

Leviathan's Latent Dimensions:
Measuring State Capacity for Comparative Political Research

Jonathan K. Hanson
Gerald R. Ford School of Public Policy
University of Michigan

Rachel Sigman
Naval Postgraduate School

Journal of Politics, forthcoming
<https://doi.org/10.1086/715066>

Date of manuscript 2020

Short title: Leviathan's Latent Dimensions

Abstract

State capacity is a core concept in political science research, and it is widely recognized that state institutions exert considerable influence on outcomes such as economic development, civil conflict, democratic consolidation, and international security. Yet, researchers across these fields of inquiry face common problems involved in conceptualizing and measuring state capacity. In this article, we examine these conceptual issues, identify three core dimensions of state capacity, and develop the expectation that they are mutually supporting and interlinked. We then use Bayesian latent variable analysis to estimate state capacity at the conjunction of indicators related to these dimensions. We find strong interrelationships between the three dimensions and produce a new, general-purpose measure of state capacity with demonstrated validity for use in a wide range of empirical inquiries. It is hoped that this project will provide effective guidance and tools for researchers studying the causes and consequences of state capacity.

Keywords: State capacity, measurement, Bayesian statistics, latent variable analysis

Replication files are available in the JOP Data Archive on Dataverse (<http://thedata.harvard.edu/dvn/dv/jop>).

Support for this research was provided by the Gerald R. Ford School of Public Policy at the University of Michigan and the Maxwell School for Citizenship and Public Affairs at Syracuse University.

In the influential volume, *Bringing the State Back In*, Evans, Rueschemeyer, and Skocpol (1985) noted a surge of interest in the state as an actor. This interest has not abated in the years since. It is widely recognized that state institutions exert considerable influence on outcomes including economic growth, human development, civil conflict, international security, and the consolidation of democracy. Along with the proliferation of theories containing state capacity as an explanatory variable, however, has come divergence in how it is conceptualized, impeding our ability to compare findings and expand our understanding of its roles. The difficulty of measuring state capacity empirically, however conceptualized, magnifies this problem.

A core question confronting scholars of state capacity is how to address the multidimensional nature of the concept. Despite a multitude of theorized, underlying dimensions of state capacity, the conceptual and empirical interrelationships among these dimensions remain poorly understood, leading to a number of potential measurement issues. First, absent clear definition of the concepts underlying state capacity, researchers may select dimensions and measures that are not relevant to their research (Berwick and Christia 2018; Cingolani 2013; Soifer 2008) or to the broader concept of state capacity. Second, measures are not always distinct from other concepts of interest such as economic development or regime type. Third, sparse geographic and temporal coverage for many measures of state capacity may prevent researchers from using the best measures possible.

In this article, we seek to address these challenges theoretically and empirically. First, we draw upon the growing literature on state capacity to identify the most fundamental functions of modern states and outline three core types of capabilities that state organizations must possess in order to fulfill those functions. Second, drawing upon a wide range of carefully selected indicators, we use Bayesian MCMC models to estimate state capacity as a latent variable. The results provide strong evidence of the interrelationship of the three dimensions, and validity tests demonstrate the utility of this variable as an aggregate estimate of state capacity.

The central result of this investigation is thus a general-purpose empirical tool built upon the idea that “state capacity” arises from the interrelationship of its most commonly identified dimen-

sions. The estimate has two key strengths for empirical analysis compared with other work in this area. First, by incorporating information from indicators related to multiple dimensions of state capacity, it is more comprehensive than measures that are oriented on a single dimension. Second, with annual estimates for every state in the Polity dataset from 1960-2015, it provides broader temporal and geographical coverage than other projects that seek to measure state capacity in a comprehensive manner. This measure is thus well-suited for a broad variety of comparative analyses, especially cross-national studies set in the post-war, post-colonial era.

Defining State Capacity

Usage of the term state capacity varies considerably across the literature in political science and related disciplines. This variation creates potential confusion for its use as a “productive, analytical concept” (Centeno et al. 2017, 4) and complicates the task of measurement. Further complications arise from an abundance of concepts that refer to other, closely-related attributes of states: strength, fragility, failure, effectiveness, efficiency, quality, legitimacy, autonomy, scope, and so on. With such a broad array of concepts in use, it is not surprising that state capacity “remains a concept in search of precise definition and measurement” (Hendrix 2010, 273).

As a starting point for a definition of state capacity that is conducive to reliable comparative measurement and avoids conflation with other concepts, we recognize that many works share the central idea that state capacity relates to the state’s ability to implement its goals or policies (Cingolani 2013). Beyond this concordance, however, lie two key areas of divergence about what it means for states to possess such abilities. The first concerns the nature of the state’s power. The second involves defining the set of functions on which state capacity should be assessed. In this section, we examine these debates and outline a definition of state capacity that embraces areas of agreement among different approaches.

State capacity embodies state power, as in the ability of one actor (the state) to get another actor (members of society) to do things they would not otherwise do (Dahl 1957). Like others working in this area, we seek a conception of state power that avoids conflation with other concepts and

eschews normative beliefs about what constitutes legitimate exercise of state power (e.g. Centeno et al. 2017; Lindvall and Teorell 2016). It is helpful to begin with Mann’s concept of infrastructural power: the capacity of the state to penetrate society and “to implement logistically political decisions throughout the realm” (Mann 1984, 189).

As Soifer (2008) describes, scholars think about infrastructural power in three ways: the state’s material capabilities, its effects on society, and its territoriality. In defining state capacity for the purpose of creating a measure that is useful for comparative research, it is more constructive to focus on the capacities that exist within the state’s organizational structures, and the territorial reach of these capacities, than on the effects of state actions on social relations and identities. Assessing state capacity based on the state’s effect on society risks entangling a decision to not deploy state power with its inability to do so. Additionally, these outcomes often serve as dependent variables in political science research.

State capabilities include material resources and organizational competencies internal to the state that exist independently of political decisions about how to deploy these capabilities. Giddens observes, for example, that “resources are the media through which power is exercised” (Giddens 1979, 141). Lindvall and Teorell (2016), similarly, describe state power as arising from access to monetary, human, and informational resources. Others direct attention to the organizational and bureaucratic competence of state institutions (Centeno et al. 2017, 4-7), which itself flows from resources, expertise, and professionalism. The territorial reach of the state is likewise central to its level of capacity, and we note there is vibrant scholarship on variation in state capacity at the subnational level (Foa and Nemirovskaya 2016; Harbers 2015; Harbers and Steele 2020; Soifer 2008).¹

Finally, we argue that Mann’s concept of despotic power – the “range of actions which the

¹Although our focus in this paper is not on subnational measurement of state capacity, we see efforts to measure state capacity in this way as complementary to our approach

elite is empowered to undertake without routine, institutionalized negotiation with civil society groups” (1984, 188) – is deeply entwined with characteristics of political regimes and issues of state autonomy that should be separated from the concept of state capacity. As Lindvall and Teorell (2016) argue, the state’s capacity is a function of the power it projects, which is conceptually distinct from mechanisms for societal involvement in political decisions regarding what outcomes states should pursue (e.g. democracy) or from the power of civil society to push back against the state (Migdal 1988).

A second issue in defining state capacity relates to the question of what functions a capable state should have the capacity to perform. The capacity to do what? Connected to this question of scope is the issue of whether we can conceive of capacity as a general characteristic of states that relates to core state functions or whether a disaggregated approach is required.

On one end of the spectrum lie approaches that define a state’s capabilities in terms of its most essential features and functions. For example, some state capacity research focuses on the concept of “stateness,” which involves the extent to which the state lives up to its Weberian definition as holding a monopoly on the legitimate use of force in its territory (Linz and Stepan 1996), and is sometimes used interchangeably with state capacity (Møller and Skaaning 2011). A specification this narrow would obscure the complexities of modern states and produce measurement strategies that are incapable of capturing important variation in contemporary state capacity. To study states in the modern, post-colonial era, it is necessary to recognize that the expected roles of states are not simply about the establishment of a monopoly on the legitimate use of force.

On the other end of the spectrum lie approaches that consider a much broader range of functions. Work in this perspective describes states as serving a potentially large number of roles such as the development and maintenance of economic systems, the provision of public services to the population, and the administration of justice (Bersch et al. 2017; Besley and Persson 2011; Rauch and Evans 2000). For example, Besley and Persson (2011) include a wide range of fiscal, administrative, public service delivery, and legal capacities in their definition.

Conceptual and measurement issues flow from this question of scope. As Levi contends, “good analysis requires differentiating among the features of the state in order to assess their relative importance; the state becomes less than the sum of its parts” (2002, 34). By this logic, a state’s capacity is assessed severally with respect to particular functions or goals. Skocpol (1985), in foundational work, uses the plural “state capacities,” noting the potential for the unevenness of state capabilities across policy areas or sectors. Recent empirical work fruitfully builds upon this perspective (Bersch et al. 2017; Foa and Nemirovskaya 2016; Gingerich 2013) by addressing variation in capacity across states agencies and regions.

Definitions that lead to assessments of state capacity across a highly disaggregated set of state functions, however, may drift from a core theoretical focus on the state’s ability to implement goals (Cingolani 2013, 36-37), capturing instead the results of “negotiations within the state and between it and other actors regarding the level, type, and form of intervention in society” (Centeno et al. 2017, 4). In other words, these approaches risk conflating the issue of the state’s capacity to implement policies in a particular sector or region with the political decision to prioritize these functions. For this reason, Fukuyama (2004, 7) distinguishes between state scope and state strength. In the United States, he explains, the state is relatively limited in terms of its scope of activities but, “within that scope, its ability to create and enforce laws and policies is very strong.”

