executive are elected by the provincial assembly. Several Latin American countries have similar arrangements.

Executives in Indonesian *provinsi-provinsi* and *kabupaten-kabupaten* run the gamut of institutional possibilities. In the first ten years after independence, governors and mayors were elected by their respective assemblies and fully accountable to them, scoring 2 on representation. In 1959, governors and mayors became dual local and central representatives, and were no longer accountable to regional assemblies. Nevertheless, they were still elected by regional assemblies, and we assess this as a dual executive. In 1974, governors were appointed by the president, and mayors followed in 1979, reducing the score to zero. The 1999 constitution restored the pre-1959 situation, and from 2005, governors and mayors became directly elected.

Ecuador had a dual executive for the briefest of times, from 1967 and 1971, when presidentially appointed *gobernadores* co-existed with directly elected *prefectos*—each with executive competences (score=1). When the military took over, *prefectos* were appointed (score=0), and from 2008, the *prefecto* became again popularly elected and the role of governor was abolished (score=2). In Canada, provincial heads responsible to regional legislatures direct the executive alongside lieutenant-governors, ceremonial posts that are too marginal to dilute the executive power of the provincial head, so we score 2.

Measurement

Belgium provides a gray case: provinces combine some features of centrally controlled and dual executives. Until 1987, the centrally appointed governor was pre-eminent in the regional executive. The governor's formal approval was required for legislation, and regionally elected executive members could not reverse this. The governor also opened and closed council meetings, determined their length, and could demand to be heard. Moreover, he saw to it that the provincial council and the executive did not break any laws or decide upon matters beyond their competences.

Is this pre-eminence enough to score provincial executive representation as zero? We think not since the six remaining members of the executive were in charge of day-to-day management and served as heads of departments. The regional members of the executive have gained some authority since 1987, but we continue to interpret it as a dual executive. A reform in that year granted the provincial executive shared executive powers with the governor and reduced the governor's role. In 1997 the governor lost voting rights in the executive. "In purely legal terms, the Belgian governor no longer has the real policy power since 1997" (Valcke et al. 2008: 254). However, the governor retains sole responsibility for public order, security, and the police. The governor is undoubtedly the junior partner in policy making, but this is not enough to tip the score to 2.

Finally, we wish to clarify the distinction between the character of the central regime and the authoritative competences of regions. There is no doubt that an authoritarian regime can destroy the autonomy of its constituent jurisdictions. But the effect of authoritarian regimes in the countries we observe varies along the dimensions of the RAI. Authoritarian regimes do not always suspend or abolish regional elections or disempower or replace elected regional governors.

Russia illustrates this. *Subyekty federacii* (federal jurisdictions) score 2 for assembly and zero for executive from 1993 to 1995, 2 and 2 from 1996 to 2004, and 2 and 1 from 2005 to 2010. The first change corresponds to Yeltsin's decision in 1996 to replace appointment of governors from Moscow with popular regional elections. The second change, a drop in executive representation from 2 to 1 in 2005, was Putin's decision to replace direct election with a procedure in which the president proposes a candidate for governor to each regional legislature.

Argentina reveals the scope for variation. The 1955 military coup ousted the national government but left subnational institutions substantially intact (Eaton 2004a: 71). By contrast the *Revolución Argentina* (1966–72) led to the replacement of elected governors by central government appointees who were put in control of provincial legislatures. The dictatorship of 1976–82 had a similarly drastic effect. Provincial assemblies were disbanded and provincial administration was divided among the army, navy, and air force (Eaton

How We Apply the Coding Scheme

2004a: 71, 117–18; Falleti 2010). Regions score 4 on representation in the first authoritarian episode, and zero in the subsequent ones.

The military regime in Brazil (1964–85) lies between these extremes. It maintained direct elections for governors and assemblies before introducing a system in which assemblies chose governors from a central government shortlist (Samuels and Abrucio 2000: 48). Elections were never canceled, but representative authority was restricted. Governors could be replaced by the military regime (and some 25 percent were in 1964 alone), and direct elections for assemblies took place under a new constitutional framework restricting political parties and civil liberties (Samuels and Abrucio 2000: 49). Our scoring reflects the contrasting strategies of the military in Argentina and Brazil: a sharp drop from the maximum to the minimum score on representation in Argentina, and an intermediate score for both assembly and executive in Brazil.

Shared Rule

A regional government may co-determine decision making at the national level. The coding scheme distinguishes five dimensions and two modes of shared rule.

A region may a) participate in making national law through its representation in the national legislature, usually in the upper chamber; b) share executive responsibility with the national government for designing and implementing policy; c) co-determine the distribution of tax revenues in the country; d) co-determine borrowing conditions and public debt management; and e) exercise authority over the constitutional set up.

A region may exercise multilateral shared rule or bilateral shared rule. Under multilateral shared rule the region relates to the central state as part of a standard tier. It is contingent on coordination with other regions in the same tier. Under bilateral shared rule the region relates to the central state directly. It can be exercised by the region acting alone. The criteria for these forms of shared rule are the same for executive, fiscal, and borrowing control, but vary when it comes to law making and constitutional reform. We detail these differences in the sections that follow.

Law Making

The legislative arena in which regions or their governments directly influence national law is usually the upper, or second, chamber. Most upper chambers came to serve as bulwarks against the principle of one citizen, one vote. They

Measurement

were conservative, sometimes reactionary, bodies representing the aristocracy, the church, corporatist groups, or territorial communities with pre-modern roots. Upper houses are in decline. Thirty-six of the eighty-one countries we observe had a bicameral parliament in 2010, whereas forty-three countries had one at the time they enter the dataset. Nineteen of these upper chambers represent territorial communities in 2010.

Multilateral or bilateral law making accounts for a total of two points in our schema. The scoring is additive in units of 0.5. First we establish whether the composition of a national legislature is primarily regional. One possibility is that its principle of representation is territorial rather than population-based. Are regions the unit of popular representation? The other possibility is that regional governments or assemblies themselves designate representatives to the national legislature. These are the first two items in the scoring scheme for law making. Unless one of these criteria is met, a region will score zero on this dimension. Only if one (or both) of these take place, do we need to assess the law making role of regions at the national level.

MULTILATERAL LAW MAKING		BILATERAL LAW MAKING
Regions are the unit of representation in a national legislature.	0.5	The region is a unit of representation in a national legislature.
Regional governments designate representatives in a national legislature.	0.5	The regional government designates representatives in a national legislature.
Regions have majority representation in a national legislature based on regional representation.	0.5	The regional government or its representatives in a national legislature are consulted on national legislation affecting the region.
The legislature based on regional representation has extensive legislative authority.	0.5	The regional government or its representatives in a national legislature have veto power over national legislation affecting the region.

To assess the regional character of a chamber's composition we need to answer three questions: a) are regions represented in the national legislature *qua* regions or in proportion to their population; b) are representatives to the national legislature chosen directly by regional governments or assemblies; and c) what is the regional role in mixed chambers?

The allocation of seats with respect to territory and population is often categorical. Many countries are divided into roughly equal political constituencies based on population or have some system of proportionality based on population. The Colombian and Peruvian (until 1993) upper chambers are elected on the basis of a single national district. Other countries, by contrast, have second chambers based on territorial representation, including Australia,

How We Apply the Coding Scheme

Russia, Switzerland, Brazil, and Indonesia. In order to score 0.5, the territorial principle must bias the population criterion. This excludes Austria where each *Land* receives seats in the upper chamber in relation to its population in accord with the prior population census. The representation of regions in the Italian senate (2010) also falls short. The 315 constituencies of the senate are distributed among the twenty Italian regions in proportion to their population, save for six seats assigned to Italians living overseas and two life-time senators.

There are some judgment calls. The Italian electoral law for the senate mentions the principles of territory and population in the same paragraph: “The Senate of the Republic is elected on the basis of the region. Except for the seats assigned to the Overseas, the seats are divided among the regions in accordance with Article 57 of the Constitution on the basis of the results of the last general census of the population.”¹² However, the allocation of seats reveals that population trumps territory. The smallest *regioni*, Valle d’Aosta and Molise, have just one and two seats, respectively. The other eighteen *regioni* range from seven to forty-nine seats in step with population.

