ON THE MEASUREMENT OF STATE CAPACITY

QUANTITATIVE QUESTIONS AND CONCEPTUAL CONSIDERATIONS

DOCTORAL DISSERTATION

By Andrea Vaccaro

Reviewers:

Rachel M. Gisselquist

Agnes Cornell

Committee:

Marina Nistotskaya

Antonio Savoia

Lyndal Keeton

SAPIENZA
UNIVERSITÀ DI ROMA

Department of Social Sciences and Economics

Sapienza University of Rome

Date of PhD Defence: October 18th, 2021

Abstract

The state has been "brought back in" and today state capacity is one of the most important research topics in a variety of social science subfields. Although conceptual, theoretical, and empirical literature on the topic has been flourishing for a while, issues related to the measurement of state capacity have attracted less interest in the scholarly debate for now. This is not to say that researchers have not recognised some of the problems related to the measurement of state capacity; quite the opposite. Nevertheless, measurement issues are seldom addressed in detail in studies on the topic. The main aim of my dissertation is to fill this gap in literature by addressing some of the most pressing questions on the measurement of state capacity.

In Chapter I, I provide a review of definitions and measures of state capacity. This introductory chapter provides a first look at the concept of state capacity and shows some of the consistencies and inconsistencies related to the definition as well as the measures of state capacity. The purpose of Chapter I is illustrative rather than exhaustive.

In Chapter II, I select some of the most established measures of state capacity for further investigation. These measures are systematically compared and evaluated with several statistical tools, but not only. I point out some interesting differences and weaknesses in the content and statistical properties of the measures. Even if I find that all the measures are highly convergent among one another, a battery of replicated studies shows that their interchangeability is weak. In other words, different measures of state capacity can lead to completely different interpretations. The primary cause of these somewhat counterpoising findings lies in divergent understandings of the level of state capacity in some countries. Since no measure of state capacity emerges as the best one in absolute terms, researchers are advised to select a measure that closely represents their theory or alternatively to test the robustness of their findings with multiple measures.

In Chapter III, I show that one way to increase our knowledge on state capacity is to dig deeper both in terms of theory and measures. I argue that the literature on the state-democracy nexus has remained largely inconclusive because of conceptual and measurement issues related to state capacity (as well as democracy). One way to push forward the research agenda on the topic is to shift the analysis to more specific aspects of the state-democracy relationship. Through a comparative statistical analysis of the relationship between impartial public officials and civic participation – two core attributes of state capacity and democracy – I provide new empirical evidence on the state-democracy nexus across countries and over time. Besides convincingly supporting the hypothesised positive effect of civic participation on impartial public officials, the findings of Chapter III indicate indeed that the research agenda on state capacity can be pushed forward by disaggregating the concept. A vibrant civic society seems to be an important prerequisite of impartial bureaucracies, but overly broad approaches on the state-democracy nexus fail to capture this "nuance".

In Chapter IV, I take the reasonable assumption that the need to conduct research on state capacity as a whole is not going to disappear. General level theories require general level measures. Scholars have plenty of measurements to choose from, but most of them were not created to capture state capacity: ill-suited measures stand in the way of accumulating knowledge on the topic. In Chapter IV, I develop and present a novel cross-national measure of state capacity to tackle the problem. A set of validity tests show persuasively that the new measure contributes to the literature on state capacity and is a useful tool to make progress on the research agenda on the topic. Chapter IV provides also novel statistical information on the key dimensions of state capacity.

In Chapter V, I sum up the main findings, discuss their potential implications for policymakers, and identify possible avenues for future research on state capacity.

Acknowledgements

Writing this dissertation would have not been possible without the immensely important help and support provided by several people who I would like to sincerely thank.

First of all, I am grateful to my PhD supervisors Luigi M. Solivetti and Emanuela Reale. Thank you for your invaluable guidance and inspiration throughout this journey.

Second, I would like to thank my colleagues at the Department of Social Sciences and Economics and the wider Sapienza community. Formal and informal intellectual exchanges that took place at our university played a key role for my dissertation project.

Third, I would like to thank all the brilliant scholars that I met during my international research visits at the University of Helsinki, at the University of Gothenburg, and at the United Nations University World Institute for Development Economics Research. Without your ideas, suggestions, and encouragement, this dissertation would have looked very much different – to be clear, much worse. All these three research visits were absolutely crucial for this project. I am particularly indebted to Rachel M. Gisselquist, Marina Nistotskaya, and Tero Erkkilä, who helped me to grow professionally and personally. I cannot thank you enough!

Fourth, I would like to thank the reviewers of my dissertation and the dissertation committee. Last but not least, I would like to thank my friends, and my family in Finland and Italy. Thank you for your wholehearted support and enormous patience.

Foreword

When I started my doctoral studies and to work on the dissertation at hand, the world was different than it is today. Back then, in late 2017, no one could have foreseen that the COVID-19 pandemic would come and completely change our lives. If back then, the state was an actor of deep concern mainly for social scientists, today, virtually everyone – also a non-specialist audience – acknowledges the fundamental role of the state in resolving some of the most demanding problems of our societies, such as the current coronavirus pandemic.

Unsurprisingly, some of the most influential social scientist in the world, have argued that state capacity is the key to win the fight against the pandemic (e.g., Fukuyama 2020). Therefore, if possible, the content of my dissertation has become even more topical and salient than it was at the time I started the writing. During the pandemic we have seen in practice how countries with dysfunctional state apparatuses have not been able to successfully implement their intended policies nor reach their objectives. Building more capable states must be then one of the most urgent objectives for policymakers around the world. Not only right now, but also in the light of future crises where effective responses by states are needed.

If this is true, gaining knowledge on state capacity, how to build it, and how it is related to other factors becomes of vital importance. Without good quantitative information on state capacity, addressing such issues becomes difficult. Numbers matter, and thus, the accurate measurement of state capacity must become one of the main concerns for both social scientists and policymakers.

Table of contents

Abbreviations

CAP	Vaccaro's State capacity index
CCSI	Core civil society index
CMI	Carbone and Memoli's index of state capacity
CPI	Corruption perceptions index
CSPART	Civil society participation index
DV	Dependent variable
EFA	Exploratory factor analysis
FE	Fixed effects
FSI	Fragile states index
GDP	Gross domestic product
HF	Heritage Foundation
HSI	Hanson and Sigman's Capacity index
ICRG	International country risk guide
ICTD	International Centre for Tax and Development
IMF	International Monetary Fund
KMO	Kaiser-Meyer-Olkin test of sampling adequacy
MAR	Missing at random
MCAR	Missing completely at random
MNAR	Missing non at random
OECD	Organization for Economic Co-operation and Development
OLS	Ordinary least squares
PCA	Principal component analysis
QoG	Quality of Government Institute
QOG	Quality of government index

SDG	Sustainable development goals
SFI	State fragility index
TI	Transparency International
UN	United Nations
UNU-WIDER	United Nations University World Institute for Development Economics Research
V-Dem	Varieties of Democracy Institute
VDEM	Impartial public officials indicator
WB	World Bank
WDI	World development indicators
WGI	Government effectiveness index

List of Figures

List of Tables

Chapter I

Introduction: Reviewing State Capacity

Definitions and Measures

Introduction

State capacity has become an increasingly important research topic in a variety of social science sub-fields. Especially in recent years, the study of state capacity, its causes, and its consequences have attracted a large number of scholars around the world. A quick search on the well-known Scopus[1] database provides compelling although crudely approximative evidence on the explosion of academic literature on the topic in the last decades. The term *state capacity* appears only 375 times in scholarly articles from 1980 to 2009 but as many as 1,060 times from 2010 onwards. The interest in broadly analogous terms such as *quality of government* and *good governance* follows a similar increase. The former appears in 42 articles from 1980 to 2009 and in 258 articles from 2010 on. The latter appears in 1,155 articles from 1980 to 2009 and in 3,330 articles from 2010 on. These numbers confirm that the star of state capacity has begun to shine brighter than ever in the just-ended decade.