In order to advance conceptual clarity, embrace the multi-functional realities of modern states, and facilitate cross-national measurement comparability, we seek a middle ground between these perspectives. We thus define state capacity as the state’s ability to perform the *core functions* most commonly deemed necessary for modern states: protection from external threats (Tilly 1990), the maintenance of internal order, the administration and provision of basic infrastructure necessary to sustain economic activity (Mann 1984), and the extraction of revenue (Levi 1988; North 1981; Tilly 1990). This approach steers clear of normative questions about what states should do, avoids conflating capacity with political priorities, and creates a viable framework for comparative analysis. It provides the basis for a measurement strategy that focuses upon key dimensions of state

capacity rather than disaggregate the concept into ever-smaller functional roles.

Dimensions of State Capacity

Even when focusing on core state functions, there remains a multitude of theorized dimensions of state capacity that relate to such functions. This array of potential dimensions creates confusion for researchers when employing the concept and selecting appropriate measures. It also raises a broader question: with so many underlying dimensions, is state capacity sufficiently coherent as a concept to be amenable to measurement?² In this section, we discuss how the literature addresses the dimensionality of state capacity and distill three essential and plausibly distinct dimensions out of the many that appear. We then consider the mutually supporting nature of these dimensions and make the argument that state capacity can be measured as a latent concept that lies at their intersection.

In a review of state capacity scholarship, Cingolani (2013) identifies at least seven different dimensions of state capacity in use: coercive, fiscal, administrative/implementation, transformative/industrializing, relational/territorial, legal and political capacities. We add several more to this assortment and illustrate the widely varying terminology across the literature (see the online appendix). In another comprehensive review, Berwick and Christia (2018) note that researchers often describe state capacity as involving only the aspects that they confront in their particular inquiry. The result is a confusing array of dimensions and insufficient attention to how specific dimensions relate to the broader concept of state capacity.

The dimensionality of state capacity appears in three basic ways across the literature. First, many approaches, explicitly or implicitly, treat these dimensions as operating independently of each other. Albertus and Menaldo (2012), for instance, argue that coercive capacity in particular undermines democratization because of the potential for effective repression of pro-democratic

²Coherence is an important feature of concept formation that refers to the internal consistency of the instances or attributes of the phenomenon (Adcock and Collier 2001; Saylor 2013).

movements. Likewise, in a study of state compliance with international human rights treaties, Cole (2015) argues that administrative capacity supports effective enforcement of such treaties, while other types of state capacity do not. Studies of this genre underscore the desirability of measuring state capacity and its dimensions in disaggregated fashion.

Second, many approaches use a measure of one dimension of state capacity as being a strong proxy for the overall concept. For example, many studies state that a government's ability to tax its population serves as a good overall representation of state capacity due to the broad range of infrastructural capabilities it requires (Brautigam et al. 2008; Harbers 2015; Rogers and Weller 2014). In other recent work, Brambor et al. (2020), D'Arcy et al. (2019), and Lee and Zhang (2017), develop measures of "legibility" and "information capacity," citing the crucial role of information as a resource for the state for taxation, conscription, growth promotion, and administration. Both sets of studies, accordingly, express the idea that one dimension of state capacity is a key dimension because of the way in which it supports and links with other dimensions.

Third, other approaches take these interrelationships one step further, conceiving of the dimensions as a set of mutually dependent underpinnings that work in tandem to enable states to perform a broader range of functions. Tilly's (1990) account of European state formation embodies this school of thought as, according to his account, the imperatives of territorial protection and conquest drove the development of states with capacities to raise revenue, build armies, and provide public goods. In a similar fashion, Besley and Persson observe strong complementarities in the development of states' fiscal and legal capacities, noting that "investments in one aspect of the state reinforce the motives to invest in the other" (2011, 15).

We note that empirical efforts to disaggregate state capacity have produced ambiguous results. Hendrix (2010), for example, makes a conceptual distinction between military capacity and administrative capacity but finds in factor analysis that indicators such as military expenditures load heavily on the same dimension (factor) as high-quality bureaucratic institutions. In another study, Fortin-Rittenberger (2014) investigates the relationship between two dimensions: infrastructural

capacity, which combines indicators of both extractive and administrative capabilities, and coercive capacity. Her results also point to the difficulty of measuring these dimensions, particularly coercive capacity. Large militaries, she finds, are equally dispersed across states with low and high infrastructural capacity thereby complicating efforts to disentangle the two dimensions. Accordingly, we argue, more attention should be devoted to the question of whether state capacity should be conceived and measured as a single concept, or whether it is more fruitful for researchers to focus on specific dimensions of state capacity.

To address the question of aggregation, while following the definitional principles laid out in the previous section, we concentrate on three dimensions of state capacity that are 1) minimally necessary to carry out the functions of contemporary states, and 2) most plausibly distinct from one another. These criteria lead us to the identification of three dimensions: extractive capacity, coercive capacity, and administrative capacity. These three dimensions accord with what Skocpol identifies as providing the “general underpinnings of state capacities” (1985, 16): plentiful resources, administrative-military control of a territory, and loyal and skilled officials.³

The extractive, coercive, and administrative aspects of state capacity are fundamental to modern states. Raising tax revenue is not only a critical function of the state to support all of its activities, but it also encompasses a particular set of capabilities that are foundational to broader powers of the state. In particular, states must be able to reach their populations, collect and maintain information, possess trustworthy agents to manage the revenue, and have enforcement capabilities to ensure compliance (Pomeranz and Vila-Belda 2019). North defines the boundaries of the state in terms of its ability to tax constituents (1981, 21), while Levi (1988) and Tilly (1990) make a direct connection between a state’s revenue and the possibility to extend its rule. Empirically, taxation is associated with property rights (Besley and Persson 2009), the reach of the state (Harbers 2015),

³These three aspects of state capacity are also similar to those examined by Soifer (2015) in his study of Latin American state building and what Berwick and Christia (2018) propose as a unifying framework.

and state legibility (Lee and Zhang 2017).

Like extractive capacity, coercive capacity is also central to the definition of the state, particularly in the Weberian tradition that defines the state as the organization possessing a monopoly on the legitimate use of force within its territory (Weber 1919). Coercion connects directly to the state's ability to preserve its borders, protect against external threats, maintain internal order, and enforce compliance with the law. To perform other functions, including the collection of revenue, a state must possess the force necessary to contain threats throughout its territory, or at least convince its rivals that this is the case. While coercion is not the only way to maintain order and evoke compliance from the population (Levi 1988), it represents a key aspect of the ability of states to survive and implement policies.

Administrative capacity is an encompassing dimension that pertains to the state's organizational capabilities with respect to developing policy, delivering public services, and regulating commercial activity. Effective policy administration is a function of capable state agents, technical competence, data collection and record-keeping, monitoring and coordination mechanisms, and effective reach across the state's territory and social groupings. In particular, Weber (1919) emphasizes the importance of professional bureaucracies that legitimize the authority of the state, manage complex affairs, and ensure efficiency, but non-Weberian forms of bureaucratic organization can also be effective (Darden 2008).

Thus, even though these dimensions are distinct conceptual lenses through which one can usefully think about state capacity, there are logical reasons to believe that, in practice, they are mutually constitutive and interrelated. It is the need for coercive capacity, according to Tilly (1990), that drives leaders to adopt tax systems and provide goods and services. Gurr (1988), for example, argues that coercive power involves the institutionalization of the means of coercion, which requires capable personnel and functional specialization of state agencies. According to Levi (1988), the keys to effective revenue extraction are measurement, monitoring, and enforcement capabilities, which in modern states often necessitate bureaucratic revenue collection backed by a coercive

apparatus. Finally, as Fjelde and De Soysa state, “governments rely on revenue to invest in the military, police, and bureaucratic apparatus, which in turn allow them to accumulate power for further penetration and extension of state rule” (2009, 8).

If there are elemental linkages between the coercive, extractive, and administrative dimensions of state capacity, we should expect they will be related to each other empirically. This logic provides the basis for a strategy to estimate state capacity as a latent variable that arises from the conjunction of its extractive, coercive, and administrative capabilities. The outcome of this investigation has important implications for the way we advance knowledge of state capacity. If state capacity dimensions are empirically inseparable from each other, research that claims to study one dimension of state capacity may actually capture a broader phenomenon. Conversely, if state capacity dimensions do not cohere into a broader construct, researchers must be especially careful to select measures that meaningfully represent the narrower concept of interest.

Measurement Strategies and Challenges

As a latent concept, state capacity (or its underlying dimensions) is not directly observable, but it is connected to a range of indicators from which we can learn information about its level. In this section, we consider various indicators that relate to the three dimensions presented above. For each dimension, we discuss a range of possible measurement strategies and explain our own selection of indicators.⁴

We apply several criteria to decide which indicators to include in our latent variable model. First, we consider conceptual fit with the three core dimensions of state capacity, avoiding those that overlap too much with other concepts. Second, with the goal of gathering sufficient information to capture variation in state capacity in many countries over five decades, we seek broad

⁴In the online appendix, we provide a list of possible state capacity indicators with coverage and descriptive data.

geographical and temporal coverage.⁵ Finally, we avoid aggregate indexes, as they may include either indicators used individually in our model or indicators that are connected more closely to other concepts. The selection criteria are described in more detail in the online appendix.