For the regional principle to prevail, seats do not have to be allocated equally across regions. What matters is the principle that is articulated in the constitution and the extent of disproportionality between seats and population. Where the constitutional principle is explicitly territorial this meets the criterion even if regions happen to be represented in rough proportion to their population. A rule of thumb for territorial representation is where the disproportion of seats per voter exceeds 5.0 between the most and least represented regions.

The German *Bundesrat* establishes regions as the unit of representation even though the number of seats per *Land* ranges from three to six. Each *Land* has at least three votes, and most have more in line with a constitutionally mandated population rule that gives four seats to *Länder* with more than two million inhabitants, five seats to *Länder* with more than six million, and six seats to *Länder* with more than seven million. The disproportion of seats to population across *Länder* reaches a whopping 1:13. This compares with less than 1:3 for the Italian senate. Between 1997 and 2006 each Thai *changwat* received between one and four seats in the senate which yields a disproportion of 1:3.5 between the most and least represented region. This is a gray case, but given that the Thai constitution does not articulate the territorial principle, we score *changwat* zero on this item.

Uncertainty can arise from thin information and abstruse legal texts. Haiti provides an illustration. Between 1950 and 1956 senators were directly elected. The constitution provisionally allocated between three and six seats

¹² Law No. 270/2005, Art. 4.

Measurement

to each *département* until a law fixed the number of senators for existing and new *départements* “taking into account the population of certain regions and, especially, their economic and political importance” (C 1950, Art. 40). Policy makers never got around to passing the law since the senate was abolished in 1957 by the Duvalier dictatorship, but elections took place in this period. Is this a region-based or population-based chamber? Given the mildness of disproportionality and the intention of the law maker to allocate seats on population, we score *départements* zero on the first component of law making. In Peru, the 1979 constitution explicitly envisages a senate elected by the *regiones*, but since this provision never came into effect, we score zero.

Direct representation of regional governments or assemblies in a second chamber is an important additional feature of regional authority because it provides institutional access to law making. This is usually clear-cut. Each German *Land* is directly represented in the *Bundesrat* by a representative designated by the *Land* itself. Regional parliaments rather than regional executives are represented in the Malaysian *Dewan Negara*, the Austrian *Bundesrat*, the Dutch *Eerste Kamer*,¹³ the Argentine senate (until 2001), and in part of the Spanish senate. In Russia, each *subyekt federacii* sends a delegate from its legislature and one from its executive to sit in the upper chamber, the *Sovet Federatsii*. Each of these variants scores 0.5.

And, finally, how should mixed chambers be evaluated? We assess regional representation as positive if one or more groups of senators are selected on the principle of regional representation or direct government representation in the chamber. We then go about estimating the extent of authority on this dimension, but we wish to pick up the role of regions in national law-making even when they do not have a majority in the chamber.

Belgium and Malaysia illustrate this. Since 1995, the Belgian senate comprises three kinds of community representatives: forty directly elected senators, twenty-one indirectly elected community senators, plus ten senators selected by these groups. The community senators are selected on the principle of regional representation (the Flemish and Francophone communities each have ten seats with one seat for the tiny German-speaking community) and they serve as delegates of the communities. We score 0.5 on each criterion even though community senators make up less than one-third of the senate. Directly elected and co-opted senators do not meet the second criterion, but arguably meet the first. Both cases are gray: equality of regional representation is finely balanced with “one citizen, one vote.” While the distribution of seats is roughly in line with population, it is fixed on territorial principles in the

¹³ The *Eerste Kamer* is a complex case because provincial representatives vote for candidates on party lists which structure the outcome. We score this 0.5 because the voters are regional representatives.

How We Apply the Coding Scheme

constitution. Contrary to the Austrian or Italian second chambers, seats are not reallocated following a census.

In Malaysia, the senate is composed of regional representatives appointed by *negeri* assemblies, and they meet both criteria. The senate also has federal appointees who meet neither criterion. Initially, *negeri* representatives had a majority but since 1964 they have been outnumbered; by 2010 they controlled just under 40 percent of the seats. We reflect the loss of a collective majority in the third criterion of law making discussed below.

Before evaluating the authoritative character of the second chamber, we need to address bilateral law making. A region, like Åland or Quebec, may be represented as a territory in the upper chamber even when the regions in its tier are not. Unlike regions in the rest of Britain, Scotland, Wales, and Northern Ireland have special caucuses in the House of Commons that convene as grand committees to discuss bills affecting their regions. Senators from Quebec are selected individually by twenty-four electoral districts within the province rather than by nomination of the prime minister. Indigenous populations in Bolivia have reserved delegates in *departamento* representation at the national level.

In contrast, the Portuguese autonomous regions have no bilateral access to law making. Regional representatives from the Azores and Madeira are no different from other Portuguese law makers in the unicameral parliament. Nor do the powerful Malaysian states of Sabah and Sarawak have bilateral shared rule. Like any other *negeri*, the parliaments of Sabah and Sarawak can send two representatives to the upper chamber, which is consulted on national legislation. Sabah and Sarawak representatives can of course weigh in on legislation relevant to their region in general proceedings, but they do not have special rights to be consulted or co-decide.

The Belgian communities are a border case; we code them as having multilateral but not bilateral law making. Multilateral law making takes place through elected and appointed representatives in the senate. There are, then, no special provisions for particular communities or regions to influence ordinary legislation affecting their territory.¹⁴ Fiscal legislation and constitutional reform require majorities of each community, but not ordinary legislation.¹⁵

We assess the extent of regional authority in shared law making for regions that are represented as territories or have institutional representation in the upper chamber. The criteria are different for multilateral law making and for bilateral law making, and we discuss them separately.

¹⁴ A partial exception is the alarm bell procedure, introduced in the 1970 constitution, which enables one language group to postpone legislation for thirty days with a three-quarters majority. Its conditions of use are highly restrictive and it has only been invoked twice since 1970.

¹⁵ We consider shared rule in fiscal policy and constitutional reform as distinct dimensions, discussed later.

Measurement

An additional half point goes to regions if their representatives constitute a majority of the chamber. The unit of analysis here is the tier or, for bilateral shared rule, the individual region. Legislatures in which regional representatives constitute a majority include the US senate, the Argentine senate, the German *Bundesrat*, the Dutch *Eerste Kamer*, the Haitian senate, and the *Dewan Perwakilan Daerah* in Indonesia. Belgian provinces, which, until 1995, were allocated one-third of the seats in the senate, fall short, as do *comunidades autónomas* in Spain, and *negeri* in Malaysia. Ecuador's pre-1978 senate meets the criterion because provincial senators outnumbered the *Senadores funcionales* who were elected by corporatist associations. In some countries, such as Bolivia, just a small number of seats are reserved for particular regions, in this case, regions with indigenous communities.

A further half point is scored if a legislature with regional representation can veto ordinary legislation or if its amendments can be overridden only by a supermajority in the other chamber. The Austrian *Bundesrat* scores zero because it can be overridden by a simple majority in the lower chamber, as can the *Županijski dom* (chamber of counties) in Croatia, which, until it was abolished in 2001, was a consultative chamber.

A legislature is judged to have extensive authority if it can veto ordinary legislation or if a supermajority in the other chamber is needed to override its veto. This applies even if the veto powers of the legislature are restricted to a subset of policies as long as these are recognized to be central to the body politic. The Belgian senate scores 0.5 on this criterion. Since the 1995 reform, the senate is conceived as a *reflectiekamer* (reflection chamber) with limited authority over ordinary legislation and none over the budget. However, it exercises equal legislative powers with the lower chamber on freedom of religion, language use, the judicial system, international treaties, and constitutional change, subjects that are close to the heart of the body politic (Deschouwer 2012; Hooghe 2004; Swenden 2006).

We must customize this criterion to tap bilateral law making. What matters here is how a region is involved in law making. A region receives a score of 0.5 if its representatives or government must be consulted on legislation affecting the region and an additional 0.5 if either can veto a legislative proposal.¹⁶

For example, the 1982 reforms gave the Corsican assembly the right to be consulted by the French government on all matters concerning Corsica. Non-binding consultation is also the rule for the Azores and Madeira. Their statutes

¹⁶ In principle a differentiated region can combine authority over multilateral and bilateral law making. In practice this appears to be extremely uncommon. There is only one instance in our dataset: Montenegro and Serbia in the Serbia–Montenegrin confederation between 2003 and 2006. We use the larger of the total scores for multilateral and bilateral law making in aggregating the score for a region.