Studying the state has become of substantial interest in itself. As put by Skocpol (2008: 122), "understanding state-building and the effects of state policies and structures have become enduringly central to the most powerful agendas of research and theorizing in the social sciences". In particular, the increased interest in the state has been boosted

---

[1] Accessed on 15 February 2021 (https://www.scopus.com/search/form.uri#basic). The search was restricted to titles, abstracts, and keywords of scientific articles indexed on Scopus.

by the widespread assumption that capable states with effective institutions are the key drivers of social, political, and economic development. For instance, nowadays, few scholars would disagree that state capacity is one of the key explanations to the gap in well-being between developed and underdeveloped countries. Hence, it is not surprising that the concept of state capacity has been embraced not only by academics and researchers but also by international institutions and the broader international development community.

That being said, it may be good to recall that "the questions of state capacity and state-building were largely absent from policy discussion in the late 1980s to early 1990s" (Fukuyama 2004: 17). The inclusion of the promotion of state capacity into the official development agenda of the United Nations (UN), however, is an evincive example of how topical the theme has become in the contemporary development debate. While the UN Millennium Declaration and its Millennium Development Goals, enacted in 2000, did not explicitly aspire to build well-functioning state institutions across countries around the world, conversely, one of the UN Agenda 2030 Sustainable Development Goals is precisely the creation of capable and effective state institutions (Goal 16). Today, building state capacity is one of the key priorities of the broader UN development framework as well as a prime concern of many specific UN development initiatives such as the UN Peacebuilding Fund (United Nations 2019).

Given the current appeal of state capacity it may seem strange that only a few decades ago the state was disregarded by most social scientists as well. As put by Krasner (1984: 223), "from the late 1950s until the mid-1970s the term state virtually disappeared from the professional academic lexicon". In the late 1960s, a seminal article that advocated the reintegration of the state as a primary concern for social scientists, said out loud that "the concept of the state is and ought to be treated as a variable in social science"

(Nettl 1968: 562). Still, even as recently as in the 1980s, some scholars went as far as suggesting that studying the state had become irrelevant. For instance, in 1986, Cassese (1986: 124) claimed that "it seems to me that the reason for this lack of interest [*in the state*] lies in the fact that there is no longer any need of it".[2] The tide has changed, however, since then. Today, we can assert quite confidently that indeed the state has been "brought back in" and it is likely to be here to stay.

In this opening chapter, I will present more thoroughly the main subject of my dissertation: state capacity. First, I will show the importance of the concept in many research agendas by briefly reviewing current empirical evidence on the causes and consequences of state capacity. As we shall see, ultimately, my dissertation speaks to all these fields of research and to anyone who uses cross-national measures of state capacity. Second, I will review some of the most common definitions of the state and state capacity. Third, I will review some of the recently used measures of state capacity and classify them according to the functional dimensions of state capacity that researchers have previously identified. Fourth and last, before moving on to the next chapters, I will present the structure of my dissertation.

The chapters of my dissertation take diverse approaches to the measurement of state capacity. Yet, the whole dissertation is dominated by a recurring "central" argument: research on state capacity is conditioned by how state capacity is quantified, and our understanding of state capacity depends on the data we have and use. Thus, to generate high-quality knowledge on the topic, we must acquire extensive information about existing data, and we need to be particularly careful in using and choosing valid measures of state capacity. Given the complexity of the concept, state capacity should be either

---

[2] Text in square brackets added by author.

disaggregated into its lower-level components and quantified with specific measures that leave few doubts on the concept-measure validity or aggregated in harmony with its higher-level dimensions and quantified with analytically useful multidimensional measures.

Even if each of the chapters contributes to slightly different strands of literatures, the dissertation as a whole seeks to speak to the comparative literature on measures of state capacity. In relation to other comparative studies on cross-national measures of state capacity (e.g., Fortin 2010; Hendrix 2010; Soifer 2012; Cingolani 2013; Hanson and Sigman 2013; Saylor 2013; Fortin-Rittberger 2014a; Savoia and Sen 2015; Hanson 2018; Ziaja, Grävingholt, and Kreibaum 2019), the study at hand pushes beyond in several important ways.

First, many of these studies focus primarily on conceptual issues. Empirical questions regarding measurement frequently remain undiscussed. My study addresses this shortcoming and focuses on measures of state capacity mainly through an empirical lens.

Second, my study provides one of the first systematic attempt to assess comprehensively questions related to different aspects of validity such as convergence, interchangeability, case-based discrepancy, and statistical dimensionality of measures of state capacity globally. Previously, some of these questions have been examined in a more limited number of countries (e.g., Fortin 2010) or with a different set of measures (e.g., Hendrix 2010). Yet, for instance, to my knowledge, no existing study has evaluated the interchangeability of commonly used measures of state capacity, or in other words, how the use of a particular measure affects research results.

Third, my study provides new practical solutions for users of measures of state capacity. I argue that a disaggregation of state capacity into its more specific components

facilitates ensuring that selected measures represent closely the selected concept. Claims favouring disaggregated measures of state capacity have been put forward previously (e.g., Savoia and Sen 2015), but practical examples on how to do so are in short supply. Chapter III provides an illustrative example on the advantages and disadvantages of such disaggregation.

I also develop an aggregate multidimensional measurement tool for researchers and policymakers who are interested in capturing the multidimensional concept of state capacity (or its higher-level dimensions) in its entirety. Somewhat similar tools with similar approaches have been created previously (e.g., Hendrix 2010; Hanson and Sigman 2013), yet the results of the study at hand provide completely new empirical evidence on the statistical dimensionality of state capacity. The new multidimensional measure developed in this study meets the requirements of multiple validity tests and represents particularly well some of the recent theoretical ideas on the core dimensions of state capacity.

To sum up, besides the more specific contributions of each chapter of my dissertation, which are discussed later on, this study as a whole contributes in particular to the comparative literature on measures of state capacity.

Existing Empirical Evidence on the Causes and Consequences of State Capacity

The causes and consequences of building state capacity across countries have been analysed in an abundant comparative social science literature.

As to the consequences, past research has shown that state capacity fosters economic growth (e.g., Evans and Rauch 1999; Dincecco 2015), prevents internal conflicts (e.g., Fearon and Laitin 2003; Fjelde and De Soysa 2009), reduces poverty (e.g., Asadullah and Savoia 2018), child deprivation (e.g., Halleröd et al. 2013), and infant mortality (Kim and

Kroeger 2018), improves public health outcomes (Holmberg and Rothstein 2011; Cingolani, Thomsson, and De Crombrugghe 2015), plays a key role in achieving both Millennium Development Goals (Joshi 2011) and Sustainable Development Goals (Asadullah, Savoia, and Sen 2020), increases welfare state generosity (e.g., Rothstein, Samanni, and Teorell 2012; Dahlström, Lindvall, and Rothstein 2013), enhances the quality and provision of public goods (D'Arcy and Nistotskaya 2017), increases respect for human rights (e.g., Englehart 2009; Cole 2016), strengthens regime stability (e.g., Andersen et al. 2014), boosts democratisation (e.g., Fortin 2012), and decreases the likelihood of electoral fraud (Fortin-Rittberger 2014b) among many other positive outcomes.

The comparative research agenda on the causes of state capacity is probably less developed than the one on its consequences. That being said, many scholars have analysed the causes of why some states have been able to build capable and effective institutions while others are trapped in a whirl of dysfunctional and weak institutions. To give a few examples on the causes of state capacity, empirical findings suggest that external conflicts (e.g., Besley and Persson 2008), external territorial threats (Gibler and Miller 2014), differences in colonial heritage (Acemoglu, Johnson, and Robinson 2001), democracy (e.g., Bäck and Hadenius 2008; Charron and Lapuente 2010; Carbone and Memoli 2015; Wang and Xu 2018), left-wing parties (Grassi and Memoli 2016a, 2016b), constraints on the executive (Ricciuti, Savoia, and Sen 2019), competition between elites (Beramendi, Dincecco and Rogers 2019), economic inequality (Cárdenas 2010), and common law systems (La Porta et al. 1999) are key determinants of well-functioning states around the world.