Indicators of Extractive Capacity

Measures of extractive capacity typically come in two main forms. First, many researchers use data on government revenue collections as a measure of state capacity.⁶ Tax revenue data are available for most countries from the early-1970s onwards, generally from the IMF's *Government Finance Statistics*.⁷ Data on different types of revenues are usually expressed as a raw amount, as a proportion of GDP, or as a proportion of total revenue collected. As Lieberman (2002) explains, there are many factors to consider when selecting revenue indicators that are appropriate for a particular purpose.

Aggregate revenue, for example, is a noisy indicator of extractive capacity. For states with relatively high extractive capacity, the level of tax revenue collection reflects a policy choice rather than extractive capacity. Additionally, different types of revenue vary significantly in terms of their administrative complexity. As Lieberman (2002) and Rogers and Weller (2014) argue, the revenue sources that are most likely to capture concepts related to state capacity include income, property and domestic consumption taxes. These taxes are more administratively complex, requiring higher levels of record-keeping, transparency, and a more sophisticated bureaucratic apparatus than other revenue sources. Taxes on international trade, on the other hand, are much easier to collect and, like rents from mineral resources, do not require significant enforcement capacity (Lieberman 2002,

⁵The temporal coverage we strive for is more limited than recent efforts to generate long-run time series of particular dimensions of state capacity (Brambor et al. 2020; D'Arcy and Nistotskaya 2017), but includes broader country coverage and a more encompassing conceptual approach.

⁶See, for example, Besley and Persson (2009) and Dincecco (2017).

⁷Prichard et al. (ICTD/UNU-WIDER 2017) have usefully standardized and compiled tax data from IMF country records. Tax data are also available from other sources such as the OECD.

98). In some cases, researchers have sought to assess the amount of tax collected relative to an estimated expected amount of revenue (Kugler 2018; Arbetman-Rabinowitz et al. 2012). Though this measure of “relative political capacity” is useful for some applications, we argue that it differs conceptually from state capacity and find in empirical tests that it correlates only weakly.

Our strategy with respect to revenue data is twofold. First, we use total tax revenues as a proportion of GDP to capture overall extractive capacity. We exclude non-tax revenues for the reasons Lieberman identifies. Second, we expect that the mixture of tax revenues – specifically taxes on income and taxes on trade – provides information about *both* the extractive and administrative capacities of the state. Given a particular level of taxation, the greater the proportion of tax revenue that comes from income taxes, the higher the expected level of administrative capacity. The opposite should be true with respect to the proportion of revenue that comes from taxes on trade, which are administratively easy to collect. We thus use the proportion of tax revenues –as opposed to taxes as a proportion of GDP– that come from these two sources as measures of the administrative capability of the state’s extractive efforts.

We also include expert-coded indicators such as the World Bank’s (2017) Country Policy and Institutional Assessment (CPIA) rating of the Efficiency of Revenue Mobilization. From Coppedge et al. (2019), we use a measure of State Fiscal Capacity (`v2stfiscap`) capturing the extent to which the state is able to fund itself through taxes that are of greater administrative complexity. Finally, we expect that some of the indicators that are logically related to the dimensions of coercive and administrative capacity will also provide information about extractive capacity. For example, a state’s ability to collect information about its citizens is relevant for extractive capacity, something we discuss in greater detail below.

Indicators of Coercive Capacity

Researchers seeking to measure coercive capacity may turn attention to military size or sophistication, as well as attributes of the state thought to promote the maintenance of order. Data on military expenditures, military personnel, and security forces are available from datasets such as the World

Development Indicators, the Stockholm International Peace Research Institute, and the Correlates of War (Singer et al. 1972). Coverage and reliability for these measures is generally quite good for most countries in the period 1960 to the present. The relationship between coercive force and a state's coercive capacity, however, is not necessarily straightforward (Hendrix 2010; Kocher 2010; Soifer and vom Hau 2008). States that have the capacity to maintain order might have effective military and/or security forces, although there are countries that maintain order with little or no military. A large military force, moreover, may be a sign of war or insecurity, both of which could deplete state capacity. We use the log value of military expenditures per million in population and the number of military personnel per thousand in the population (Singer et al. 1972; World Bank Group 2020) as indicators of military capacity. We also include a measure of the size of the police force obtained from the United Nations Office on Drugs and Crime.

In light of potential issues with indicators of personnel or spending, we also include other, expert-coded indicators of coercive capacity. From the Bertelsmann Transformation Index (Bertelsmann Stiftung 2006), we adopt a measure that assesses the degree to which the state has a monopoly on the use of force. We also include ratings from The Political Risk Services' (PRS) International Country Risk Guide on "law and order" which assesses the strength and impartiality of the legal system, and the popular observance of the law (Howell 2011).

Finally, two indicators tap the dimension of coercive capacity by capturing the state's level of institutionalization or presence in the territory (i.e. stateness). First, we use V-Dem's (Coppedge et al. 2019) measure of State Authority over Territory ($v2svsterr$), which measures the percentage of territory controlled by the central state.⁸ Second, we extend the State Antiquity Index developed by Bockstette et al. (2002) to code 27 additional countries and to reflect territorial changes and sovereignty post-1950.⁹ This measurement strategy is based on work showing the importance of

⁸The corresponding proportion is converted to the inverse of the cumulative standard normal distribution.

⁹We extend the original measure with annual coding of its three components – presence of a

historical roots of the state in its territory (Boone 2003; Herbst 2014; Wimmer 2016).

Indicators of Administrative Capacity

Since administrative capacity is a broad dimension of state capacity, a number of different measurement strategies exist. A common way to measure administrative capacity is to look at the outcomes of public goods and service delivery such as the percentage of children enrolled in primary schools, infant mortality rates, or literacy rates. These measures are attractive for their broad coverage and comparability, but assessing capacity based on measures of this kind poses several problems. First, as discussed above, a state may not prioritize the particular outcome being measured, such as schooling or health or infrastructure. Second, using these measures may compromise analytical leverage, since these types of outcomes are closely linked to economic development, the nature of the political regime, or participation in international programs with policy conditions.

Among indicators of administrative capacity, two of the most popular are the Government Effectiveness rating from the Worldwide Governance Indicators (Kaufmann et al. 2003) and the International Country Risk Guide's (ICRG) Bureaucratic Quality rating (Howell 2011). Both measures have come under scrutiny. The WGI, for example, are frequently criticized for their aggregation procedures and for the fuzzy analytical boundaries that characterize their different governance indices.¹⁰ In our case, using the WGI scores would be duplicative because the set of constituent indicators overlaps with others we employ. The ICRG Bureaucratic Quality ratings, on the other hand, may be prone to measurement errors based on subjective analyst perceptions of economic or social outcomes rather than bureaucratic quality *per se* (Henisz 2000). We include the ICRG Bureaucratic Quality rating in our analysis, however, since it is one of the few measures with relatively broad coverage that focuses on strength of the bureaucracy, including mechanisms of recruitment

state, percentage of territory under the control of that state, and whether that state is sovereign – for each year from 1950 through 2015.

¹⁰There are debates about the validity, reliability, and aggregation of the WGI. For an overview and response to critiques see Kaufman et al. (2007).

and training. Our estimation procedures expect some noise in the component indicators.

We also include several measures of administrative capacity from various sources: Administrative Efficiency (Adelman and Morris 1967), the Weberianness index (Rauch and Evans 2000), and ratings of Quality of Budgetary and Financial Management and Quality of Public Administration from the World Bank's CPIA index. None of these ratings covers a long period of time, but the combination covers significant portions of the 1960-2015 time period with at least one indicator.¹¹ Finally, we include the measure of impartial public administration developed by V-Dem (Coppedge et al. 2019), a Bayesian item response theory model based on expert survey ratings on a five-point scale related to the extent to which the law is fully respected by public officials.

Additionally, we include a set of measures aimed at capturing the information-gathering capabilities of states. First, we derive a measure of census frequency calculated with data on country censuses provided by the U.S. Census Bureau.¹² As argued in Centeno (2002b) and Soifer (2013), countries that can conduct censuses have not only the capacity to collect information exhibit higher levels of territorial reach. These data cover 173 countries throughout the 1960-2015 time period. Second, we use the measure of information capacity developed by Brambor et al. (2020), which is derived from indicators of whether a state has a statistical agency, a civil register, a population register, and its capabilities relative to producing a census and a statistical yearbook. The information capacity index covers 70 countries during the 1960-2015 time period. Finally, we include the World Bank's Statistical Capacity measure, which assesses the extensiveness of statistical systems in up to 139 countries annually from 2004-present.

¹¹We code Administrative Efficiency as covering the years 1960-1962 and Weberianness as covering the period 1970-1990 based on the objectives of their creators.

¹²We have annualized this measure by looking forward and backward in time from a given year to find the nearest censuses. The longer the gaps between censuses, the lower the Census Frequency measure.