How We Apply the Coding Scheme

specify that the Portuguese parliament is constitutionally bound to consult the regional assemblies, and each regional assembly can submit amendments or legislative drafts on taxation, environmental policy, criminal law, law and order, regional planning, and social security.

The Korean island of Jeju is a gray case that illustrates the lower bound for bilateral shared rule. Jeju does not have special representation in the legislature, but the governor “may present his/her opinion on any matter he/she considers necessary to deliberate on legislation concerning the Province upon obtaining consent from two-thirds of the incumbent Provincial Council Members” (2006 Special Act, Art. 9.1). These views are then presented to a “Supporting Committee,” a thirty-member body comprised of heads of central government departments and chaired by the prime minister, which negotiates on behalf of Jeju. Hence the Jeju government has a right to put legislative proposals on the agenda but it is held at arm’s length from the negotiations. Still, the right is legally embedded. We score bilateral shared rule only if it has a legal basis in the constitution, the statute, a law, or an executive decree.¹⁷

There are just five cases in the dataset where an individual region has formal veto rights over national legislation affecting its territory: Montenegro and Serbia in the former Yugoslav confederation (2003–06), and the special regions of Northern Ireland (since 2000) and Scotland and Wales (since 1999).

Montenegro and Serbia had a veto because ordinary legislation required a double majority: a majority of representatives of each republic and an overall absolute majority. Note the difference with Belgium, where only laws concerned with the fiscal framework and constitutional change require a majority in both large language groups.

The three UK regions have a veto over national legislation pertaining to their region on account of the Sewel convention which states that the “UK Parliament would not normally legislate with regard to devolved matters except with the agreement of the devolved legislature” (Cairney 2006; Devolution Guidance Notes Nos. 8–10 2014). The Sewel convention was written into a memorandum of understanding between the UK and its regional parliaments in 1999 (Memorandum of Understanding 2002 paragraph 13; 2013 paragraph 14).

It is interesting that no other autonomous region has veto power over ordinary legislation. Greenland, the Farøer islands, and the Åland islands narrowly miss. The governments of Greenland and the Farøer islands are required to be consulted on all national bills, administrative orders, and statutes of importance to them before the legislation can be put before the

¹⁷ Two other regions can propose (or oppose) legislation in the national parliament: Vojvodina in Serbia, and London in the United Kingdom.

Measurement

Danish parliament. In case of disagreement, the question is tabled before a board consisting of two members nominated by the Danish government, two members nominated by the home-rule authorities, and three judges nominated by the Danish highest court. This falls short of giving the islands a veto. Similarly, the Åland government must be consulted by the Finnish parliament on any act of special importance to the islands, but national legislation is not conditional upon its assent. The Åland government also participates in EU decision making for matters within its powers, and the parliament of Åland must give its consent to international treaties in areas under its competence.

Executive Control

Regional governments may share executive authority with the central government in the context of intergovernmental meetings. To score on this dimension, such meetings must be routinized, not ad hoc. To score the maximum two points, such meetings must be authoritative, i.e. reach decisions that formally bind the participants. The criteria are the same for bilateral and multilateral executive control.

EXECUTIVE CONTROL

- 0: no routine meetings between the central government and the regional government(s) to negotiate national policy affecting the region;
- 1: routine meetings between the central government and the regional government(s) without legally binding authority;
- 2: routine meetings between the central government and the regional government(s) with legally binding authority.

The distinctions on this dimension are illustrated in the history of German intergovernmental relations from the early days of the Federal Republic (Benz 1999; Scharpf, Reissert, and Schnabel 1976). In 1947, a first consultative meeting was held between *Land* premiers (*Ministerpräsidenten*) and the federal chancellor, but it was one-off. In 1954, the *Ministerpräsidentenkonferenz*, which combines all *Land* presidents, became a standing, but still consultative, meeting. It scores 1 in our schema. In 1964, the two government levels agreed to negotiate on joint policy tasks in routine, binding intergovernmental meetings. In 1969, these were anchored in a revision of the Basic Law concerning joint federal-*Länder* tasks. In most meetings unanimity is the rule but some can make majoritarian binding decisions (with thirteen of sixteen *Länder*), scoring 2.

Executive control in Germany from 1969 fully meets the criteria for a maximum score. Meetings between regional and central governments are highly institutionalized, general purpose in policy scope, and produce legally

How We Apply the Coding Scheme

binding agreements. Few cases are so clear-cut. We draw on several examples to explicate how we adjudicate ambiguities, in particular when the rules differ across policy. We conclude with examples of bilateral executive control.

Executive shared rule involves routinized negotiation among regional and central governments. There are several requirements for a positive score. Both central and regional governments—not professional or sectoral groups—must be involved. Regional governments must be able to select their own representatives. Negotiation must be institutionalized. The framework must be general purpose governance. Let us engage each in turn.

The system of *conselhos* in Brazil illustrates that the criterion of government involvement is not always black and white. Since the early twentieth century, *conselhos* composed of professional groups have existed in health and education, but with tenuous connections to *estado* governments. We therefore score the *estados* zero for the first decades. In 1990 a routinized system of multilevel governmental *conselhos* emerged. Local *conselhos* are represented in *estado conselhos*, which are in turn represented in a nation-wide *conselho* (Pogrebischini and Santos 2009). While the *conselhos* convene societal users and providers, they are led by government representatives. The system is most developed in the health sector, but is also present in education, transport, and other areas. We score *estados* 1 from 1990.

Executive power sharing must be vertical, that is, it must include both regional and national government. Horizontal coordination among regions does not amount to shared national control of policy making. Intergovernmental coordination in Switzerland is instructive. This chiefly takes the form of inter-cantonal *concordats*, which often lead to binding agreements among cantons, but rarely include the federal government (Blatter 2010; Sciarini 2005). However, from 1978 vertical cantonal–federal coordination was organized through the *Kontaktgremium Bund-Kantone*; and this was replaced in 1997 with the twice-yearly *Föderalistischer Dialog* (federal dialogue). A constitutional revision of 2008 opened the door to binding, not just voluntary, cooperation. Article 48a of the constitution authorizes the confederation to declare inter-cantonal agreements binding or require cantons to participate in inter-cantonal agreements in nine constitutionally defined domains, including tertiary education, urban public transport, and waste processing. The confederation can initiate binding cooperation only at the request of the cantons. The reform facilitates inter-cantonal conventions with federal involvement and the equalization of burdens among cantons (Cappelletti, Fischer, and Sciarini 2014). Cantons score 1 on executive control until 2007 at which point they score 2.

Mexico provides a gray case which we score zero because the vertical component is weak. Since 1999 Mexican governors have held meetings to discuss decentralization in health and education. These became formalized as a

Measurement

standing *conferencia nacional de gobernadores* (national conference of governors, CONAGO) with regular meetings, a permanent secretariat, and executive committees. Although federal representatives sometimes attend, CONAGO meetings are inter-state (Falleti 2010). CONAGO has brokered a few binding agreements, but there is no formal role for the federal government.

Executive control may enhance regional authority only if regional governments can select their own representatives. Colombia is a negative example. *Departamentos* have been consulted since 1991 on economic development through a standing body, the National Planning Council. However, the five members representing the *departamentos* are selected by the president from a list of governors submitted by the *departamentos*.

A positive score requires that executive control is routinized on a legal basis. Since the 1990s, Mexican *estados* have organized occasional informal meetings to put pressure on the federal government. Such meetings led to health care decentralization in 1996. However, none of these initiatives has thus far generated a routinized system that encompasses both *estados* and the federal government (Jordana 2001; Falleti 2010). This is a fairly clear example. Italy provides a gray case. Intergovernmental conferences between the central government and *regioni* took place in 1983, 1984, and 1985, with none the following year. In 1987 the constitutional court ruled that the principle of “fair cooperation” should guide regional–national relations, which prompted a 1988 law creating a standing conference on state–regional relations with routinized bi-annual meetings (Ceccherini 2009). We score 1 from the time of the first meeting in 1989.