Furthermore, state capacity seems to play a particularly important role in conditioning the relationship between political regimes and various development

outcomes. For instance, it has been empirically shown that state capacity affects the relationship between democracy and income inequality (Soifer 2013), democracy and human development (Hanson 2015; Grundholm and Thorsen 2019), democracy and economic growth (Knutsen 2013), democracy and civil wars (Rossignoli 2016), democracy and natural disaster management (Persson and Povitkina 2017), democracy and human rights protection (Cole 2018), and democracy and workers' rights (Berliner et al. 2015).

In fact, the relationship between state capacity and democracy has become a thriving strand of literature on its own. To be more specific, the study of the state-democracy nexus is at the core of the so called "institutional sequencing" literature, which revolves around questions related to the association between the state and democracy, and in particular, to the causal "sequentiality" of the association. These questions are especially important for the study at hand, because in the next chapters of my dissertation I will use the relationship between state capacity and democracy as an illustrative framework for the analysis of the measurement of state capacity. Since the sequencing literature has a salient role in my dissertation, as a consequence, my dissertation contributes also to the sequencing literature.

We have seen above that state capacity has been analysed extensively in relation to a great number of factors. Given the premises on the overarching importance of state capacity, however, it can be surprising to realise how many obstacles there still are to make great strides in the field of research. To a certain extent these obstacles exist at the conceptual and theoretical levels, but in particular, they exist in the measurement of state capacity.

An abundant literature on the causes and consequences of state capacity has emerged. Sometimes we still disagree on very definition of state capacity, but more often

than not, we disagree on how state capacity should be measured. Even if the primary focus of my dissertation is the measurement of state capacity, before shifting my analysis to measurement issues, it can be useful to review some of the most relevant parts of the conceptual discussion on state capacity. Next, to make clearer what state capacity is and how it has been defined, I present some of the most common definitions of the state and state capacity.

On Definitions of the State and State Capacity

Without a state there cannot be state capacity. This means that the definition of state is pivotal to the definition of state capacity. Hence, as a first step, before understanding state capacity, it is essential to understand what a state is.

Arguably, the most common starting point to understand the state and its capacity is Weber's famous conception of the state, which has inspired virtually all definitions of the modern state. According to Weber (1946: 78), the state is "a human community that (successfully) claims the monopoly of the legitimate use of physical force within a given territory". An obvious implication of this commonly accepted definition of the state is that a capable state must have, at least up to a certain extent, such monopoly of the legitimate use of force to establish order within its borders. A state with no monopoly of the use of force within its territory cannot exert power nor enforce order, and in fact, is not even a state in the Weberian sense of the word.

Besides the monopoly of the use of force and the resulting capacity to enforce order within a given territory, a modern state needs an administrative apparatus to maintain power over time (Weber 1946). In modern states, according to Weber (1978), the ideal type of such administrative apparatus is a bureaucracy (i.e., competent, meritocratic, professional, and impartial public administrative organisation with a salaried long-term

career structure). A Weberian approach to the state indicates that without effective control over its territory and without a well-functioning administrative apparatus a state has a hard time to implement any of its decisions. Therefore, both the monopoly of the use of force within a given territory and a rational-legal authority in the form of bureaucracy must be at the core of any conceptualisation of state capacity.

Much of the well-known "pioneering" state capacity literature (e.g., Skocpol 1979, 1985; Mann 1984, 1986, 1988; Tilly 1975; Migdal 1988a, 1988b) is deeply rooted in these Weberian ideas.

Skocpol (1985) conceives the state as an organisation that has territorial control and can formulate policy goals autonomously. States operate by extracting resources from the society using these resources to build and maintain coercive and administrative organisations (Skocpol 1979). State capacity, which she understands as the ability of the state to implement policy goals formulated by the state, implies "sheer sovereign integrity and the stable administrative-military control of a given territory" (Skocpol 1985: 16). Besides Weber's essential coercive and administrative characteristics of the state, Skocpol brings in the basic functions of extracting and deploying resources, which are fundamental for any modern state. To effectively achieve policy goals, then, the state needs especially "loyal and skilled officials and plentiful financial resources" Skocpol (1985: 16). Recent conceptualisations on the dimensions of state capacity have often drawn on Skocpol's ideas (e.g., Hanson and Sigman 2013; Hanson 2018).

The importance of financial resources is emphasised also in Tilly's (1992) historical account on the formation of European states, although his famous statement "war made the state, and the state made war" (Tilly 1975: 42) is frequently associated with only coercive aspects of the state. Tilly (1992: 20) argues that large scale warfare required inevitably the extraction of resources from the people, and to extract these resources, in

turn, states needed to develop "an infrastructure of taxation, supply, and administration that requires maintenance of itself". Therefore, given the historical frequency of wars in Europe, the emergence of capable state institutions was almost an inevitable by-product of competition and survival. Tilly's account on state-formation in Europe stresses the interaction between different aspects of state. Coercive capacity might be essential, but so are the state's administrative and extractive abilities as well.

Mann (1984: 187) reaffirms the importance of "resource-supported administrative and coercive organizations" but tries to understand further the autonomous power of the state vis à vis civil society. Mann (2012: 5) defines power as "the capacity to get other to do things that they would otherwise not do" and argues that ultimately the power of states derives from the need of most societies to have monopolistically set rules. He distinguishes between two types state power: despotic and infrastructural. The former refers to the unilateral power of the state, which allows it to act without consultation with the civil society. The latter, instead, refers to "the capacity of the state to actually penetrate civil society, and to implement logistically political decisions throughout the realm" (Mann 1984: 189). Some of the most prominent recent accounts on state capacity (e.g., Slater 2010; Fortin 2012; Soifer 2012; Lindvall and Teorell 2016) anchor their definition of the concept to Mann's vision of infrastructural power.

Migdal (1988a) tries to provide an answer to why so many countries in the Global South have weak and dysfunctional state apparatuses, and in parallel to Mann, he argues that social control is the key to state strength. According to Migdal (1988b: 25), the ideal type of a state is "an organization with a monopoly over the principal means of coercion, autonomy from existing social groups in what rules to make, and the ability to get its population to follow those rules". To get a population to follow the rules set by the state, strong states need to have the capacity "penetrate society, regulate social relationships,

extract resources, and appropriate or use resources in determined ways" (Migdal 1988a: 4). For Migdal (1988a), these are ultimately the capabilities that states need to achieve changes in the society through its infrastructure.

A more recent but equally influential perspective on the origins of state capacity that insists on the importance of controlling the society is Scott's (1998) groundbreaking study on state "legibility" (i.e., the ability of the state to acquire information about the society). Without trying to define the concept of state capacity in itself, Scott recognises that a modern well-functioning state needs information about its citizens to tax its population or to enforce rules on its population. Legibility plays a key role for successful policy implementation and "an illegible society, then, is a hindrance to any effective intervention by the state" (Scott 1998: 78). By showing that legibility is an essential characteristic of capable states, Scott's account has inspired many scholars (Soifer 2013; D'Arcy and Nistotskaya 2017; Lee and Zhang 2017; Brambor et al. 2020) to develop measures of state capacity that capture different aspects related to legibility.

The so called "new institutional economics" literature reiterates the importance of state revenues and the capacity to collect taxes. As put by Levi and North (1982: 318), "a state is any organization with an absolute advantage in violence extending over a geographic area whose boundaries are determined by its power to tax constituents". Nevertheless, besides the fiscal functions of the state, new institutional economists call attention to the legal and regulatory functions of states. In particular, they point out the fundamental role of the modern state in protecting and enforcing property rights. In fact, according to Levi and North (1982: 320), "the most essential and universal benefit provided by a state is the arbitration of conflicts over property rights". In a simplified model of the state, thus, rulers offer protection and justice to the population in return for taxes (North 1990). By drawing on this literature, some of the most influential recent

accounts on state capacity focus especially on the legal and fiscal functions of the state (e.g., Besley and Persson 2008, 2009, 2011).