Table 1: Indicators of State Capacity

Indicator	Countries	Years
Administrative Efficiency (Adelman and Morris 1967)	69	1960–1962
Bureaucratic Quality (Political Risk Services)	141	1984–2015
Census Frequency (calculated from UN 2011)	173	1960–2015
Efficiency of Revenue Mobilization (World Bank CPIA)	72	2005–2015
Fiscal Capacity (V-Dem v9)	174	1960–2015
Information Capacity (Brambor et al. 2020)	70	1960–2015
Law and Order (Political Risk Services)	141	1984–2015
(log) Military Personnel per 1,000 in population (COW, WDI)	176	1960–2015
(log) Military Expenditures per capita (SIPRI, COW)	176	1960–2015
Monopoly on Use of Force (Bertlesmann Transformation Index)	129	2006–2015
(log) Police Officers per 1,000 in population (UN)	121	1973–2015
Quality of Budgetary and Financial Management (World Bank CPIA)	72	2005–2015
Quality of Public Administration (World Bank CPIA)	72	2005–2015
Rigorous and Impartial Public Administration (V-Dem v9)	177	1960–2015
State Antiquity Index, based on Bockstette et al. (2002)	172	1960–2015
State Authority over Territory (V-Dem v9)	177	1960–2015
Statistical Capacity (World Bank)	127	2004–2015
Taxes on Income as % of Taxes (ICTD, IMF)	168	1963–2015
Taxes on International Trade as % of Taxes (ICTD, IMF)	167	1960–2015
Total Tax Revenue as % of GDP (ICTD, IMF, OECD)	167	1960–2015
Weberianness (Rauch and Evans 2000)	34	1970–1990

Indicators overall

Altogether, we employ 21 different indicators related to the three key dimensions of state capacity (Table 1). The indicators span 56 years (1960-2015) and up to 163 countries in a given year, with 94,135 data points in total. In 99% of country-years, at least 6 indicators are available, and the median number of indicators per country-year is 12. By adopting a latent variable analysis of the kind employed to assess measures of democracy (Pemstein et al. 2010; Treier and Jackman 2008) and governance (Arel-Bundock and Mebane 2011; Bersch and Botero 2014) we can use these multiple measurements of the same underlying concept to gain information about the distribution of the latent parameters that generate the observed indicators.

Latent Variable Analysis

We employ the latent variables estimation approach developed by Arel-Bundock and Mebane (2011) that uses Bayesian Markov-Chain Monte Carlo (MCMC) techniques to identify underlying factors. This technique, based on earlier work by Lee (2007), has many advantages over traditional factor analysis including robustness to missing data. By incorporating indicators of state capacity drawn from multiple sources, we seek to provide annual measures of state capacity for the set of all countries that appear in the Polity dataset (Marshall and Jaggers 2016) during the 1960-2015 time period.

Specifically, each observed indicator x_k for country i in time t is a linear function of J latent variables and a disturbance ϵ_k :

$$x_{kit} = c_k + \sum_{j=1}^J \lambda_{kj} \xi_{jit} + \epsilon_{ki} \quad (1)$$

In Equation 1, ξ_{jit} is the latent value of the j th dimension of state capacity for country i in time t , and λ_{kj} is the linear effect of the j th dimension on the observed indicator x_k . Overall, then, the various observed indicators are linear functions of the latent values of state capacity in each dimension measured with some error. Since there are k observed indicators measured in many countries over several years, we have multiple data points with which to obtain the posterior distributions of the latent parameters.¹³ We assign standard normal priors to the latent factors. The intercepts c_k have independent, diffuse normal priors, and the disturbance terms ϵ_k have independent uniform priors with mean zero. In general, diffuse normal priors are used for each λ_{kj} .

¹³The greater the number of indicators, the more information we have about the values of latent dimensions of state capacity in country i at time t . The larger the number of country-years, the more information we have to uncover λ_{kj} , the effect of dimension j on indicator k , which is treated as constant over time.

To facilitate identification, one of the parameters λ_{kj} is fixed at 1 for each of the J dimensions in the analysis. In these cases, the intercepts c_k are fixed at 0. Additionally, truncated (positive) normal priors were applied to facilitate identification where we had a strong prior belief that the relationship between a given indicator (x_k) and the parameter representing Capacity (ξ_j) is positive. In our main model (with $J=1$) truncated, normal priors are applied in the following cases: census frequency, state antiquity, taxes on income, Weberianness, the World Bank's statistical capacity index, information capacity, the V-Dem public administration measure, PRS law and order, and the administrative efficiency rating of Adelman and Morris (1967).¹⁴

The MCMC is implemented in JAGS through the package `rjags` (Plummer 2012) for R statistical software. The algorithm tours the parameter space specified by the sets of equations represented by Equation 1. Successive draws lead to descriptions of the posterior distributions of the remaining parameters that produce the observed indicators of state capacity. A typical MCMC run included five chains with an adaptation phase of 5,000, a burn-in phase of 10,000 iterations, and a sampling phase of 5,000 iterations. Samples were thinned with a setting of 5 to alleviate memory/storage constraints.

In order to test whether the three theorized dimensions are discernible in the data independently, we run multiple analyses, letting the number of dimensions J range from 1 to 3. The parameter estimates that emerge from choosing a particular number of dimensions need not bear any particular relationship to the theoretical dimensions we describe. As with traditional factor analysis, we would rely on analysis of which indicators align with the resulting parameters to interpret the dimensions. One possibility is that each successive dimension captures more marginal aspects of variation in the observed indicators rather than clear dimensions.

¹⁴Otherwise, some chains would simply take on the opposite signs of other chains.

Latent Variable Analysis Results

In repeated tests, we found that a one-dimensional model ($J = 1$) was the only model to converge consistently. Attempts to identify a second or third dimension did not bear fruit. Typically, the different chains would fail to converge, and the posterior distributions for some parameters would exhibit strong non-normality. These outcomes arise when the MCMC routine does not produce a stationary distribution for various parameters. In other words, given a particular set of observed indicators, and a specification of multiple dimensions of state capacity that are connected to these indicators, the routine does not yield information about the relative probabilities for different levels of state capacity in these dimensions and the parameters that connect these levels to the observed indicators. Consequently, in the sections that follow we present results that reflect a single, latent dimension that we call Capacity.

Accordingly, we believe that the results are consistent with the theoretical perspective that extractive, coercive and administrative dimensions of state capacity, though distinct conceptually, are interrelated in practice. Extractive capacity supports both coercive power and provides the resources needed to sustain a sophisticated administrative bureaucracy. Likewise, states that lack coercive and administrative capabilities are likely to find revenue extraction more difficult. Finally, although state coercion can take many different forms, some of them very simple, coercive power is facilitated by a well-organized, administratively-sophisticated coercive apparatus. These interrelationships make it difficult empirically to disaggregate state capacity into separate dimensions, which has been noted in previous efforts to understand the relationship between these dimensions (Fortin-Rittenberger 2014; Hendrix 2010).¹⁵

¹⁵We also conduct a dimensionality test using traditional factor analysis. The results, provided in the online appendix, also produce latent factors lacking any clear relationships to the three dimensions.

Exploring the Aggregate Measure

As a latent variable that lies at the conjunction of state capacity's core dimensions, Capacity plausibly captures the concept more comprehensively than previous work that is focused on a single indicator or dimension. Additionally, with 8,254 observations in total, the Capacity estimates have much broader coverage than the most commonly used general indicators of state capacity for research in the post-war, post-colonial era. The measure can thus serve to fill an important niche in comparative cross-national research involving state capacity, particularly for large-sample analysis. A comparison of coverage with other measures is presented in the online appendix. The Capacity measure is scaled from -2.31 to 2.96, with a mean of .26 and a standard deviation of .95.

To understand what factors drive the Capacity estimates, we first examine their correlation with the observed indicators included in the estimation procedure, presented in Table 2. Overall, capacity appears to be a general-purpose measure of state capacity that draws from indicators representing all three theorized dimensions. The indicators most strongly associated with Capacity are: the World Bank's measure of Statistical Capacity ($r = .83$); the PRS Bureaucratic Quality ($r = .81$); V-Dem's Rigorous and Impartial Public Administration ($r = .80$); PRS Law and Order ratings ($r = .77$); the CPIA's Quality of Public Administration rating (.74); BTI's Monopoly on Use of Force rating ($r = .74$); and the measure of State Fiscal Capacity from V-Dem ($r = .73$). Most of the indicators are correlated with Capacity at the .5 level or greater (or less than -.5 in the case of taxes on trade).

The indicators with weakest correlation to Capacity are the measures related to military and police personnel. Since other measures of coercive capacity are strongly correlated with Capacity, the pattern appears to be limited to security personnel. A few explanations seem plausible. First, the Capacity measure misses aspects of coercive capacity that arise from state employment of security personnel. Second, rulers of weaker states, or those engaged in conflict, tend to expand their security forces in response to this weakness, thereby further weakening the relationship. Third, it is not the numbers of security personnel that matter but their level of capability as measured by

Table 2: Correlation of Capacity with Base Indicators

Indicator	r	N
Statistical Capacity	0.83	1492
Bureaucratic Quality	0.81	4089
Rigorous and Impartial Public Administration	0.80	8252
Law and Order	0.77	4089
Quality of Public Administration	0.74	724
Monopoly of Force	0.74	1247
Fiscal Capacity	0.73	7673
Qual. of Budgetary and Financial Management	0.71	724
Administrative Efficiency	0.70	199
(log) Military expenditures per capita	0.70	7925
Efficiency of Revenue Mobilization	0.67	724
State Authority over Territory	0.66	8237
Total Tax Revenue as % of GDP	0.66	6413
Information Capacity	0.66	3591
Weberianness	0.59	714
Census Frequency	0.59	8201
Taxes on Income as % of Tax Revenue	0.57	5854
State Antiquity Index	0.42	8032
(log) Military Personnel per 1,000 in population	0.26	8116
(log) Police Officers per 1,000 in population	0.03	1569
Taxes on International Trade as % of Tax Revenue	-0.67	6270

their administrative organization or technological sophistication.