Consistent with our focus on general purpose rather than task-specific governance, executive control must cover significant policies to warrant a positive score. At the margin are a handful of cases where we score executive control with limited policy coverage, but where the policies are central to the authority of regional governments. Argentina illustrates this. Executive coordination was virtually non-existent in the 1950s and 1960s. In 1972 the *Consejo Federal de Educación* (Federal Council of Education) was created to coordinate provincial and federal educational policy (Falleti 2010). Meetings between *provincias* and the central government were routinized and took place at least once a year, but their scope was narrow. In 1979 the council was expanded to include culture, at which point we score 1. When its decisions became legally binding in 2006, *provincias* score 2.

Coordination can be binding (score=2) or non-binding (score=1). Where there are multiple meetings with different decision rules, we score the predominant pattern. Malaysian federalism is characterized by numerous national councils that interweave state and federal policy making on a broad range of issues, and only two of these produce legally binding decisions: the National Land Council and, since 1986, the National Council on Local

How We Apply the Coding Scheme

Government (Loh 2010). The councils make binding nation-wide policy on matters that are constitutionally reserved for the *negeri*, so shared rule counterbalances federal usurpation of *negeri* self-rule. Though land use and local government have gained importance, they are much less central to *negeri* authority than the regulation of religious and cultural life, which remains largely in the realm of non-binding coordination (Harper 1999; Reid 2010b). A score of 1 reflects the predominance of non-binding executive shared rule.

Finally, it is useful to compare bilateral and multilateral executive shared rule. To score 1 or 2 either type must be routinized, general purpose, and government-dominated. And for a provision to receive a positive score, it needs to be in operation. The difference between bilateral and multilateral is whether the meetings with the central government involve a single region or all regions in a particular tier.

The five indigenous *comarcas* in Panama have bilateral meetings that arrive at binding decisions with the central government in the *Consejo Nacional de Desarrollo Indígena* (National Council on Indigenous Development). Panama's *provincias* are not involved. The two indigenous *regiones autónomas* in Nicaragua are consulted on, and can veto, national executive decisions on natural resources and communal land. Several autonomous regions have non-binding bilateral control, including the Åland islands, Greenland and the Farøer islands, Scotland, Wales, and Northern Ireland, and Azores and Madeira.

Bilateral executive shared rule normally has a legal basis in statute, the constitution, a special law, or executive order. It is, for example, explicitly set out in the Åland Act, the Greenland Act, and the special statutes for Azores and Madeira, all of which are enforceable in court. Despite the informality of its constitution, the United Kingdom is no exception. The devolution acts mandate statutory consultation by the British government.

The US states have an unusual form of bilateral executive control, which we assess to be binding. States can opt to accept or reject regulations or programs that the federal government offers within concurrent policy areas such as health, environment, or transport (Bakvis and Brown 2010). The implementation of many national laws in these areas hinges on one-to-one bilateral agreements with state governments. While there is no particular passage in the constitution, law, or executive order that regulates these meetings, the legal basis for the right to be consulted on (and veto) the implementation of many federal policies lies in the Commerce Clause, the Fifth and the Fourteenth Amendment, and in Supreme Court jurisprudence (Christensen and Wise 2009; Wright 1988).

Mexican *estados*, Aceh, and the Spanish *comunidades* illustrate the distinction between routinized and ad hoc consultation. The predominant mode of coordination in Mexico has been ad hoc bilateral agreements between the federal government and an *estado* (Jordana 2001). There is no formal legal

basis and we score zero. In Indonesia, the 2006 Aceh statute promised non-binding consultation on law making, administrative policy, and international relations (Art. 8). Once this was implemented in a presidential decree of 2008, Aceh scores 1 for bilateral executive control.¹⁸

Finally, Spain demonstrates complex interplay between bilateral and multi-lateral shared rule. In the first decade after democratic transition, bilateral negotiations between the national government and individual *comunidades autónomas* predominated (Bolleyer and Thorlakson 2012). These lacked predictability and structure, which translates into a score of zero. Since the 1980s, Spain has shifted to a multilateral frame including routinized conferences producing binding decisions in health and European affairs which sustain a score of 2. Routinized bilateral shared rule is limited to taxation policy for the Basque Country (and its provinces) and Navarre.

Fiscal Control

Shared rule on taxation is a special case of legislative and executive shared rule. Scoring fiscal control requires a few ground rules. First, we conceptualize fiscal policy as distinct from executive policy or borrowing policy. Second, we identify the institutional framework for fiscal control. Regional influence on fiscal policy may employ one of two institutional routes: a *Bundesrat*-type chamber composed of regional government representatives or a routinized intergovernmental forum. Third, we explain what happens when both routes are present. We conclude with a brief discussion of bilateral fiscal control.

We assess regional fiscal shared rule as the role of regional governments in legislation or executive regulation regarding the collection and allocation of taxes. The collection and allocation of taxes includes distribution keys, tax rates, tax bases, intergovernmental transfers, grants, and annual or multi-annual central budgets. We assess regional debt management and borrowing in a separate dimension. To qualify as shared rule, coordination must be encompassing; it cannot be limited to consultation on a particular fund or grant. For example, Uruguayan *departamentos* score 1 on fiscal control because they are consulted on the percentage of tax revenue to be shared—not because they provide input on how to spend some 25 percent of the *Fondo de Desarrollo del Interior* (Fund for the Development of the Interior).

Two routes are available for regional governments to influence the generation and distribution of national tax revenues. The executive route provides direct access via intergovernmental meetings. The legislative route gives indirect access through a national chamber with regional representation. If

¹⁸ The 2006 legislation included bilateral law making, but this was excluded from the 2008 presidential decree, and Aceh scores zero on this dimension (Ahtisaari 2012; Suksi 2011: 363–5).

How We Apply the Coding Scheme

regional governments negotiate over the distribution of tax revenues via either channel they score 1; if they have a veto, they score 2. The box below summarizes the three alternatives.

FISCAL CONTROL

- 0: neither the regional government(s) nor their representatives in the national legislature are consulted over the distribution of tax revenues;
- 1: the regional government(s) or their representatives in the national legislature negotiate over the distribution of tax revenues, but do not have a veto;
- 2: the regional government(s) or their representatives in the national legislature have a veto over the distribution of tax revenues.

To score 1 via the legislative route, the legislature must have authority over the distribution of tax revenues. If the representatives of regional governments constitute a *majority* in a legislature and the legislature has a *veto* on the distribution of tax revenues, this scores 2. This avenue requires that regional governments (not their populations through the ballot box) send representatives to the legislature.

Dutch *provincies* and Swedish *landstinge* (until the abolition of the upper chamber in 1971) meet the conditions for a score of 2: they form or formed a majority in the upper chamber with that chamber having a veto on tax revenue allocation. Spanish *comunidades* score 1 both because they are a minority in an upper chamber and because that chamber can be overridden by a majority in the lower chamber. Belgian provinces were (until 1995) represented in an upper chamber with a tax veto, but they never constituted a majority and also score 1. However, Belgian communities (1970–95) and regions (1980–95) did have a majority in the senate by virtue of their institutional representation through the so-called double mandate. Senators wore two hats in addition to their national mandate: as members of a community council (linguistic affiliation) and of a regional council (residence-based). Since the senate could veto financial regulations, communities and regions score 2. Since 1995, community senators constitute a minority and can influence but not block fiscal decisions.

To score 1 via the executive route, regional governments must be directly involved in negotiation and to score 2, they must be able to exercise a veto. Such involvement could, in principle, be exercised through a peak association if that association could bind its members, but this is rare. Denmark and Sweden provide gray cases. Peak associations of regional and local governments meet with the central government, but we score zero for fiscal control because these associations are best seen as lobby groups rather than negotiators. Similarly, the Ecuadorian *Comisión Nacional de Descentralización y Organización Territorial* (National Commission on Decentralization and Territorial

Measurement

Organization, CONADE) does not provide regions with direct involvement. The eight-member committee is headed by a government official, and includes representatives of the municipal and provincial associations alongside sectoral and legislative representatives. The peak organizations cannot legally commit their members and the national parliament reserves the right to take unilateral action.