The main contribution of new institutional economists to the broader state capacity literature is the recognition that the effective protection of property rights and the provision of justice entail not only resolving conflicts among citizens but also limiting arbitrary behaviour by the state itself. If not constrained, rulers will try to extract as much revenue as possible from the population (Levi 1988; Olson 1993). In a seemingly insoluble dilemma, impartial law-enforcement requires a state that has the coercive force to effectively protect property rights and enforce contracts, but "if the state has coercive force, then those who run the state will use that force in their own interest at the expense of the rest of the society" (North 1990: 59). Thus, on the one hand, the state has coercive force by definition, but on the other hand, this coercive force must be constrained to ensure that the state commits to the neutral enforcement of the rules it has set.

If some scholars (e.g., O'Donnell 1993) stress that the state should not be conflated with its bureaucracy, others (e.g., Geddes 1994; Evans 1995) instead emphasise that the capacity of the state hinges primarily exactly on its administrative abilities. In theorising the political underpinnings of well-functioning developing states, Geddes (1994: 14) argues that "the capacity to implement state-initiated policies" depends ultimately on "the existence of effective bureaucratic organizations". In examining the characteristics of "developmental states", in addition to the "embeddedness" of state and society, Evans (1995) highlights the importance of a Weberian bureaucracy. According to Evans (1995) modern states are responsible for economic transformation and a Weberian bureaucracy is precisely the kind of organisation that can intervene effectively in the economy and constrain predatory rulers. Bureaucracy-centric views have encouraged many researchers to assert that administrative capacity is the most important dimension of state capacity

(e.g., Knutsen 2013; Cole 2016; Acemoglu and Robinson 2019). Some scholars have even argued that it would be conceptually better to understand state capacity exclusively as "the organizational and bureaucratic ability to implement governing projects" (Centeno, Kohli, and Yashar 2017: 3).

Rothstein and Teorell's (2008: 165) influential theory on quality of government as the "the impartiality of institutions that exercise government authority" draws upon some of the insights of the "new institutional economics" literature and the bureaucracy-centric accounts. According to the authors, the principle of impartiality in the exercise of public power entails the rule of law, but not only; it implies impartial procedures also in a broader set of state functions (Rothstein and Teorell 2008). For instance, the principle of impartiality rules out all forms of corruption, and in fact, it can be understood as the opposite of corruption (Rothstein 2014). "The idea about the impartial civil servant goes back to the most central figure in bureaucratic theory, namely Max Weber" (Rothstein and Teorell 2008: 173), and thus, the theory of quality of government is strictly rooted in the Weberian conception of bureaucracy. Rothstein and Teorell's intuition has inspired many scholars to consider impartiality (or the absence of corruption) a fundamental ingredient of state capacity (e.g., Charron and Lapuente 2010, 2011; Bizzarro et al. 2018; Grundholm and Thorsen 2019).

Rothstein and Teorell's (2008) proposal should be seen also as an answer to the all-embracing "good governance" agenda promoted by some of the key actors in the international development community. The World Bank (1994) identifies three distinct dimensions of governance: (1) the type of political regime, (2) the manner in which power in exercised for development, and (3) the policy implementation capacity. Such definition of governance has been criticised because it conflates the notion with both development and political regimes. As to conflating development and governance, Rothstein and

Teorell (2008) warn that if good governance is to be understood in a functionalist manner as good for economic development, it is impossible to conceptualise at a general level what good governance actually is, besides being just good for economic development. The type of political regime and governance should not be either equated with each other, for the simple reason that the relationship between the two is not straightforward (Rothstein and Teorell 2012).

In similar vein, Fukuyama (2013) stresses that the type of political regime should be excluded from the definition of governance because we might want to analyse the relationship between democracy and governance. Additionally, however, he points out an important distinction within the concept of governance, namely the difference between sound procedures and enforcement power. He defines governance as the "government's ability to make and enforce rules, and to deliver services" (Fukuyama 2013: 350) and posits that "it is possible to have an impersonal, merit-based bureaucracy that nonetheless is extremely poor at getting things done" (Fukuyama 2013: 353). Finally, in an apparent response to Rothstein and Teorell's idea of the principle of impartiality, Fukuyama (2013) argues that impartial procedures alone cannot guarantee that a state gets things done as intended, because a capable state needs also resources and a professional bureaucracy.

Some of the most interesting recent conceptual analyses on state capacity have elaborated on the aforementioned distinction between the power to implement policies and the procedural norms in the policy-implementation process. According to Lindvall and Teorell (2016: 5) state capacity should be understood solely in terms of "the projection and exercise of power", or in other words, "the state's ability to 'get things done'". Procedural "constraints" on the exercise of this power, on the contrary, should not be conflated with the power exercised by the state in the first place (Lindvall and Teorell 2016). Somewhat conversely, according to Sekeris (2015: 553), state power

should be understood as "the state's ability to broadcast power", whereas state capacity should be understood as "capacity of the state to collect taxes and enforce property rights".

Johnson and Koyama (2017) distinguish between power and procedural constraints as well. They argue that state capacity in itself has to do with the law enforcement and revenue raising activities of the state, but procedural constraints are the linkage between state capacity and economic performance (Johnson and Koyama 2017). D'Arcy and Nistotskaya (2017) posit that well-functioning states need to have both power and constraints to this power. In further disentangling the different aspects of the state, they point out the difference between enforcement power, procedural constraints or impartiality on the "output" side of the political process, and the type of political regime on the "input" side of the political process (D'Arcy and Nistotskaya 2021).

We have reviewed above some of the traditional and more recent approaches to the state and state capacity. As we have seen, Weber's influential ideas on the state have influenced more or less all the presented subsequent analyses on the state. Moreover, as we have noticed, in one way or another the "classic" accounts on state capacity provide a framework in which more recent conceptualisations of state capacity are anchored to. The above review should not be considered by any means an exhaustive assessment of definitions of state capacity. It should be considered instead as an illustrative sample and a good starting point for the next parts of my study.

To sum up some of the differences between the above accounts of state capacity, besides Weber's coercive and administrative preconditions, scholars such as Skocpol and Tilly highlight the importance of extractive functions of the state, scholars such as Mann, Migdal, and Scott focus on the ability of the state to control society, and new institutional economists stress especially the significance of impersonal and credible law enforcement.

More recent accounts focus in particular to the administrative functions of the state, to how the state exercises its implementation power, and to whether such power exists to begin with.

Table 1.1 provides a non-exhaustive sample list of definitions of state capacity in chronological order. While it becomes clear that there is no complete agreement on the exact definition of state capacity, the entries in the table indicate that virtually all the conceptualisations have to do in one way or another with the ability to implement intended policies or official policy goals either explicitly (e.g., Skocpol 1985; Sikkink 1991; Geddes 1994; Buhaug 2010; Soifer 2013; Knutsen 2013; Rogers and Weller 2014; Dincecco 2018; Acemoglu and Robinson 2019; Williams 2021) or implicitly (e.g., Levi 1988; Evans and Rauch 1999; Fukuyama 2004; Bäck and Hadenius 2008; Charron and Lapuente 2010; Mazzucca and Munck 2014; Acemoglu et al. 2020).

In addition to such ability to implement policies, in conceptualising state capacity, some scholars focus on other more specific factors like public goods provision (e.g., Levi 1988; Evans and Rauch 1999; Mazzucca and Munck 2014; Geloso and Salter 2020), transparency (Fukuyama 2004), effectiveness (Hulme, Savoia, and Sen 2015), sovereignty (Gibler and Miller 2014; Rossignoli 2016), efficiency and corruption (Charron and Lapuente 2010), appropriateness (Cronert and Hadenius 2020), the mobilisation of resources (Tikusis et al. 2015; Carment et al. 2015).