Validity Checks

The broader coverage of countries and years is welcome, provided that the measures perform well. The goal of this section is to investigate whether the Capacity measure behaves in the expected manner, and whether it will be useful for investigating theoretical questions regarding state capacity. Following guidance from Adcock and Collier (2001), Seawright and Collier (2014) and McMann et al. (Forthcoming), we examine the new measure in terms of its face validity, content validity, convergent validity, and nomological validity.

Figure 1 displays the mean and standard deviation of each country's Capacity posterior distribution in the year 2015, ranked from the highest to the lowest. In terms of face validity, the countries we might expect to have strong state capacity are found to have higher scores, while those that

are experiencing or have recently experienced war or have notoriously weak capacity are found to have the lower scores. That Singapore ranks among the twenty-five highest Capacity scores help us to know that these measures do not capture concepts more closely related to democratic governance than to capacity itself. At the lower end of the scale, we see states such as Somalia, Yemen and Central African Republic that are embroiled in conflict, lacking state structures, or both. Plots similar to Figure 1 for several other years are included in the online appendix.

To examine over-time variation, Figure 2 plots Capacity scores for all the countries in the dataset, with 1975 scores on the x-axis, 2015 scores on the y-axis, and a 45 degree line between the two. As theory would predict, the relationship between the Capacity variables in different years is strongly positive. Most countries starting with high scores in 1975 also have high scores in 2015. Overall, Capacity rose in most countries, rising the most in Uganda, Bolivia, Rwanda, Lesotho, and Nicaragua. Countries where Capacity decreased the most include Somalia, Libya, Venezuela, Syria, Kuwait, and Iraq. In the online appendix, we show the evolution of Capacity over time in Chile, Haiti, Iraq, and Singapore to further illustrate the face validity of the measure's temporal variation.

Given the latent nature of the Capacity measure, we check convergent validity by comparing the Capacity variable with other measures that were not used in the MCMC process in order to assess whether it accurately taps the intended concept of state capacity. We choose a variety of other indicators, most of which are other indexes, constructed using different methodologies. If Capacity is a valid measure, we should observe strong correlation with other attempts to measure this concept.

As can be seen in Table 3, the Capacity measure is quite strongly correlated in the expected direction with a broad range of these other measures in pairwise tests. Among the those most strongly correlated with Capacity, for example, are the WGI Government Effectiveness index ($r = 0.91$), the WGI Rule of Law index ($r = .88$) ratings, the Fragile States Index ($r = -.88$), and the WGI Regulatory Quality index ($r = .86$). Among these measures, Capacity is the least correlated

Figure 1: Posterior Distribution of Capacity in the Year 2015

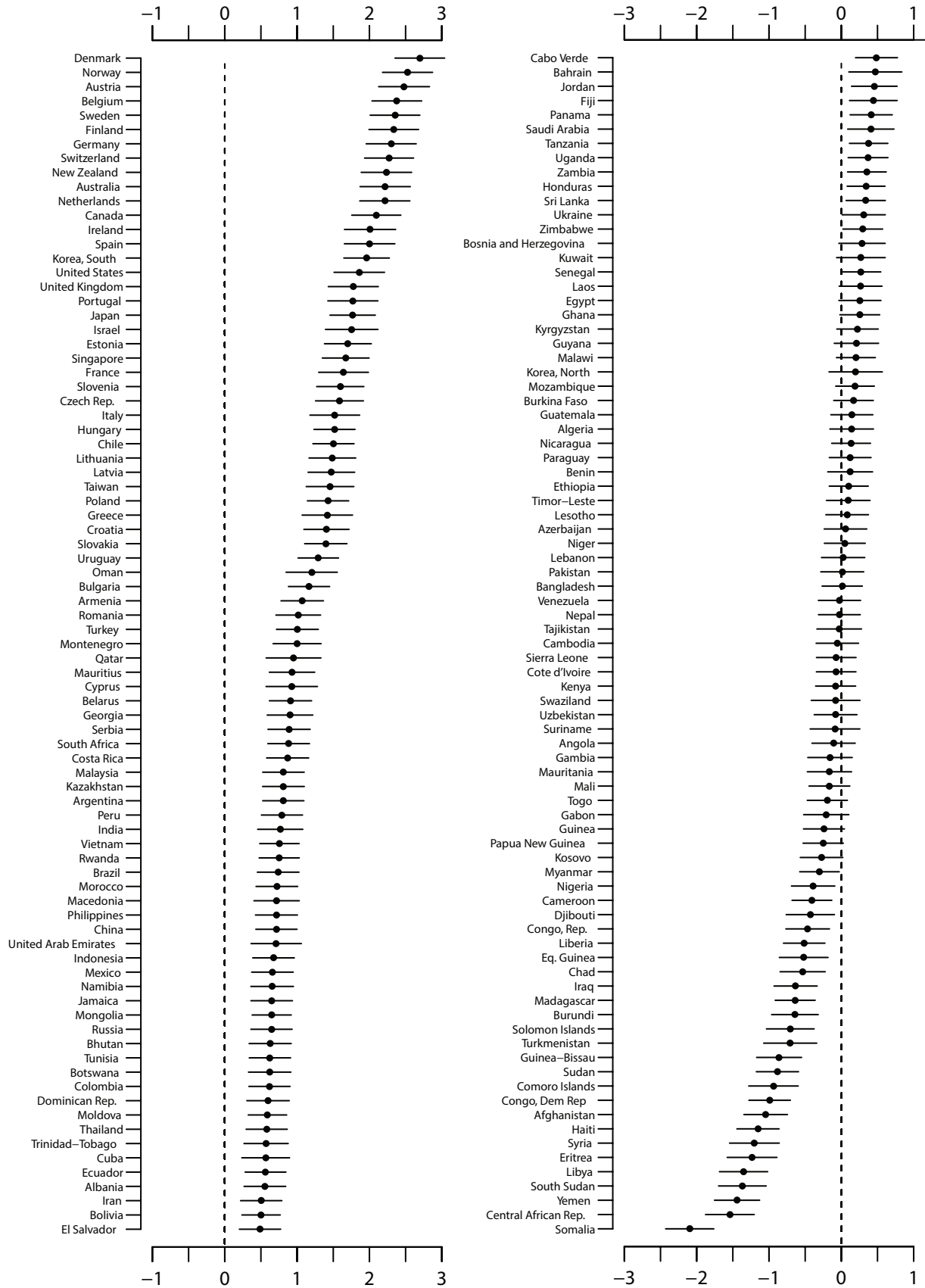
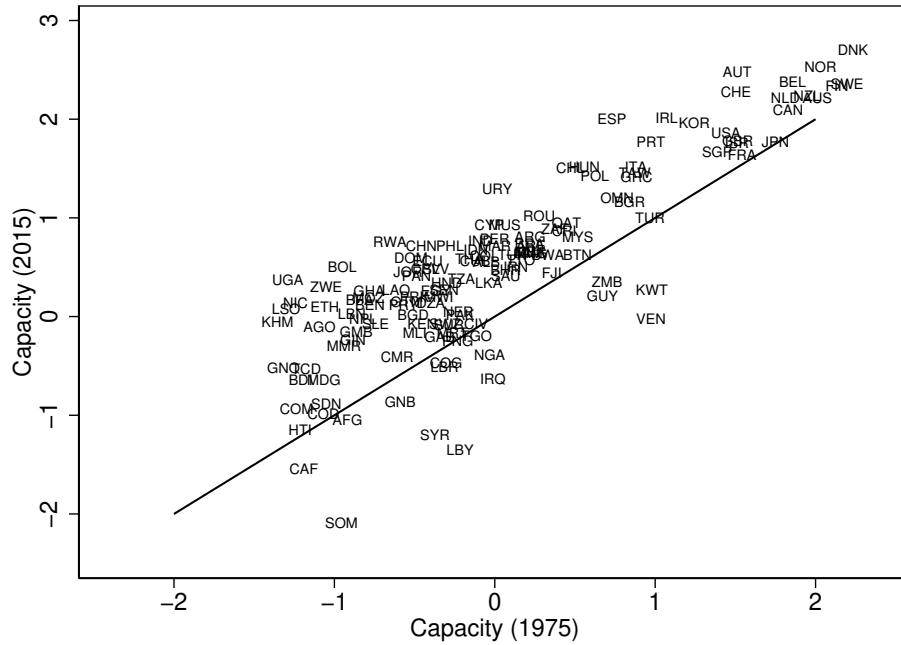


Figure 2: Scatter Plot of Capacity 1975 and 2015



with the BTI Management Index ($r = .66$).