Uruguayan *departamentos* meet the criteria for a score of 1 through their participation in the *Comisión Sectorial de Descentralización* (Sectoral Commission on Decentralization, COSEDE), which advises the national government on the percentage of revenue to be shared. The *Comisión* is composed of representatives of national and regional governments and makes non-binding recommendations (Eaton 2004a).

Some regions have access to both the legislative and the executive routes, in which case we count the route that produces the highest score. Until 2001, Argentine *provincias* could operate along both routes, barring authoritarian periods. A senate composed of provincial delegates wielded a veto over taxation and intergovernmental grants, which we score 2. When direct elections for the senate replaced institutional representation of *provincias* in 2001, the score for the legislative route becomes zero. However, *provincias* also had access to an institutionalized system of regular intergovernmental negotiations, formalized in a 1951 law, which produced binding *co-participación* agreements on national revenue sharing. Both this system and the 1994 constitutionalized arrangement of binding *co-participación* agreements with a provincial veto score 2.

We conclude by emphasizing the criterion of routinization. We assess Brazilian *estados* to have neither multilateral nor bilateral fiscal control. There is no standing collective body in which *estados* and federal government convene to discuss fiscal policy and, since the senate is composed of directly elected senators rather than regional government delegates, there is also no legislative route. Moreover, no *estado* has legally protected bilateral fiscal control. This induces *estados* to engage in bilateral deals with the federal government in time of need, but these deals typically provide one-off transfers, and we score them zero (Diaz-Cayeros 2006; Dillinger and Webb 1999a; Rodden 2004).

Borrowing Control

Shared rule on borrowing is a special case of executive control. The scoring rules are parallel: we assess the representation of regions in meetings with the central government, the extent to which they are institutionalized, and the extent to which they make binding decisions. Here, however, we are

How We Apply the Coding Scheme

concerned with the national regulatory framework on government borrowing and debt management. The box below lays out the categories.

We begin by outlining the distinction between borrowing control and fiscal control on the one hand, and between borrowing control and constitutional reform or law making, on the other. We then explain how we operationalize bilateral borrowing control.

BORROWING CONTROL

- 0: regional government(s) are not routinely consulted over borrowing constraints;
- 1: regional government(s) negotiate routinely over borrowing constraints but do not have a veto;
- 2: regional government(s) negotiate routinely over borrowing constraints and have a veto.

This dimension encompasses subnational and national borrowing or debt control. It considers fiscal policy only to the extent that fiscal decisions affect borrowing, and does not include raising or spending taxes. There is minimal overlap between this dimension and others that we assess independently. Hence the association between regional authority in borrowing and in fiscal policy can be investigated empirically.

Drawing the line between fiscal and borrowing policy can be tricky. Rules that constrain spending or revenues are technically within the remit of fiscal policy, but they can affect debt levels (Schaechter et al. 2012). Our approach is to examine the authoritative connection between routinized coordination on fiscal rules and subnational borrowing. We begin with two clear, but contrasting cases: Australia and Argentina. Australia's Loan Council is the venue for routinized coordination on fiscal as well as borrowing policy. It is composed of one federal representative and one representative of each state. It approves state borrowing and determines, with the consent of the states, the amount of borrowing, and the interest rate. Its second role is to advise the premiers' conference on fiscal matters. We score 2 points on borrowing and 1 on fiscal control.¹⁹

In contrast, Argentina has separate intergovernmental fora: the *Comisión Federal de Impuestos* (Federal Tax Commission), a long-standing body, deals only with taxation and intergovernmental transfers, while the *Consejo Federal de Responsabilidad Fiscal* (Federal Fiscal Responsibility Commission), created in 2004, monitors budgetary transparency and borrowing. Both consist of federal and provincial governmental representatives, but while the former has binding authority based on regional agreement, and scores 2, the latter does not, and scores 1.

¹⁹ Until 1999, when the score for fiscal control becomes 2 following the creation of a Ministerial Council for Commonwealth–State Financial Relations.

Measurement

Austria illustrates how fiscal rule making can affect borrowing authority, and is best coded as borrowing control. In an effort to meet the Maastricht stability criteria for the European Monetary Union, in 1999 all levels of government agreed to a domestic stability pact with far-reaching fiscal targets. *Länder* as a group must achieve an annual budgetary surplus of 0.75 percent of gross domestic product (GDP). Each *Land* is given a target but can transfer surplus or deficit rights to other *Länder*, and sanctions are applied in case of non-compliance (Balassone, Franco, and Zotteri 2003; Joumard and Kongsrud 2003). A commission composed of *Bund-Land* municipality representatives takes decisions by unanimity. While the pact does not address borrowing constraints directly, the intended effect was to impose collective binding control over *Land* (and *Bund*) borrowing. We score 2 on borrowing control.

Contrast this with Bolivia. The *Consejo Nacional para las Autonomías y la Descentralización* (National Council on Autonomy and Decentralization) is a forum for the national government, *departamentos*, municipalities, indigenous communities, and autonomous regions. It meets twice a year to advise on, among other things, fiscal policy, but congress remains the venue for borrowing policy (Frank 2010).

The overlap between borrowing and constitutional reform or law making is minimized by focusing on the intergovernmental arena. It is not uncommon for constitutions to have provisions on subnational borrowing. The authority of regions to influence these rules is assessed under constitutional reform. Similarly, since the 1990s, several countries have passed fiscal responsibility laws with the aim of constraining subnational borrowing (Liu and Webb 2011). We code these under borrowing control only if they are accompanied by an institutionalized intergovernmental forum that monitors, regulates, or sanctions. Otherwise this falls under law making.

Early examples of institutionalized intergovernmental coordination are the Australian Loan Council, regulating multilevel borrowing since 1923, and the Malaysian National Finance Council set up in 1957 to advise on “the annual loan requirements of the Federation and the States and the exercise by the Federation and the States of their borrowing powers; the making of loans to any of the States” (C 1957, Art. 108). The German *Finanzplanungsrat*, created in 1968 to coordinate federal and subnational budgetary planning, is another early example, though it became binding with respect to *Länder* borrowing in from 2010.²⁰

Subnational borrowing was on the backburner until the debt crises of the 1980s and 1990s (Rodden 2002: 670). In 1989 Belgium reformed its *Hoge Raad van Financiën* into a body with equal federal–community representation

²⁰ It was renamed the *Stabilitätsrat* (Stability Council) in 2010.

How We Apply the Coding Scheme

and tasked it to advise on subnational and national borrowing. In 1980 Spain created the *Consejo de Política Fiscal y Financiera* (Fiscal and Financial Policy Council) composed of national and *comunidad* finance ministers to coordinate fiscal policy and, since 1992, set deficit and debt ceilings. In 1999 Austria established a committee with equal *Land* and federal representation with the authority to fine *Länder* that violate budget and borrowing targets.

The criteria for bilateral borrowing control are the same as for multilateral borrowing. Hence, a positive score requires evidence of institutionalization. In our dataset we detect only one instance, the Argentine *provincias*, including Buenos Aires. This case epitomizes the gray zone between bilateral and multilateral shared rule. In 2004, congress passed a fiscal responsibility law which in principle applies to provincial as well as the national government, and created a federal council for fiscal responsibility composed of the national and provincial ministries of finance. The law has a covenant format, i.e. provincial governments must actively consent one by one for it to be binding. There is, then, no collective contract, though initially twenty-one of twenty-four *provincias* and the city of Buenos Aires signed up. For those who sign up, the law creates a routinized system for intergovernmental coordination and monitoring on budgets and borrowing (Liu and Webb 2011). We code this as bilateral rule because individual *provincias* retain the right to withdraw at any time, though the *modus operandi* is multilateral.

Constitutional Reform

Constitutional authority is fundamental for it concerns the rules of the game. Subnational control over the constitution is often seen as the defining characteristic of federalism (e.g. Riker 1964). Here we suspend this assumption and explore how the constitutional role of regions can be estimated in non-federal and federal countries.