Some scholars add into their conceptualisation of state capacity some sort of requirements of social and/or economic development. For instance, according to Evans and Rauch (1999: 752) state capacity has to do with the "ability to deliver the collective goods that constitute the state's *potential contribution to economic growth*". For Norris (2012: 15) state capacity "is the ability of regime authorities to perform functions essential for *collective well-being*, including, most essentially, maintaining security and

Table 1.1. Sample list of definitions of state capacity

Definition of state capacity	Source
"The capacity of the state to actually penetrate civil society, and to implement logistically political decisions throughout the realm."	Mann (1984: 189)
"The 'capacities' of states to implement official goals."	Skocpol (1985: 9)
"Capacity to exert control over local economic resources."	Evans (1985: 194)
"State capacity – the differential ability of states to assert control over political outcomes."	Ikenberry (1986: 106)
"The ability of government to penetrate society and extract resources."	Rouyer (1987: 453)
"The capacities to *penetrate* society, *regulate* social relationships, *extract* resources, and *appropriate* or use resources in determined ways."	Migdal (1988: 4)
"The state's ability to provide collective goods."	Levi (1988: 1)
"The ability of state institutions to effectively implement official goals."	Sikkink (1991: 11)
"The manner in which power is exercised in the management of a country's economic and social resources for development."	World Bank (1992: 1)
"The capacity to implement state-initiated policies."	Geddes (1994: 14)
"The ability of the state to undertake collective actions at least cost to society."	World Bank (1997: 77)
"The ability to deliver the collective goods that constitute the state's potential contribution to economic growth."	Evans and Rauch (1999: 752)
"The degree of control that state agents exercise over persons, activities, and resources within their government's territorial jurisdiction."	McAdam, Tarrow, and Tilly (2001: 78)
"The ability of states to plan and execute policies and to enforce laws cleanly and transparently."	Fukuyama (2004: 7)
"The ability to perform appropriate tasks effectively, efficiently and sustainably."	Hildebrand and Grindle (2005: 445)
"The extent to which interventions of state agents in existing non-state resources, activities, and interpersonal connections alter existing distributions of those resources, activities, and interpersonal connections as well as relations among those distributions."	Tilly (2007: 16)
"Impartial institutions and a solid capacity to develop, legislate, and implement effective policies."	Carothers (2007: 19)
"The capacity of the state organs to maintain sovereignty" and "how well the state organs are able to carry out their tasks."	Bäck and Hadenius (2008: 3)
"The impartiality of institutions that exercise government authority."	Rothstein and Teorell (2008: 165)
"The capability of the state to provide public goods in diverse parts of the country."	Acemoglu (2008: 1)
"The willingness and capability of the state apparatus to carry out government policy."	Englehart (2009: 167)
"The ability of a state to perform its activities in an *efficient way* and *without corruption*."	Charron and Lapuente (2010: 443)
"State capacity refers to the regime's ability to implement preferred policies."	Buhaug (2010: 109)
"The institutional capability of the state to carry out various policies that deliver benefits to households and firms."	Besley and Persson (2011: 6)

"The ability of regime authorities to perform functions essential for collective well-being, including, most essentially, maintaining security and managing the delivery of public goods and services, measured by the quality of bureaucracy, lack of corruption, and rule of law."	Norris (2012: 15)
"A government's ability to make and enforce rules, and to deliver services, regardless of whether that government is democratic or not."	Fukuyama (2013: 350)
"The ability of the state to effectively implement its chosen policies."	Soifer (2013: 2)
"The ability of state managers to penetrate society by extracting resources, exercising coercion, and deploying authority via political alliances, the military, or an administrative apparatus."	Saylor (2013: 370)
"The ability of the state to project its power throughout its territory."	Bell et al. (2013: 241)
"Capacity of state institutions to implement public policies."	Knutsen (2013: 1)
"We use the concept of state capacity to stand for a range of capabilities that are needed for the state to function effectively."	Besley and Persson (2014: 932)
"State capacity to deliver certain public goods."	Mazzucca and Munck (2014: 1233)
"The ability of a state to implement public policy."	Rogers and Weller (2014: 185)
"Capacity, imperfectly defined, includes the ability of a state to perform essential functions such as, but not exclusively, the ability to extract taxes and obtain census information."	Rotberg (2014: 514)
"The ability of the government of enforce its sovereignty across all its lands."	Gibler and Miller (2014: 635)
"The extent to which the state possesses the coercive and administrative capacities to penetrate and regulate society and extract and appropriate resources."	Andersen et al. (2014: 1310)
"State capacity – the extractive and productive power of states."	Dincecco (2015: 901)
"The ability of a state to provide its citizens with basic life chances."	Grävingholt, Ziaja, and Kreibaum (2015: 1290)
"The capacity to enforce law and order, regulate economic activity, and provide public goods."	Acemoglu, García-Jimeno, and Robinson (2015: 2364)
"To effectively and efficiently deliver government output."	Joshi, Hughes and Sisk (2015: 287)
"The power of leaders to mobilize resources for productive and defensive purposes."	Tikusis et al. (2015: 567)
"Potential for a state to mobilise and employ resources towards productive ends."	Carment et al. (2015: 1317)
"The capacity of the state to implement rules and policies effectively."	Hulme, Savoia, and Sen (2015: 87)
"Quality of a state's bureaucratic institutions."	Cole (2016: 2)
"The power of the state's agents to get members of society to do things that they would not otherwise do."	Lindvall and Teorell (2016: 9)
"The effectiveness of state sovereignty over its territory and population."	Rossignoli (2016: 427)
"State capacity is a multidimensional concept that encompasses not only extractive abilities of a state but also economic development and quality of institutions."	Rudolfsen (2017: 6)
"We take state capacity to mean the organizational and bureaucratic ability to implement governing projects."	Centeno, Kohli and Yashar (2017: 3)
"The ability of a state to collect taxes, enforce law and order, and provide public goods."	Johnson and Koyama (2017: 2)

"The state's ability to effectively exercise authority, as reflected in its ability to implement its chosen policies throughout the territory over which it claims to rule."	Luna and Soifer (2017: 888)
"The ability of states to solve collective action problems by efficiently monitoring individual contributions to collective endeavours and punishing free-riding."	D'Arcy and Nistotskaya (2017: 1)
"Human and physical capital investments that led to the creation of a bureaucratic authority, independence, efficacy, and penetration within a national territory."	Bustikova and Corduneanu-Huci (2017: 281)
"The state's ability to accomplish its intended policy actions."	Dincecco (2018: 2)
"The institutional capability of states to deliver policies benefiting their citizenry."	Asadullah and Savoia (2018: 71)
"It refers, at minimum, to a state's ability to maintain its territorial monopoly on the legitimation of force within fixed borders, and more generally, to the capability of the state's administrative apparatus to carry out its core tasks."	Cappelen and Sorens (2018: 195)
"State capacity is the power and capability to use state resources to implement political initiatives."	Walther, Hellström, and Bergman (2019: 584)
"The institutional capability of the state to carry out various policies that deliver benefits to households and firms."	Ricciuti, Savoia, and Sen (2019: 1)
"The state's ability to effectively implement policies throughout the state's territory."	Grundholm and Thorsen (2019: 382)
"The capacity of a state is its ability to achieve its objectives."	Acemoglu and Robinson (2019: 12)
"The ability of governments to raise tax revenue and execute the tasks it sets out to accomplish."	Murphy (2019: 936)
"The demonstrated ability of the state to provide basic services to its population."	Ziaja, Grävinholt, and Kreibaum (2019: 303)
"State capacity refers to the state's physical and logistic ability to penetrate society and implement decisions."	Andersen and Doucette (2020: 2)
"State capacity can be viewed as the product of a strong fiscal administration allowing the state to produce public goods."	Geloso and Salter (2020: 374)
"The capacity of a state is defined by its ability to set and enforce collectively binding rules."	Dimitrova et al. (2020: 1)
"The state's ability to take action in an appropriate and effective way."	Cronert and Hadenius (2020: 7)
"The degree to which a national state has achieved the Weberian monopoly of violence over its territory, developed the authority and the capacity to enforce laws, maintain law and order, and raise taxes and provide public goods."	Acemoglu et al. (2020: 749)
"The ability of the state bureaucracy to implement government's policy choices."	Williams (2021: 339)

Note: Entries in the table include definitions of state capacity and closely related terms such as state capability and quality of government, which are often used interchangeably among one another.

managing the delivery of public goods and services". Rudolfsen (2017: 123) argues that "state capacity is a multidimensional concept that encompasses not only extractive abilities of a state but also *economic development* and quality of institutions". For Grävingholt, Ziaja and Kreibaum (2015: 2364) it has to do with "the ability of a state to provide its citizens with *basic life chances*". For Besley and Persson (2011: 6) state capacity is "the institutional capability of the state to carry out various policies that deliver *benefits and services to households and firms*". For Tikusis et al. (2015: 567) the capacity of the state "refers to the power of leaders to mobilize resources for *productive and defensive purposes*".