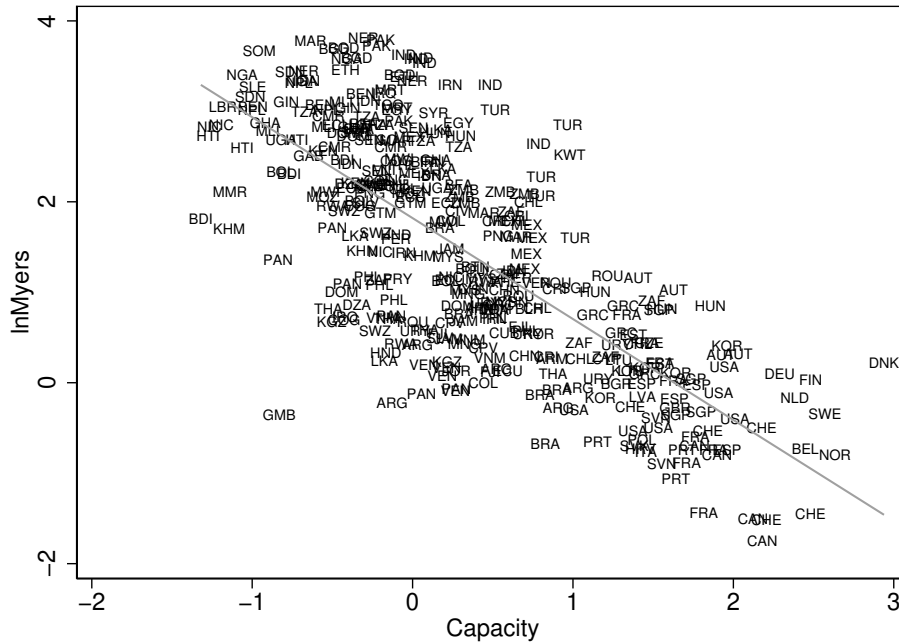
In a recent notable work, Lee and Zhang (2017) develop a measure of legibility – the extent of state information about citizens that is available in standardized forms – built upon the accuracy of age-reporting in national censuses. Where the state creates little reason to know one’s age exactly, citizens tend to report their ages in numbers that end with zeros or fives. The degree of “heaping” creates a way to measure legibility: the Myers index, which Lee and Zhang show to have a moderately strong correlation to other measures of state capacity. Although constructed in a very different manner, the Capacity measure developed here is correlated more strongly with both the log Myers index ($r=-.74$) and many of those other measures than they are with each other. Figure 3 illustrates this relationship.

A valid measure should also discriminate between the concept of interest and other concepts. Table F.1 in the online appendix presents correlations between Capacity and measures such as elections, conflict, and economic growth that may be correlated with state capacity but represent distinct concepts. The correlations are particularly low for indicators of conflict, population, oil

Table 3: Correlations of Capacity with Other Measures

Indicator	r	N
Government Effectiveness (WGI)	0.91	2782
Rule of Law (WGI)	0.88	2784
Regulatory Quality (WGI)	0.86	2783
Rule of Law Index (Bertlesmann TI)	0.68	1592
Impartial Public Administration (Rothstein & Teorell)	0.80	50
Public Sector Management Average (CPIA)	0.82	724
Rational-Legal (Hendrix)	0.84	1408
CPIA Index (CPIA)	0.80	724
Stateness Index (Bertlesmann TI)	0.77	1592
Management Index (Bertlesmann TI)	0.66	1588
(log) Myers index (Lee and Zhang 2018)	-0.74	345
Public Services indicator (Rice and Patrick)	-0.86	1719
Fragile States Index (Rice and Patrick)	-0.88	1719

Figure 3: Scatter Plot of lnMyers and Capacity



production per capita, urbanization, and the number of consecutive presidential elections. The relationship between Capacity and measures of other regime traits are slightly higher, though still not as high as the alternative state capacity indicators presented in Table 3. Unsurprisingly, the correlation between Capacity and log GDP per capita is fairly high ($r = .79$), but it is reassuring that these two variables are not capturing exactly the same thing.

We further demonstrate validity in tests where we use the Capacity measure as a predictor of various outcomes widely associated with state capacity. Table 4 presents the results from six regression models that test whether Capacity predicts development outcomes even after controlling for log GDP per capita. In each of these tests, Capacity is a substantively strong and statistically significant (at the 99% level) predictor.

Table 4: Construct Validity Tests for Capacity

	(1) InformalEcon	(2) lnMyers	(3) PublicServ	(4) Letters	(5) AveDays	(6) eGov
Capacity	-1.96** (0.21)	-0.47** (0.10)	-0.32** (0.09)	20.83** (3.95)	-69.04** (13.59)	0.11** (0.01)
lnGDPcap	-5.95** (0.24)	-0.61** (0.10)	-0.62** (0.10)	2.16 (2.39)	-16.28* (8.21)	0.08** (0.01)
Constant	83.58** (1.94)	6.47** (0.76)	11.23** (0.85)	28.00 (18.33)	409.09** (63.04)	-0.25** (0.06)
N	1350	345	1719	150	150	164
R^2	0.99	0.91	0.97	0.45	0.51	0.86
Fixed Effects?	Yes	Yes	Yes	No	No	No

[^] $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

In Model 1, the dependent variable is a measure of the size of the shadow economy as a percentage of GDP (Schneider et al. 2010).¹⁶ We find that each one point increase in Capacity is associated with a reduction in the size of the shadow economy by 1.96 percentage points of GDP, controlling for the log level of GDP per capita. Models 2 and 3 use as dependent variables

¹⁶The shadow economy includes “all market-based legal production of goods and services that are deliberately concealed from public authorities” (Schneider et al. 2010, 444).

the log Myers Index and the Fragile States Index Public Services indicator, which is a measure of the state's capability to carry out core functions (higher values mean less capability). Even after controlling for log GDP per capita, Capacity is strongly associated with both of these measures. A one-unit increase in Capacity is connected with a reduction in the log Myers index to about 47% of its previous size and a decrease in the Public Services indicator by .32 points.

In Models 4 and 5, we draw upon a study conducted by Chong et al. (2014) to assess the efficiency of government in 159 countries by measuring how long it would take the country's postal service to return undeliverable mail to an international address. They sent 10 letters to each country and found that about 60% of letters were returned. The mean number of days it took to return a letter was about 228. For Model 4, the dependent variable is the percentage of letters sent to a country that were returned. Where Capacity is one point higher, the percentage of letters returned is about 20.8 percentage points higher. Similarly, in Model 5 where the dependent variable is the average number of days it takes to return a letter, a one-point increase in Capacity is associated with a reduction of about 69 days in how long it takes for the letter to be returned.

Finally, Model 6 uses data from United Nations E-Government Development Database, which tracks the e-governance readiness of each UN member country's government and the extent of citizen e-participation in government. The scale runs from 0 to 1, with higher scores meaning greater preparedness. We find that each one-point increase in Capacity predicts a .11 point increase in the e-Government Development Index, which is about one-half a standard deviation in the index.

To demonstrate the utility of the Capacity measure, we conduct a set of tests using the level of Capacity as measured in 1960, or the earliest available year for a country, as a predictor for the year-2010 levels of different development indicators.¹⁷ We consider this a very challenging test, since we control for the initial level of GDP per capita (logged), the mean level of Democracy during the period (using polity2 rescaled from 0-1), and the mean level of tax revenues as a percentage of

¹⁷Except for access to basic water services, which is from 2012.

GDP.¹⁸ As seen in Table 5, each test nevertheless shows Capacity to be strongly related to these outcomes.

Table 5: Illustrative Tests Using Capacity

	(1) InfMort	(2) LifeExp	(3) Roads	(4) Water	(5) Hospitals	(6) lnGDP/cap ₁₀
Capacity ₆₀	-12.15** (3.13)	5.32** (1.19)	0.41* (0.18)	6.92** (2.11)	1.52** (0.41)	0.59** (0.11)
lnGDP/cap ₆₀	-7.27** (1.69)	2.83** (0.64)	0.18 [^] (0.10)	5.49** (1.14)	0.26 (0.21)	0.63** (0.06)
Democracy	-9.40 (7.39)	4.19 (2.80)	-0.02 (0.40)	4.98 (4.98)	-1.90* (0.95)	-0.23 (0.27)
TaxRev	-0.26 (0.28)	-0.06 (0.10)	-0.01 (0.02)	0.12 (0.19)	0.12** (0.04)	0.01 (0.01)
Constant	95.29** (14.32)	45.33** (5.42)	-0.64 (0.80)	37.63** (9.65)	0.28 (1.82)	3.74** (0.52)
N	148	148	106	148	112	149
R ²	0.53	0.54	0.22	0.50	0.39	0.73

[^] $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Cross-sectional OLS regression with standard errors in parentheses. The dependent variables, measured in the year 2010 for all but Water (2012), are Roads (km of road per 100 squared km), Water (% of population using at least basic water services), Hospitals (number of hospital beds per 1,000), InfMort (infant mortality rate), LifeExp (level of life expectancy), and log GDP per capita. The independent variables are lnGDP/cap₆₀ (log level of GDP per capita in 1960), Democracy (mean level during the period 1960-2010), TaxRev (mean level of tax revenue as a percentage of GDP over the period 1960-2010), and Capacity in 1960.