The coding scheme attaches greater weight to regional governments (or their representatives in the legislature) than to other regional actors (i.e. electorates or regionally elected representatives), and it rates binding authority (i.e. veto power) as more authoritative than non-binding involvement. For multilateral control over constitutional reform the schema is as follows: a score of 1 if regional electorates or their representatives can raise the hurdle for constitutional change; 2 if regional governments can raise the barrier for constitutional change; 3 if regional electorates or their representatives can veto constitutional change; and 4 if regional governments can veto constitutional change. The box below details this. Since bilateral constitutional reform requires different criteria, it will be discussed separately.

MULTILATERAL CONSTITUTIONAL REFORM

- 0: the central government or national electorate can unilaterally change the constitution;
- 1: a national legislature based on regional representation can propose or postpone constitutional reform, raise the decision hurdle in the other chamber, require a second vote in the other chamber, or require a popular referendum;
- 2: regional governments or their representatives in a national legislature propose or postpone constitutional reform, raise the decision hurdle in the other chamber, require a second vote in the other chamber, or require a popular referendum;
- 3: a legislature based on regional representation can veto constitutional change; or constitutional change requires a referendum based on the principle of equal regional representation;
- 4: regional governments or their representatives in a national legislature can veto constitutional change.

Scoring multilateral constitutional reform poses several challenges. Under what circumstances does it make sense to say regional intervention raises the hurdle for central actors to pass reform? What is an appropriate floor for scoring regional authority in constitutional reform. What is an appropriate ceiling? Finally, we discuss scoring rules for four sources of ambiguity arising where regions have more than one option for constitutional shared rule, where constitutions have more than one amendment procedure, where constitutional reform is unwritten, and where formal rules and political practice diverge.

We score zero when regional actors or regional governments cannot legally veto or raise the hurdle for constitutional reform. Being consulted or having the right to propose reforms is not sufficient to score 1. For example, until 2001 the Croatian upper chamber, composed of *županija*-appointed representatives, was consulted on constitutional reform but could not amend or raise the hurdle.

A non-blocking minority is insufficient. In Spain, *comunidad*-appointed senators make up less than 20 percent of the senate, too few to block constitutional reform or raise the hurdle in the other chamber, and therefore score zero. Directly elected senators from Spanish *provincias*, by contrast, can veto constitutional bills and consequently score 3. Since the reorganization of the Belgian senate in 1995, the twenty-one senators elected from community parliaments make up 30 percent of the senate and cannot raise the hurdle or veto constitutional reform, which requires a two-thirds majority in both chambers. Belgian communities/regions do not have the institutional representation to warrant a positive score. However, there are also forty popularly elected senators from Belgian communities and regions. Hence a legislature based on regional representation can veto constitutional change and the communities and regions score 3.

The criterion for a regional veto depends on the rules of a chamber in which constitutional reform is decided. For example, *negeri* currently occupy

How We Apply the Coding Scheme

37 percent of the seats in the Malaysian senate, but this is sufficient to score 4 because reform requires a two-thirds majority.

Regions can raise the hurdle for central actors to change the constitution in several ways. They may be able to require a referendum, force a second vote, change the voting rule in the other chamber, or postpone reform. We score 1 when a legislature with regional representation (via the regional electorate) is involved, and 2 when regional governments act through their delegates. In Australia, the Northern Territory and the Canberra Capital Region score 1 between 1975 and 1977 because they had elected representatives in a chamber with regional representation (the senate) which could raise the hurdle, but not veto, a reform of the Australian constitution.²¹ Until 1984, Austrian *Länder* score 2 because they had, through their delegates in the federal council, the power to delay: they could demand a second vote in the first chamber or require a national referendum.

It is useful to specify the floor for a score of 1 or 2. Minimally, this requires that regional intervention is part of a legal process in which regional proposals must be discussed in a parliamentary committee, debated in plenary session, or formally considered by the central government. Portugal provides a clear example. The regional assemblies of Madeira and Azores must initiate the process of revising their statute (C 1976, Art. 228). If the national assembly amends the draft, it is sent back to the regional assembly for consultation. However, the final word lies with the Portuguese parliament. Hence they score 2.

To contribute to regional shared rule, referenda must be regional, that is, preferences are aggregated on the principle of regional, not individual, representation. This is the case in Switzerland and Australia, where constitutional reform requires a double majority in a referendum—a majority of voters in a majority of regions as well as in the country as a whole. This is not so in the Philippines, Ireland, South Korea, Bolivia, Colombia, Peru, or Venezuela, where constitutional amendments require approval by a nation-wide referendum without a regional hurdle.

Scores of 3 or 4 require the authority to veto. We conceptualize the maximum score for the constitutional role of regions in terms of the veto rather than their positive capacity to impose their will on the central government because this would be an almost empty category. The one case that arguably meets the bar of regional imposition was the short-lived confederation of Serbia-Montenegro (2003–06). Constitutional change required the consent of both

²¹ A negative vote in the senate triggers a reflection period of three months. Thereafter, an amendment can pass over the objections of the senate if it obtains an absolute majority in the lower house followed by a referendum in which a majority of states and a majority of the Australian electorate endorse the reform. Until 1978, residents of the Northern Territory and Canberra could not participate in such a referendum, and could influence constitutional change only through the senate.

Measurement

republics' legislatures in addition to a double majority in the unicameral legislature: a majority of representatives of each republic, and an overall absolute majority. Since the unicameral parliament was regional—all its members were delegates from the republics' assemblies—one can argue that the constituent republics could indeed impose constitutional change upon a very weak center. That included the right of each republic to secede unilaterally, which Montenegro promptly did in 2006.

Few other cases are gray. Australian states cannot impose constitutional reform upon the central government. They have an indirect veto over unilateral federal imposition because their representatives can require a binding referendum based on the principle of equal regional representation, and therefore score 3. Similarly, Malaysian *negeri* are in no position to reform the constitution by themselves, which requires approval by two-thirds of the members of each chamber, but they can collectively block amendments—just. *Negeri* representatives now make up twenty-six of the seventy seats in the upper chamber, which gives them two seats to spare for a collective veto. *Negeri* score 4 on constitutional reform. Mexican *estados* cannot initiate reform, which requires a two-thirds majority in the congress. However, they can block because amendments require approval by a majority of *estado* legislatures. They also score 4.

There are several possible sources of ambiguity. First, more than one option for constitutional shared rule may apply. The simple rule is to take the highest score. In Australia constitutional amendments require absolute majorities in both chambers of parliament and then must pass referenda in a majority of states/territories while obtaining an overall majority of the Australian electorate. If there is disagreement between the house and the senate, the objections of the senate can be overridden provided the amendment passes the house by absolute majority after a reflection period of at least three months and after it passes a national referendum. So there are three options: raising the hurdle by requiring a three-month cooling-off period *and* a regional referendum (=1); veto via a regional referendum after both houses pass the amendment (=3); veto via a regional referendum after the lower house passes the amendment (=3). We take the higher score.

Along similar lines, a declaration to reform the Haitian constitution must be approved by two-thirds of each national legislature. Revisions require final approval of at least two-thirds of the national assembly (C 1987, Arts. 281.1–282). The ratio of senators to deputies has changed over time. Until 2000, senators made up more than a third of the national assembly, and hence could block constitutional change. In the 2010 parliament, this is no longer the case (thirty of ninety-nine MPs), but since senate consent is required to initiate constitutional reform (first step of the process), we continue to code the senate as having veto power over constitutional change, giving Haitian regions a score of 3.

How We Apply the Coding Scheme

A second source of ambiguity is that constitutions may contain more than one amendment procedure. One might be for partial constitutional reform and the other for comprehensive reform. These may engage regional actors differently. Our rule is to score the most authoritative route. Bolivia and Venezuela provide examples. Until 2002, the Bolivian constitution regulated only partial reform, that is, reform that did not involve the fundamental principles and rights in the constitution. Such reform required a two-thirds majority in the senate, giving *departamentos* a veto. A revision in 2002 inserted a path for comprehensive constitutional reform which bypasses the senate in favor of a two-thirds vote in the combined congress. The senate contributes just twenty-seven of 157 seats in this congress and *departamentos* are unable to propose or postpone reform. Because they retained their veto role in one of the two procedures for constitutional reform, we continue to score *departamentos* 3 until a 2009 reform eliminated the partial reform process.