Despite some differences, the reviewed definitions do provide us an understanding of what state capacity is at its core. Most scholars working on the topic would arguably agree – at least in principle – with Acemoglu and Robinson (2019: 12), according to whom "the capacity of the state is its ability to achieve its objectives". One can imagine, however, that even if everyone would agree on such definition, its accurate operationalisation is not an easy task by any means. Next, we shift our attention to how state capacity has been measured in recent studies. As we shall see, plenty of different measures have been used to quantify state capacity.

On Measures of State Capacity

State capacity has been operationalised and measured in countless ways in comparative social science literature. The use of different measures to capture the same concept is not exceptional, since even more established social science concepts such as democracy have been measured with various different indicators. Hendrix (2010), Cingolani (2013), Saylor (2013), Fortin-Rittberger (2014a), Savoia and Sen (2015), and Hanson (2018) provide comprehensive reviews of measures that have been used to quantify state

capacity. All these reviews categorise measures of state capacity according to different functions of the state.

Hendrix (2010) maps measures of state capacity into three categories: military capacity, administrative capacity, and political institutional quality. Cingolani (2013) maps measures of state capacity into seven categories: coercive, fiscal, administrative, transformative/industrialising, relational/territorial, legal, and political. According to Saylor (2013), measures of state capacity can be classified by four conceptual dimensions: extractive capacity, penetrative capacity, coercive power, and administrative scope. Fortin-Rittberger (2014a) divides measures of state capacity into two branches: infrastructural capacity and coercive capacity. Savoia and Sen (2015) map measures of state capacity into five empirically relevant dimensions of state capacity: administrative capacity, legal capacity, infrastructural capacity, fiscal capacity, and military capacity. Hanson (2018) defines state capacity as a tri-dimensional concept, consisting of coercive, extractive, and administrative capacity, and discusses a number of measures that capture these three dimensions.

The "field" of measures of state capacity evolves rapidly. Hence, it can be useful to review measures of state capacity that have been recently employed in cross-national comparative social science literature. I classify these measures according to Savoia and Sen's (2015) functional dimensions, which seem to reflect well the most commonly thought dimensions of state capacity. These functional dimensions are (1) administrative capacity, (2) legal capacity, (3) infrastructural capacity, (4) fiscal capacity, and (5) military capacity. In Savoia and Sen's (2015: 442–443) classification, administrative capacity concerns the state's bureaucratic apparatus, legal capacity concerns the rule of law, infrastructural capacity concerns the "territorial reach of the state", fiscal capacity concerns the "state's ability to raise revenues", and military capacity concerns the

effective establishment of the state's monopoly on the use of force. In referring to this last dimension, I will use the terms military capacity and coercive capacity interchangeably. I will also use the terms fiscal capacity and extractive capacity interchangeably.

Besides the above mapping into functional dimensions, I will classify the listed measures according to the type of underlying data (subjective, objective, or both) and their scope. The latter is classified into three categories following Fabra Mata and Ziaja's (2009) distinction of indicators of state fragility into single proxies, unidimensional (aggregate) indices, and multi-dimensional (composite) indices. As one can expect, unidimensional indices capture only one dimension of state capacity, whereas multi-dimensional indices cover multiple dimensions of state capacity. Table 1.2 provides a non-exhaustive sample of recently used measures of state capacity.

As we have previously discussed, according to some scholars, the administrative dimension of state capacity is the most important one. Indeed, without a functioning bureaucracy the state is not likely to be able to implement successfully its intended policies nor reach its objectives. International Country Risk Guide's *Bureaucracy Quality* (e.g., Knutsen and Fjelde 2013; Bauhr and Grimes 2017; Franco Chuaire et al. 2017; Reinsberg et al. 2019), Bertelsmann Stiftung's *Basic Administration* (e.g., Carbone and Memoli 2015), Worldwide Governance Indicators' *Government Effectiveness* (e.g., Cárdenas 2010; Charron and Lapuente 2010, 2011; Böhmelt et al. 2019), and Transparency International's *Corruption Perceptions Index* (e.g., Englehart 2009; Lin 2015) are some of the most commonly used measures that capture mainly the capacity and quality of the administrative functions of the state. Some procedural measures from the Varieties of Democracy dataset (V-Dem) focus also on administrative aspects of the

Table 1.2. Sample list of recently used measures of state capacity

Measure	Producer	Used in (e.g.,)	Dimension(s)	Type of data	Scope
Authority over population	Varieties of Democracy	Rossignoli (2016)	Coercive Infrastructural	Subjective	Single
Authority over territory	Varieties of Democracy	Seeberg (2019)	Coercive Infrastructural	Subjective	Single
Basic administration	Bertelsmann Stifstung	Carbone and Memoli (2015)	Administrative	Subjective	Single
Bureaucratic quality	ICRG (PRS Group)	Bauhr and Grimes (2017)	Administrative	Subjective	Single
Bureaucratic autonomy	Cingolani, Thomsson, and De Crombrugghe (2015)	Nistotskaya and Cingolani (2016)	Administrative	Objective	Aggregate
Cadastral index	D'Arcy and Nistotskaya (2017)	D'Arcy and Nistotskaya (2018)	Infrastructural	Objective	Single
Capacity	Hanson and Sigman (2013)	Grassi and Memoli (2016a)	Administrative Legal Extractive Infrastructural Coercive	Both	Composite
Census	Soifer (2013)	Soifer (2013)	Infrastructural	Objective	Aggregate
Constraints on the executive	Center for systemic peace	Asadullah and Savoia (2018)	–	Subjective	Single
Contextualized index of statehood	Schlenkrich, Lemm, and Mohamad-Klotzbach (2016)	Schlenkrich, Lemm, and Mohamad-Klotzbach (2016)	Administrative Coercive	Both	Composite
Contract enforcement	Doing Business (WB)	Besley and Persson (2014)	Legal	Subjective	Single
Contract intensive money	Clague et al. (1999)	Fortin-Rittberger (2014a)	Legal	Objective	Single
Control of corruption	World Bank	Melville and Mironyuk (2016)	Administrative Legal	Subjective	Composite
Corruption perceptions index	Transparency International	Lin (2015)	Administrative Legal	Subjective	Composite
Direct taxes/GDP	International Monetary Fund	Dincecco and Prado (2012)	Extractive	Objective	Single
Expropriation risk	ICRG (PRS Group)	Cárdenas (2010)	Legal	Subjective	Single
External intervention	Fund for Peace	D'Arcy and Nistotskaya (2017)	Coercive	Subjective	Single
Fragile states index	Fund for Peace	Møller and Skaaning (2011b)	Administrative Legal Extractive Infrastructural	Both	Composite

			Coercive		
GDP/capita	World Bank (WDI)	Ha and Rogers (2017)	–	Objective	Single
Governance capacity index	Joshi, Hughes, and Sisk (2015)	Joshi, Hughes, and Sisk (2015)	Administrative Extractive	Both	Composite
Government effectiveness	World Bank	Serikbayeva, Abdulla, and Oskenbayev (2020)	Administrative Infrastructural	Subjective	Composite
Impartial public officials	Varieties of Democracy	Grundholm and Thorsen (2019)	Administrative	Subjective	Single
Implementation and enforcement	Berkman et al. (2009)	Cárdenas (2010)	Administrative Legal Fiscal	Subjective	Composite
Income tax revenue/GDP	Baunsgaard and Keen (2010)	Cárdenas (2010)	Extractive	Objective	Single
Income taxes/total taxes	Rogers and Weller (2014)	Rogers and Weller (2014)	Extractive	Objective	Single
Index of state capacity	Fortin (2010)	Fortin (2012)	Administrative Legal Infrastructural Extractive	Both	Composite
Information capacity	Brambor et al. (2020)	Brambor et al. (2020)	Infrastructural	Objective	Aggregate
Law and order	ICRG (PRS Group)	Dietrich and Bernhard (2015)	Legal Coercive	Subjective	Composite
Legibility index	Lee and Zhang (2017)	Lee and Zhang (2017)	Infrastructural	Objective	Single
Military personnel/capita	Correlates of War	Wang and Xu (2018)	Coercive	Objective	Single
Military spending/capita	Correlates of War	Andersen et al. (2014)	Coercive	Objective	Single
Millennium governance index	Joshi (2011)	Joshi (2011)	Administrative Extractive	Both	Composite
Monopoly of violence	Bertelsmann Stifstung	Carbone and Memoli (2015)	Coercive	Subjective	Single
Night-time luminosity	Koren and Sarbahi (2018)	Koren and Sarbahi (2018)	Infrastructural	Objective	Single
Non-resource taxes/GDP	Mansour (2014)	Cappelen and Sorens (2018)	Extractive	Objective	Single
Political corruption	Varieties of Democracy	Povitkina and Bolkvadze (2019)	Administrative Legal	Subjective	Composite
Political stability	World Bank	Melville and Mironyuk (2016)	Coercive	Subjective	Aggregate
Public sector corruption	Varieties of Democracy	Wang and Xu (2018)	Administrative	Subjective	Aggregate
Quality of government index	Quality of Government Institute	Charron and Lapuente (2011)	Administrative Legal	Subjective	Composite