In Model 1, the dependent variable is a country's infant mortality rate. Where Capacity was one-unit higher in 1960, mortality in 2010 is about 12.2 deaths lower per 1,000 infants, all the other variables being held constant. Similarly, as Model 2 shows, a one-unit higher 1960 Capacity score is associated with 5.3 years longer life expectancy in 2010. Models 3 through 5 present tests in which the dependent variables are measures of national infrastructure and health care facilities. Where Capacity was one unit higher in 1960, there are about .41 more kilometers of road per 100

¹⁸We control for the initial level of GDP per capita since 1960 levels of Capacity could affect subsequent economic growth and thus bias the estimates

square km land area, the percentage of citizens using at least basic water services in 2012 is about 6.9 points higher, and the number of hospital beds per 1,000 people to be 1.5 greater. Finally, as Model 6 shows, Capacity in 1960 is associated with greater GDP per capita 50 years later, even after controlling for the initial level of GDP per capita. Each one-unit increase in Capacity in 1960 is associated with GDP per capita in 2010 being 59% higher. The robustness of these results to controlling for country wealth and democracy, we argue, provides confidence that the Capacity measure is indeed capturing something that is distinct from these other concepts.

Summary

Ultimately our understanding of the causes and consequences of state capacity depends on our ability to measure it in valid, reliable, and practical ways. That state capacity is composed of multiple dimensions, fundamentally latent, and closely related to a range of concepts presents a particularly complicated set of challenges that researchers must overcome. In focusing on the use of state capacity across political science research, identifying its core theoretical dimensions, and systematically analyzing the best available data for these dimensions, we hope to have advanced the discussion of the conceptual and measurement issues related to state capacity, addressed recently by, among others, Centeno et al. (2017), Hendrix (2010), Fukuyama (2013), Lindvall and Teorell (2016), Rogers and Weller (2014), Soifer (2008).

In particular, our analysis has provided new insight into the empirical manifestations of state capacity. First, our findings suggest that the dimensions of state capacity are mutually constitutive and interrelated, meaning that attempts to isolate specific types of capacity may be difficult to achieve. Surely, states differ in which capabilities are most strongly developed, but significant strength in any one dimension likely requires at least some strength in the others. Second, given the interrelationship between these the dimensions, we hope that these data will facilitate research about how state capacity has developed since the decline of European colonialism in the mid-twentieth century. For example, researchers may want to pursue the “chicken and egg” question of state capacity: which dimension comes first, if any?

Second, the broader geographic and temporal coverage provided by these estimates can support research to provide new insights on a range of familiar questions. The value of this exercise is evidenced by a wide range of published research utilizing earlier versions of the Capacity measure. For example, the data have been used to produce new knowledge on state building processes (Grassi and Memoli 2016), resilience in electoral authoritarian regimes (van Ham and Seim 2018), the relationship between democracy and state capacity (Wang and Xu 2018), and even stock market development (Guillén and Capron 2016). The Capacity measure has also been used extensively as a control variable in cross-country regressions (Graham et al. 2017; Houle 2017). We hope that the measure can be useful in many longstanding debates about the relationship between institutions, economic growth and development outcomes.

With expanding data and sustained interest in the state as a conceptual variable in political science research, we are confident that measurement options will grow in the coming years. To make meaningful improvements on the data currently available, however, we recommend careful consideration of the issues laid out in this article, particularly as they relate to the need to focus on core functions of the state, to expand coverage of existing measures, and to eschew definitions of state capacity that relate too closely to decision-making procedures. The result, we contend, will be progress on assessing the effects of state institutions on a broad variety of outcomes.

Acknowledgements

We would like to thank many individuals for feedback with this project, especially Michael Beckstrand, Stephen Knack, Monika Nalepa, David Patel, Hillel Soifer, Vivek Srivastava, Jan Teorell, and attendees of conferences and workshops hosted by the American Political Science Association, Duke University, the ESID Research Centre, the Gerald R. Ford School of Public Policy, Lund University, Syracuse University, the University of Gothenburg, and the World Bank. We greatly appreciate technical help from Vincent Arel-Bundock, Bennet Fauber, Kyle Marquardt, and Juraj Medzihorsky. Finally, we thank anonymous reviewers for suggesting fruitful revisions, Sunghye Cho for research assistance, and Erica De Bruin and Jun Koga Sudduth for providing data.

References

- Adcock, Robert and David Collier. 2001. "The American Political Science Review." *Measurement Validity: A Shared Standard for Qualitative and Quantitative Research* 95(3):529–546.
- Adelman, Irma and Cynthia Taft Morris. 1967. *Society, Politics, & Economic Development: A Quantitative Approach*. Baltimore: The Johns Hopkins Press.
- Albertus, Michael and Victor Menaldo. 2012. "Coercive capacity and the prospects for democratization." *Comparative Politics* 44(2):151–169.
- Arbetman-Rabinowitz, Marina, Jacek Kugler, Mark Abdollahian, Kristin Johnson, and Kyungkook Kang. 2012. *Performance of Nations*. Lanham, MD: Rowman & Littlefield Publishers, Inc.
- Arel-Bundock, Vincent and Walter Mebane. 2011. "Measurement Error, Missing Values and Latent Structure in Governance Indicators." Presented at the Annual Meeting of the American Political Science Association, Seattle.
- Bersch, Katherine and Sandra Botero. 2014. "Measuring Governance: Implications of Conceptual Choices." *The European Journal of Development Research* 26(1):124–141.
- Bersch, Katherine, Sérgio Praça, and Matthew M. Taylor. 2017. "Bureaucratic Capacity and Political Autonomy Within National States: Mapping the Archipelago of Excellence in Brazil." In Centeno et al. (2017), pages 157–183.
- Bertelsmann Stiftung. 2006. "Bertelsmann Transformation Index."
- Berwick, Elissa and Fotini Christia. 2018. "State Capacity Redux: Integrating Classical and Experimental Contributions to an Enduring Debate." *Annual Review of Political Science* 21:71–91.
- Besley, Timothy and Torsten Persson. 2009. "The Origins of State Capacity: Property Rights, Taxation, and Politics." *The American Economic Review* 99(4):1218–1244.
- Besley, Timothy and Torsten Persson. 2011. *Pillars and Prosperity: The Political Economics of Development Clusters*. Princeton: Princeton University Press.
- Bockstette, Valerie, Areendam Chanda, and Louis Putterman. 2002. "States and markets: The advantage of an early start." *Journal of Economic Growth* 7(4):347–369.
- Boone, Catherine. 2003. *Political Topographies of the African state: Territorial authority and institutional choice*. Cambridge: Cambridge University Press.
- Brambor, Thomas, Agustín Goenaga, Johannes Lindvall, and Jan Teorell. 2020. "The Lay of the Land: Information Capacity and the Modern State." *Comparative Political Studies* 53(2):175–213.
- Brautigam, Deborah, Odd-Helge Fjeldstad, and Mick. Moore. 2008. *Taxation and State-Building in Developing Countries: Capacity and consent*. Cambridge: Cambridge University Press.
- Centeno, Miguel. 2002b. *Blood and Debt: War and the Nation-State in Latin America*. University Park: Pennsylvania State University Press.

- Centeno, Miguel A, Atul Kohli, Deborah J Yashar, and Dinsha Mistree. 2017. *States in the Developing World*. Cambridge: Cambridge University Press.
- Chong, Alberto, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2014. "Letter Grading Government Efficiency." *Journal of the European Economic Association* 12(2):277–299.
- Cingolani, Luciana. 2013. "The State of State Capacity: A Review of Concepts, Evidence and Measures.", UNU-MERIT Working Paper Series on Institutions and Economic Growth.
- Cole, Wade M. 2015. "Mind the Gap: State Capacity and the Implementation of Human Rights Treaties." *International Organization* 69(2):405–441.
- Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard, M. Steven Fish, Adam Glynn, Allen Hicken, Anna Lührmann, Kyle L. Marquardt, Kelly McMann, Pamela Paxton, Daniel Pemstein, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Steven Wilson, Agnes Cornell, Lisa Gastaldi, Haakon Gjerløw, Nina Ilchenko, Joshua Krusell, Laura Maxwell, Valeriya Mechkova, Juraj Medzihorsky, Josefina Pernes, Johannes von Römer, Natalia Stepanova, Aksel Sundström, Eitan Tzelgov, Yi ting Wang, Tore Wig, and Daniel Ziblatt. 2019. "V-Dem [Country-Year/Country-Date] Dataset v9." Varieties of Democracy (V-Dem) Project.
- Dahl, Robert A. 1957. "The Concept of Power." *Behavioral Science* 2(3):201–215.
- D’Arcy, Michelle and Marina Nistotskaya. 2017. "State First, Then Democracy: Using Cadastral Records to Explain Government Performance in Public Goods Provision." *Governance* 30(2):193–209.
- D’Arcy, Michelle, Marina Nistotskaya, and Robert Ellis. 2019. "Mapping the State: Measuring Infrastructural Power through Cadastral Records." In *Proceedings of the International Federation of Surveyors’ Working Week*.
 URL https://www.fig.net/resources/proceedings/fig_proceedings/fig2019/papers/ts04j/TS04J_nistotskaya_darcy_et_al_9784.pdf
- Darden, Keith. 2008. "The Integrity of Corrupt States: Graft as an Informal State Institution." *Politics & Society* 36(1):35–60.
- Dincecco, Mark. 2017. *State Capacity and Economic Development: Present and Past*. Cambridge: Cambridge University Press.
- Evans, Peter, Dietrich Rueschemeyer, and Theda Skocpol. 1985. *Bringing the State Back In*. Cambridge: Cambridge University Press.
- Fjelde, Hanne and Indra De Soysa. 2009. "Coercion, Co-optation, or Cooperation? State Capacity and the Risk of Civil War, 1961-2004." *Conflict Management and Peace Science* 26(1):5–25.
- Foa, Roberto Stefan and Anna Nemirovskaya. 2016. "How State Capacity Varies within Frontier States: A Multicountry Subnational Analysis." *Governance* 29(3):411–432.
- Fortin-Rittenberger, Jessica. 2014. "Exploring the Relationship Between Infrastructural and Coercive State Capacity." *Democratization* 21(7):1244–1264.