Venezuela had two tracks with separate rules until 1999. Partial reform required a positive vote in two-thirds of the *estado* assemblies, while comprehensive reform required a majority in the senate and ratification by national referendum. So the former route produces a score of 4, and the latter a score of 3. We take the highest score. Under the 1999 constitution, reform requires a two-thirds majority in the combined assembly (where senators hold less than one-third of the seats) and a simple majority in a nation-wide referendum, neither of which give the *estados* traction in proposing or postponing reform.

Constitutional norms may be unwritten or dispersed across written documents as in Britain and some of its former colonies. Canada provides an instructive example. Until 1982, the ultimate authority for constitutional change in Canada was vested in the British parliament with the formal understanding (recognized in the 1949 British North America Act) that reform would be proposed by the parliament of Canada. There was also a precedent from 1940 that amendments would need the consent of at least a majority of provinces. When in 1980, Prime Minister Pierre Trudeau sought to patriate the constitution without provincial consent, several provinces objected. In the famous patriation reference of 1981 (SCR 753), the Canadian Supreme Court ruled that federal unilateralism, though legal in a narrow sense, violated a constitutional convention. This persuaded the federal government to negotiate the consent of nine of the ten provinces. The 1982 Canadian constitution consolidated the precedent of Article 38 which states that most amendments require the consent of at least two-thirds of the provincial legislatures representing at least 50 percent of the population.²² Hence we score 4 from 1950, even though the legal status of a collective provincial veto was clarified only in 1982.

²² The consent of Quebec is not legally necessary, although Quebec, along with other provinces, can veto constitutional change regarding English and French language use.

Measurement

Finally, ambiguity can arise where formal rules and established practice diverge. Established practice must be generally regarded as having the force of law if it is to substitute for existing legislation (or its absence) in our assessment. In Canada, neither the Northwest Territories nor Nunavut has a formal role in multilateral constitutional negotiations. Both were full partners in the Charlottetown negotiations of 1992, but until this becomes institutionalized in practice or recognized by the courts we do not assume that they have the rights of Canadian provinces and we do not upgrade their score from zero to 4. In Australia, the Northern Territory does not have the formal right to be consulted on reforming its statute. While the federal government has been receptive to negotiation, it has insisted on keeping the final decision with the Commonwealth parliament, and we score the territory zero on bilateral constitutional reform.²³

We conclude this section with a discussion of bilateral constitutional reform. The criteria are parallel to those for multilateral constitutional reform, and the target becomes the constitutional position of the region, rather than the regional tier. No region can be expected to gain a majority in a national chamber, but a regional government or a regional electorate might be able to propose, postpone or even veto reform of its constitutional position.

Two further issues need clarification: how do we define bilateral constitutional reform, and how do we adjudicate cases with access to bilateral and multilateral reform?

BILATERAL CONSTITUTIONAL REFORM

- 0: the central government or national electorate can unilaterally reform the region's constitutional relation with the center;
- 1: a regional referendum can propose or postpone reform of the region's constitutional relation with the center;
- 2: the regional government can propose or postpone reform of the region's constitutional relation with the center or require a popular referendum;
- 3: a regional referendum can veto a reform of a region's constitutional relation with the center;
- 4: the regional government can veto a reform of the region's constitutional relation with the center.

The bilateral constitutional relationship between a region and the center is usually specified in a special statute, law, or section of the constitution and

²³ Statehood for the Northern Territory has long been in prospect. In 1978, Prime Minister Fraser anticipated statehood within five years. In August 1998, Prime Minister Howard announced Commonwealth support for the territory becoming a state. In 2009–12, the federal government expressed its support for a new attempt to grant the Northern Territory statehood, but the government put the plans on ice when popular support in the Northern Territory appeared to slip, partly because it seemed unlikely that the Northern Territory would be given the same number of senate seats as the other six states.

How We Apply the Coding Scheme

enforceable in a court, independent committee, and occasionally in international law. This precludes two things: (a) the authority of a region to write its own constitution, which is part of self-rule; (b) the capacity of an individual region to affect nation-wide constitutional reform, which falls under multilateral constitutional reform. Especially (a) deserves note, because, almost by definition, every constituent unit in a federation has the authority to write its own constitution, and many regions in decentralized countries do as well. That is starkly different from a region's right to redefine the bilateral constitutional relationship with the center, which is at the core of bilateral constitutional control.

How do we adjudicate cases with access to multilateral as well as bilateral constitutional reform? Our dataset contains just four regions in that situation: the Malaysian special regions of Sabah and Sarawak, and Serbia and Montenegro in the Yugoslav federation until 2002. Sabah and Sarawak have full bilateral rights because no constitutional change on existing legislative authority, powers over judicial administration, religion, language, immigration, and residence within the state shall be made "without the concurrence of the Yang di-Pertua Negeri of the State of Sabah or Sarawak or each of the States of Sabah and Sarawak concerned" (C 1957, Art. 161E). They are also full participants in multilateral constitutional reform, and their votes are pivotal in the senate to block unilateral federal reform of the constitution. These regions therefore have both full multilateral and bilateral scores.

Serbia and Montenegro (1992–2002) is more ambiguous. A change in the constitution required a two-thirds majority in both federal chambers, which is multilateral shared rule. But some key constitutional articles, including those relating to secession, boundaries, the federal character of the state, and competence allocation, fall under stricter, bilateral control: they require legislative majorities in each republic as well as a two-thirds majority in the lower house of the federation. These provisions allow an individual republic to block change to its one-on-one relationship with the center. In a two-member federation, the differences between bilateral and multilateral shared rule shrink. In 2003, Serbia-Montenegro becomes a confederation, and from then on, constitutional change requires the consent of both republics' legislatures, which we interpret to be bilateral. Serbia and Montenegro score 4 on both multilateral and bilateral constitutional reform until 2002.

Bosnia and Herzegovina is a clear-cut example of multilateral shared rule. The upper house has a veto on constitutional amendments; there is no vote in the *Republika Srpska* or the *Federacija Bosne i Hercegovine*, the individual entities. The entities score 4 on multilateral constitutional reform through their delegates in the upper house. A possible complexity may come from the fact that an ethnic group can invoke an alarm bell procedure in the upper house, which then requires that a law (including a constitutional law)

be supported by a majority in each of the three ethnic groups in order to pass. However, since all of this neither requires a regional referendum nor intervention by the entity governments, this does not amount to bilateral control.

No other regions combine multilateral and bilateral shared rule. Differentiated regions in Bolivia, Denmark, Finland, Spain, Italy, Panama, the Philippines, Portugal, Nicaragua, the United Kingdom, and the US have bilateral but no multilateral control over constitutional reform. In Spain, each *comunidad* can veto changes to its statute of autonomy, which regulates its particular relationship with the center within the confines of the constitution. A revised statute requires a supermajority in the *comunidad* assembly (two-thirds to three-fifths, depending on the *comunidad*) as well as a majority in both chambers of the legislature. In *comunidades* that took the fast track to autonomy, changes also need to be ratified by regional referendum. Bilateral shared rule is balanced by the fact that the *comunidades* do not have multilateral shared rule.

Types of Regions

We indicate four types of region in the appendix using the notation *S Y A D*.²⁴

- A *standard region* (*S*) is part of a regional tier and has a multilateral association with the central state. Standard regions have a uniform institutional set up within a tier, and we estimate them as such.
- An *asymmetric region* (*Y*) is embedded in a national tier, yet has distinctive authority on one or several dimensions of the RAI. Asymmetry is usually specified in an executive decision, constitutional article, or special clause in framework legislation.
- An *autonomous region* (*A*) is exempt from the country-wide constitutional framework and receives special treatment as an individual jurisdiction. It operates mostly in a bilateral setting with the central state alone. The arrangement is laid down in a special protocol, statute, special law, or separate section of the constitution.
- A *dependent region* (*D*) is not part of a standard tier, but is governed hierarchically by the central state. It has a separate government with no, or very little, authority.

²⁴ This analytical framework is developed in Volume II of this study (Hooghe and Marks forthcoming).