			Coercive		
Regulatory quality	World Bank	Melville and Mironyuk (2016)	–	Subjective	Aggregate
Relative political capacity	Arbetman-Rabinowitz and Johnson (2008); Kugler and Tammen (2012)	Wang and Xu (2018)	Extractive	Objective	Aggregate
Relative political reach	Kugler and Tammen (2012)	Ha and Rogers (2017)	Infrastructural	Objective	Aggregate
Rule of law	World Bank	Melville and Mironyuk (2016)	Legal	Subjective	Aggregate
State antiquity index	Bockstette, Chanda, and Putterman (2002)	Hanson (2015)	–	Objective	Single
State capability index	Andrews, Pritchett, and Woolcock (2017)	Andrews, Pritchett, and Woolcock (2017)	Administrative Legal Infrastructural Coercive	Subjective	Composite
State capacity index	Besley and Persson (2014)	Besley and Persson (2014)	Legal Fiscal	Both	Composite
State fragility index	Center for Systemic Peace	Hiilamo and Glantz (2015)	Coercive	Both	Composite
State weakness index	Rice and Patrick (2008)	Dincecco (2015)	Administrative Legal Infrastructural Coercive	Both	Composite
Total tax revenue/GDP	Baunsgaard and Keen (2010)	Cárdenas (2010)	Extractive	Objective	Single

Note: This is a non-exhaustive list of recently (>= 2010) used cross-national measures of state capacity in the relevant literature. Measures that have been used exclusively as sub-components of composite/aggregate indices are excluded.

state. For instance, Grundholm and Thorsen (2019) operationalise state capacity with *Rigorous and impartial public officials* from V-Dem.

Many other scholars suggest that the extractive dimension of state capacity is the most important one, because without any revenues the state has a hard time to perform any of its intended activities. Objective and very specific measures capturing different forms of tax extraction such as *total taxes/GDP* (e.g., Besley and Persson 2011; Dincecco and Prado 2012; Andersen et al. 2014; Seeberg 2018), *income taxes/GDP* (e.g., Cárdenas 2010; Dincecco and Katz 2016), *direct taxes/GDP* (e.g., Dincecco and Prado 2012), *non-resource taxes/GDP* (Cappelen and Sorens 2018) and *income taxes/total taxes* (Rogers and Weller 2014) have been used as proxies of a state's extractive capacity. *Relative Political Capacity* index, developed by Arbetman-Rabinowitz and Johnson (2008) and Kugler and Tammen (2012), measures the difference between actual levels of tax extraction and predicted levels of tax extraction and has been used in numerous studies as a measure of state capacity (e.g., Braithwaite 2010; Wang and Xu 2018).

The legal dimension of state capacity refers to the state's ability to enforce laws and contracts. In particular, the protection of property rights is often seen as the most important "contract" to safeguard. The legal functions of the state have been quantified with subjective measures of property rights protection, such as ICRG's *Expropriation of Property* (e.g., Cárdenas 2010; Besley and Persson 2011) and Heritage Foundation's *Property Rights Enforcement* (e.g., Fortin 2010), or more generally the rule of law such as Worldwide Governance Indicators' *Rule of Law* (e.g., Melville and Mironyuk 2016). Clague et al.'s (1999) *Contract-Intensive Money* index, computed as the ratio of non-currency money to the total money supply, is often considered as an objective measure of property rights protection and contract security. It has been used in several studies on

state capacity (e.g., Fjelde and De Soysa 2009; Fortin-Rittberger 2014a) to quantify the legal capacity of the state.

The coercive or military dimension captures another commonly acknowledged basic function of the state: the state's ability to maintain political order, to provide security to its citizens, and ultimately, to retain the monopoly on the physical use of force. Subjective measures with a focus on the coercive dimension include indicators such as Bertelsmann Stiftung's *Monopoly on the Use of Force* (e.g., Carbone and Memoli 2015) and Worldwide Governance Indicators' *Political Stability and Absence of Violence* (Melville and Mironyuk 2016). Objective measures of coercive capacity instead focus on different aspects related to the army. Data from the Correlates of War, Stockholm International Peace Research Institute, and World Development Indicators databases regarding *military expenditures* (e.g., Fjelde and De Soyza 2009; Andersen et al. 2014), *military personnel* (e.g., Fjelde and De Soyza 2009; Buhaug 2010; Cunningham 2011; Gibler and Miller 2014) have been often used to capture coercive aspects of state capacity.

The infrastructural dimension refers to the state's ability to "penetrate" society and project power throughout its territory. Many recent scholarly attempts to quantify state capacity in terms of its ability collect and process information about its citizens can be categorised in this group of measurements. To give a few examples, Soifer (2013) measures state capacity with the regularity of census administration; Lee and Zhang (2017) develop the *Legibility Index*, which is a measure of state capacity based on the accuracy of age data; D'Arcy and Nistotskaya (2017) collect data on cadastral maps and based on this data develop a measure of historical state capacity; Brambor et al. (2020) measure the information capacity of the state with an aggregate index based on data on national censuses, civil registers, population registers, the existence of statistical agencies, and the publications of statistical yearbooks. Some scholars have used other

"facts-based" proxies such as the *share of urban population* (Cole 2015) and *night-time luminosity* (Koren and Sarbahi 2017) to quantify the state's territorial reach.

Besides instruments that focus mainly on only one of these dimensions of state capacity, there are various measures that capture multiple dimensions of state capacity. Single perception-based proxies such as V-Dem's *State Authority over Population* (Rossignoli 2016) and *State Authority over Territory* (Seeberg 2019) or ICRG's *Law and Order* (e.g., Dietrich and Bernhard 2015) seem to cover more than just one single dimension of state capacity. The first two indicators capture at least aspects related to the coercive and infrastructural dimensions, whereas the last one captures aspects related to the coercive and legal dimensions of state capacity. Other widely used measures of state capacity such as Teorell et al.'s (2019) *Quality of Government Index* and Hanson and Sigman's (2013) *Capacity* are based on multiple sub-indicators concerning multiple functions of the state.

Some multidimensional measures may even capture an excessively broad understanding of state capacity, undermining their analytical utility in many relevant research agendas. Some of the broadest objective measures of state capacity have been based in one way or another on the state's age. One example is Bockstette, Chanda, and Putterman's (2002) *State Antiquity Index*, which has been widely used as a general proxy of state capacity (e.g., Hanson 2015; Asadullah and Savoia 2018; Bizzarro et al. 2018). Some other extremely broad measures synthesise both objective and subjective data. For instance, indices close to the "realm" of state fragility that have been used to measure state capacity, such as the Fund for Peace's *Fragile States Index*, Center for Systemic Peace's *State Fragility Index*, Rice and Patrick's *Index of State Weakness*, fit well in the latter category of measures.