- Fukuyama, Francis. 2004. "The Imperative of State-Building." *Journal of Democracy* 15(2):17–31.
- Fukuyama, Francis. 2013. "What is Governance?" *Governance* 26(13):347–368.
- Giddens, Anthony. 1979. *Central Problems in Social Theory: Action, Structure, and Contradiction in Social Analysis*, vol. 241. Berkeley: University of California Press.
- Gingerich, Daniel W. 2013. "Governance Indicators and the Level of Analysis Problem: Empirical Findings from South America." *British Journal of Political Science* 43(3):505–540.
- Graham, Benjamin AT, Michael K Miller, and Kaare W Strøm. 2017. "Safeguarding democracy: Powersharing and democratic survival." *American Political Science Review* 111(4):686–704.
- Grassi, Davide and Vincenzo Memoli. 2016. "Political determinants of state capacity in Latin America." *World Development* 88:94–106.
- Guillén, Mauro F and Laurence Capron. 2016. "State capacity, minority shareholder protections, and stock market development." *Administrative Science Quarterly* 61(1):125–160.
- Gurr, Ted Robert. 1988. "War, Revolution, and the Growth of the Coercive State." *Comparative Political Studies* 21(1):45–65.
- van Ham, Carolien and Brigitte Seim. 2018. "Strong states, weak elections? How state capacity in authoritarian regimes conditions the democratizing power of elections." *International Political Science Review* 39(1):49–66.
- Harbers, Imke. 2015. "Taxation and the Unequal Reach of the State: Mapping State Capacity in Ecuador." *Governance* 28(3):373–391.
- Harbers, Imke and Abbey Steele. 2020. "Subnational Variation Across States: A Typology and Research Agenda." *Latin American Politics and Society* 62(3):1–18.
- Hendrix, Cullen. S. 2010. "Measuring State capacity: Theoretical and empirical implications for the study of civil conflict." *Journal of Peace Research* 47(3):273–285.
- Henisz, Witold. 2000. "The Institutional Environment for Multinational Investment." *Journal of Law, Economics, and Organization* 16(2):334–364.
- Herbst, Jeffrey. 2014. *States and Power in Africa: : Comparative Lessons in Authority and Control*. Princeton: Princeton University Press.
- Houle, Christian. 2017. "Inequality, ethnic diversity, and redistribution." *The Journal of Economic Inequality* 15(1):1–23.
- Howell, Llewellyn D. 2011. "International Country Risk Guide Methodology." PRS Group, Inc.
- ICTD/UNU-WIDER. 2017. "Government Revenue Dataset."
URL <https://www.wider.unu.edu/project/government-revenue-dataset>
- Kaufmann, Daniel, Art Kraay, and M. Mastruzzi. 2003. "Governance matters III: Governance Indicators for 1996-2002." World Bank Policy Research Working Paper No. 3106.
- Kaufmann, Daniel, Art Kraay, and M. Mastruzzi. 2007. "The Worldwide Governance Indicators Project: Answering the Critics." World Bank Policy Research Working Paper 4149.

- Kocher, Matthew Adam. 2010. "State Capacity as a Conceptual Variable." *Yale Journal of International Affairs* 5:137–145.
- Kugler, Jacek. 2018. *Political Capacity and Economic Behavior*. New York: Routledge Press.
- Lee, Melissa M. and Nan Zhang. 2017. "Legibility and the Informational Foundations of State Capacity." *The Journal of Politics* 79(1):118–132.
- Lee, Sik-Yum. 2007. *Structural Equation Modeling: A Bayesian Approach*. Hoboken, NJ: Wiley.
- Levi, Margaret. 1988. *Of Rule and Revenue*. Berkeley: University of California Press.
- Levi, Margaret. 2002. "The State of the Study of the State." In *Political Science: The State of the Discipline*, Ira Katznelson and Helen V. Milner, eds., New York: WW Norton.
- Lieberman, Evan S. 2002. "Taxation Data as Indicators of State-Society Relations: Possibilities and pitfalls in cross-national research." *Studies in Comparative International Development* 36(4):89–115.
- Lindvall, Johannes and Jan Teorell. 2016. "State Capacity as Power: A Conceptual Framework." STANCE Working Paper Series, No. 1. Department of Political Science, Lund University.
- Linz, Juan and Alfred Stepan. 1996. *Problems of Democratic Transition and Consolidation: Southern Europe, South America, and post-communist Europe*. Baltimore: Johns Hopkins University Press.
- Mann, Michael. 1984. "The autonomous power of the state: its origins, mechanisms and results." *European Journal of Sociology* 25(02):185–213.
- Marshall, Monty G. and Keith Jagers. 2016. "Polity IV Project: Dataset Users' Manual." Center for Systemic Peace, George Mason University.
- McMann, Kelly, Daniel Pemstein, Brigitte Seim, Jan Teorell, and Staffan Lindberg. Forthcoming. "Assessing Data Quality: An Approach and An Application." *Political Analysis* .
- Migdal, Joel. 1988. *Strong Societies and Weak States : State-Society Relations and State Capabilities in the Third World*. Princeton: Princeton University Press.
- Møller, Jørgen and Svend-Erik Skaaning. 2011. "Stateness first?" *Democratization* 18(1):1–24.
- North, Douglass. 1981. *Structure and Change in Economic History*. New York: Norton.
- Pemstein, Daniel, Stephen A Meserve, and James Melton. 2010. "Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type." *Political Analysis* 18(4):426–449.
- Plummer, Martyn. 2012. "rjags: Bayesian Graphical Models using MCMC." Version 3.5.
- Pomeranz, Dina and José Vila-Belda. 2019. "Taking State-Capacity Research to the Field: Insights from Collaborations with Tax Authorities." *Annual Review of Economics* 11:755–781.
- Rauch, James and Peter Evans. 2000. "Bureaucratic Structure and Bureaucratic Performance in Less Developed Countries." *Journal of Public Economics* 75(1):49–71.

- Rogers, Melissa Ziegler and Nicholas Weller. 2014. "Income Taxation and the Validity of State Capacity Indicators." *Journal of Public Policy* 34(2):183–206.
- Saylor, Ryan. 2013. "Concepts, Measures, and Measuring Well: An Alternative Outlook." *Sociological Methods & Research* 42(3):354–391.
- Schneider, Friedrich, Andreas Buehn, and Claudio E. Montenegro. 2010. "Shadow Economies All over the World." The World Bank. Policy Research Working Paper No. 5356.
- Seawright, Jason and David Collier. 2014. "Rival Strategies of Validation Tools for Evaluating Measures of Democracy." *Comparative Political Studies* 47(1):111–138.
- Singer, J David, Stuart Bremer, and John Stuckey. 1972. "Capability Distribution, Uncertainty, and Major Power War, 1820-1965." In *Peace, War, and Numbers*, pages 19–48, Beverly Hills: Sage.
- Skocpol, Theda. 1985. "Bringing the State Back In: Strategies of Analysis in Current Research." In Evans et al. (1985), chap. 1, pages 3–37.
- Soifer, Hillel. 2008. "State Infrastructural Power: Approaches to Conceptualization and Measurement." *Studies in Comparative International Development* 43:231–251.
- Soifer, Hillel and Matthias vom Hau. 2008. "Unpacking the *Strength* of the State: The Utility of State Infrastructural Power." *Studies in Comparative International Development* 43(3/4):219–230.
- Soifer, Hillel David. 2013. "State Power and the Economic Origins of Democracy." *Studies in Comparative International Development* 48(1):1–22.
- Soifer, Hillel David. 2015. *State Building in Latin America*. Cambridge: Cambridge University Press.
- Tilly, Charles. 1990. *Coercion, Capital, and European States, AD 990-1990*. Cambridge: Blackwell.
- Treier, Shawn and Simon Jackman. 2008. "Democracy as a Latent Variable." *American Journal of Political Science* 52(1):201–217.
- Wang, Erik H and Yiqing Xu. 2018. "Awakening Leviathan: The effect of democracy on state capacity." *Research & Politics* 5(2):2053168018772398.
- Weber, Max. 1919. *Politics as a Vocation? In From Max Weber: Essays in Sociology, HH Gerth and C. Wright Mills*. London: Routledge.
- Wimmer, Andreas. 2016. "Is Diversity Detrimental? Ethnic Fractionalization, Public Goods Provision, and the Historical Legacies of Stateness." *Comparative Political Studies* 49(11):1407–1445.
- World Bank Group. 2017. "CPIA 2017 Criteria."
URL <http://pubdocs.worldbank.org/en/203511467141304327/CPIA-Criteria-2017v2.pdf>
- World Bank Group. 2020. *WDI online*. Washington, DC: World Bank Group.
URL <https://databank.worldbank.org/source/world-development-indicators>

Biographical Statements

Jonathan Hanson is the Lecturer in Statistics for Public Policy at the Gerald R. Ford School of Public Policy, University of Michigan, Ann Arbor, MI, 48109.

Rachel Sigman is an Assistant Professor in the Department of National Security Affairs at the Naval Postgraduate School, Monterey, CA, 93943.