How We Apply the Coding Scheme

Two key features underpin these distinctions. The first concerns how a region stands in relation to other regions. Is the region part of a tier (*S*); is it part of a tier, yet has distinctive authority (*Y*) (e.g. Quebec or Catalonia); is the region anomalous (*A*) (e.g. Scotland or Aceh); or is the region excluded from a regional tier (*D*) (e.g. Misiones, Isla de la Juventud, Labuan)? The second feature concerns how a region stands in relation to the central state. Is the association multilateral, as part of a tier (*S* and *Y*); is it bilateral, so that the region relates to the central state individually (*A*); or is the relationship a unilateral one in which the region is governed by the central state (*D*)?

It is not uncommon for two or more of these types to co-exist in a country. Contemporary Canada has all four: standard regions (nine provinces and a lower tier of counties in Ontario and regional conferences in Quebec), asymmetry (Quebec), autonomy (Northwest Territories, Yukon, Nunavut, Self-governing Aboriginal Peoples), and dependency (Indian Act bands).

The status of individual regions may change over time. In 1950, Argentina had ten dependent territories. In the next decade eight of these became standard provinces and Tierra del Fuego followed in 1991. In 1996 Buenos Aires became autonomous. Sometimes a region switches back and forth between one or the other status. Aceh became a standard *provinsi* of Indonesia in 1957. It was granted an autonomous statute two years later, which was rescinded when the region was re-absorbed as a standard *provinsi* in 1966. In 2001 Aceh regained its special autonomous status. Northern Ireland alternated between home rule and dependency four times in thirty-five years.

Most regions fit clearly into this typology, but there are some gray cases. A distinction that appears translucent in theory can become opaque when applied to Belgium. Belgium is the only country in our dataset that has a regional tier with no standard regions. Each of the five jurisdictions in its upper tier has distinct competences. The Flemish community combines regional and community competences that are exercised separately by the Francophone community and the Walloon region. The German community exercises some bilateral shared rule, and is not a routine partner in intergovernmental meetings on executive policy (though it can send a representative if it maintains that its competences are affected). Because these regions/communities are regulated by the same constitutional provisions and the same special laws we consider them to be asymmetric rather than autonomous. However, we consider the Brussels region to be autonomous because it is governed by its own special law, has a unique consociational governance structure, and has distinct legal output (ordinances instead of decrees or laws). It is also subject to special federal tutelage to safeguard its role as an international capital, which is the foundation for a direct bilateral link with the federal government. It is also exempt from (or denied) institutional representation in the senate, and it has no role in constitutional reform—either multilaterally or bilaterally.

Measurement

Many dependencies have shed their subordination to the center and acquired self-governance. We observe forty-three dependencies in 1950 and just nine in 2010. Most have been transformed into standard provinces, states, or departments in big bang reforms, as in Argentina, Brazil, Colombia, and Venezuela. However some dependencies gain autonomy in steps, and this poses the question: at what point do we assess the transition away from dependence?

The Australian Northern Territory provides an example. The Northern Territory became a dependency in 1910 when South Australia ceded the territory to the federal government. At first it was run by the federal government, but over time the territory received some autonomy. We regard the decisive break from a dependent to an autonomous region to be the Northern Territory (Self-Government) Act which set up “separate political, representative and administrative institutions and . . . control over its own Treasury” (Preamble, Northern Territory Act 1978). The territory gained authority over the same range of policies as states (including health, education, social welfare, criminal and civil law, local government, residual powers, and concurrent powers over economic policy), except for control over immigration, uranium mining, and Aboriginal lands. Like states, the territory can set the base and rate of minor taxes, and it can borrow under the same rules. We classify the Northern Territory as an autonomous rather than a standard region chiefly because its relationship with the center remains primarily bilateral—and somewhat unequal: it has only one senator (against six for a state), its powers are not constitutionally guaranteed, the governor-general may withhold assent or recommend amendments to proposed territory laws, and, in contrast to standard Australian states, the territory’s autonomy statute can be changed unilaterally by the federal parliament.

The Philippine region of Mindanao has shifted from dependency to autonomy, but only after some false starts. The initial step was the internationally brokered Tripoli Accord of 1976, which set out extensive autonomy for thirteen provinces. However, implementation was lacking. The *Batas Pambansa BLG. 20* Act of 1979 divided the area in two regions, the regions of Central and Western Mindanao, each with a region-wide partially elected assembly and a dual executive, but it did not put decentralization into effect. After democratic transition a new attempt was made to grant autonomy. The key document is the Organic Act of 1989, which recognizes a single region as the Autonomous Region of Muslim Mindanao (ARMM) with boundaries to be determined by referendum. This Act installed a directly elected assembly and governor, devolved taxation powers, and gave Mindanao competences in regional and urban development. The new constitution of 1990 formally enshrined autonomy for “Muslim Mindanao” and introduced Sharia law in some parts of Muslim Mindanao’s justice system. Most scholars date autonomy in 1990 to coincide with the constitutional reform and the first elections (Bertrand 2010:

How We Apply the Coding Scheme

178). We begin coding ARMM as an autonomous region from 1990, and we introduce the two Mindanao regions as dependencies in the dataset when they were set up under President Marcos.

Aggregating the Scores

We score at the level of the individual region, or, in the case of standard regions, at the level of the regional tier, and we provide annual scores for ten dimensions.

Self-rule (0–18)

Institutional depth	Policy scope	Fiscal autonomy	Borrowing autonomy	Representation	
				Assembly	Executive
0–3	0–4	0–4	0–3	0–2	0–2

Shared rule (0–12)

Law making						Executive control		Fiscal control		Borrowing control		Constitutional reform				
L1	L2	L3	L4	L1	L2	L5	L6									
Multilateral	Bilateral			Multilateral	Bilateral	Multilateral	Bilateral	Multilateral	Bilateral	Multilateral	Bilateral	Multilateral	Bilateral			
0–2	or	0–2			0–2	or	0–2	0–2	or	0–2	0–2	or	0–2	0–4	or	0–4

The RAI for an individual region is the sum of scores for self-rule and shared rule. Self-rule is the sum of scores for institutional depth, policy scope, fiscal autonomy, borrowing autonomy, and representation (assembly representation plus executive representation). Shared rule is the sum of scores for law making, executive control, fiscal control, borrowing control, and constitutional reform.²⁵

Under exceptional circumstances a region or regional tier may receive a score for both multilateral and bilateral rule. For each shared rule dimension we use the greater of the multilateral or bilateral score in aggregating a region’s RAI. The

²⁵ We design the intervals within the ordinal scale to be equivalent and hence arithmetically summable. Chapter One finds that the RAI is robust when we vary weights across self-rule and shared rule.

Measurement

maximum regional score for self-rule is 18 and the maximum score for shared rule is 12, yielding a maximum RAI of 30 for a region or regional tier.

We aggregate regional scores to country scores in three steps. First, we calculate a score for each standard tier and each non-standard region. Second, we weight scores by population for each tier.²⁶ Third, we sum the weighted regional scores for each tier.²⁷

The data is accessible on the project's website (<http://www.falw.vu/~mlg/>). It consists of the following:

- A dataset and codebook, "RAI regional scores," with annual scores for 240 regional governments/tiers in sixty-five countries for the period 1950–2010.
- A dataset and codebook, "RAI country scores," with annual scores for all countries in the regional dataset plus scores for sixteen countries that do not have regional governance.
- Three calculation datasets with population figures and aggregation formula.

This chapter concludes the discussion of the general principles that guide our measurement. It is now time to introduce the reader to the implementation of those principles. We have designed an instrument for measuring regional authority. Will it fly? That is to say, will it produce estimates that make sense both to experts on particular countries and regions and to comparativists who may find it useful to summarize a vast amount of information in a systematic and accessible way?

²⁶ Where a tier is composed of regions with different RAI scores, we weight each region's score by its share in the national population. Where lower level regions exist only in a subset of higher level regions or where scores for lower level regions vary across higher level regions, the lower level scores are weighted by the population of the higher level regions of which they are part. We use population figures for 2010 or the nearest year except in the rare case that a country gains or loses territory or if the country is partitioned. A robustness check indicates RAI estimates using 2010 population data are not measurably different from estimates using decadal census data.

²⁷ Hence, the more regional tiers a country has, the greater the country score, all else equal.