Almost none of the above reviewed measures were originally created to quantify the concept of state capacity. To cope with this problem, a few scholars have tried to develop more "theoretically grounded" measures to capture the multidimensional concept of state capacity. Arguably the most well-known effort in this sense is the already mentioned Hanson and Sigman's (2013) index of state capacity, which is based on over twenty extant measures from a variety of sources. Anyhow, there have been also other attempts to better capture state capacity in its entirety.

To give some illustrative examples, inspired by Mann's idea of infrastructural power, Fortin (2010) creates an index of infrastructural state capacity by aggregating five indices: taxes as a share of GDP, property rights protection, contract intensive money, corruption, and infrastructural reform. Joshi (2011) creates a composite index of state capacity by aggregating six different indices regarding public revenue collection, public administration, and public service delivery. Besley and Persson (2014) aggregate income taxes/GDP, contract enforcement, school attainment, and life expectancy into an index of state capacity. Norris (2012) creates an index of state capacity by retaining the latent factor of some of the Worldwide Governance Indicators and the Quality of Government index. Bäck and Hadenius (2008) aggregate two indicators from the International Country Risk Guide (ICRG) dataset: Bureaucratic quality and Corruption. Berliner et al. (2015) and Cole (2016) instead create an index of bureaucratic capacity with three indicators from ICRG: Bureaucratic quality, Corruption, and Law and order. These three indicators are also the same that have been compiled by Teorell et al. (2019) in the already mentioned *Quality of Government index*.

My review of recently used measures of state capacity has shown that researchers working on the topic have plenty of measures to choose from. In theory, this is positive. Especially so, if users of the measures have all the relevant information that allows them

to pick out rationally a valid measure that closely represents their theory or matches their conceptualisation. In practice, however, this is seldom the case. The multitude of measures of state capacity can be then a problem for the generation and accumulation of knowledge on the topic. In particular, this is true if "datasets that supposedly represent the same thing are used to support different interpretations" (Jerven 2013: 119). Hence, if a particular measure is not chosen on clear-cut theoretical grounds, it becomes essential to understand comprehensively how equivalent these measures are and whether they can really lead to different interpretations or not.

In the next chapter, to address the above problem, I will evaluate some of the most established measures of state capacity and analyse their similarities and differences from multiple perspectives. Before moving on from the introductory chapter at hand to Chapter II, however, I briefly present the structure of my dissertation and in particular the structure of the next chapters.

Structure of the Dissertation

The objective of this first chapter has been to provide the reader a basic understanding on the state, state capacity, its measures, and the importance of the topic. I have reviewed some of the most common conceptualisations of state capacity and presented some of the differences between one definition and another. Furthermore, I have provided a review of recently used measures of state capacity and introduced also some of the less widely used measures created by researchers working on the topic.

In the next chapter of my dissertation, I will shift the attention to a mainly quantitative but also qualitative analysis of some of the most established measures of state capacity. The primary objective of Chapter II is to examine the advantages and disadvantages of frequently used measures and to assess their equivalence, in terms of

content, statistical properties, convergence, interchangeability, and individual country scores. As a by-product of this analysis, through a set of replications, we will gain new knowledge also on the relationship between state capacity and democracy. Chapter II identifies some of the problems concerning measures of state capacity that are addressed subsequently in Chapters III and IV.

To be more specific, in the third and fourth chapters of my dissertation, I provide effective solutions to some of the problems identified in Chapter II. In the third chapter, I argue that one solution to some of the identified measurement problems is to start analysing disaggregated aspects of state capacity. The primary aim of Chapter III is to examine the utility of such disaggregation, both in terms of measurement and theory, through a practical example on the state-democracy nexus. To test the main argument, I disaggregate the two concepts and conduct an analysis of the relationship between two specific aspects of state capacity and democracy, namely impartial public officials and civic participation.

In Chapter IV, I argue that another solution to some of the measurement problems identified in Chapter II is to develop a novel multidimensional measure of state capacity. The aim of the fourth chapter, thus, is to create a new set of state capacity measures that has some advantages compared to the commonly used measures. The novel measure(s) are scrupulously built, validated, and tested through empirical examples. My approach to the development of the new measure(s) combines theoretical knowledge and data-driven information. As a by-product of the analysis conducted in Chapter IV, we are able to gain interesting knowledge on the "true" statistical dimensions of state capacity.

Finally, in Chapter V, I present the conclusions of my dissertation. In particular, I provide a summary of the main findings, a discussion of the policy implications of my study, and set forth potential avenues for future research on state capacity and beyond.

Chapter II

Measuring State Capacity

An Assessment of Frequently Used Cross-National Measures

Introduction

The introductory chapter of this dissertation has shown that state capacity has been defined and measured in several different ways. State capacity remains a somewhat contested concept, but in the end, most social science concepts are contested. Hence, it should not take our breath away to recognise that there are some conceptual disagreements on state capacity; conceptual disagreement is common in the social sciences, and state capacity is no exception.

Despite some conceptual disagreements, we have seen in Chapter I that numerous scholars seem to agree that state capacity has to do at minimum with the ability of the state to achieve its policy goals. We have seen as well in Chapter I that abundant cross-national comparative empirical work shows that well-functioning state institutions are associated to various social, political, and economic issues. Despite proliferating quantitative cross-national work on state capacity, however, the analysis and comparison of cross-national measures of state capacity remains largely overlooked. This is true especially when it comes to the empirical evaluation of these measures. Existing comparative considerations of measures of state capacity rely mainly on qualitative techniques (Soifer 2012; Cingolani 2013; Saylor 2013; Hanson 2018) – with some notable exceptions (Hendrix 2010; Fortin-Rittberger 2014a; Savoia and Sen 2015) – and are predominantly devoted to conceptual issues.

There is no doubt that, overall, measures of other related or similar concepts such as good governance (e.g., Knack, Kugler, and Manning 2003; Arndt and Oman 2006; Thomas 2010; Bersch and Botero 2011, 2014), human rights (e.g., Bollen 1986; Barsh 1993; Landman 2002, 2004; Skaaning 2009), rule of law (e.g., Skaaning 2010; Møller and Skaaning 2011a, 2014; Versteeg and Ginsburg 2017), and corruption (e.g., Knack 2007; Treisman 2007; Donchev and Ujhelyi 2014; Charron 2016; Chabova 2017; Qu et al. 2019) have been analysed and compared much more extensively. In particular, there is an abundant literature on the measurement of democracy (e.g., Bollen 1980, 1993; Gleditsch and Ward 1997; Bollen and Paxton 2000; Elkins 2000; Casper and Tufis 2003; Bowman, Lehoucq, and Mahoney 2005; Hadenius and Teorell 2005; Knutsen 2010; Pemstein, Meserve, and Melton 2010; Giebler 2012; Högström 2013; Pickel, Stark, and Breustedt 2015; Elff and Ziaja 2018; Skaaning 2018; Boese 2019; Vaccaro 2021).

By drawing on the above studies, the chapter at hand contributes to the nascent literature on the measurement of state capacity and provides one of the first systematic comparative assessments of some of the most established cross-national measures of state capacity. My research strategy relies partly on Vaccaro's (2021) recent comparative study of democracy indices, in which measures are evaluated for their statistical properties, convergent validity, and interchangeability. In addition to these three criteria, I evaluate qualitatively the content validity of the selected measures and analyse quantitatively possible country-level discrepancies in scores across the measures. The chosen research strategy, thus, focuses more on the empirical implications of these measures rather than the data-generation process itself.

A comparative assessment of common measures of state capacity is a valuable task in itself, for the simple reason that our empirical knowledge about the similarities, divergencies, and possible shortcomings of these measures is limited. Increasing such

knowledge will also help users of measures of state capacity to choose the most appropriate measure for their needs. Therefore, to begin with, my research contributes to reduce the "informational uncertainties" (Mudde and Schedler 2010: 410) faced by the users of these measures.

Nonetheless, my analysis has also additional implications for the broader research agenda on the topic, because "we often lack the concrete knowledge of how the specific measures we select affect the empirical inferences we draw" (Mudde and Schedler 2010: 410). More specifically, if common measures of state capacity are empirically similar to each other, scholars can be ensured that selecting one measure instead of another is not likely to cause major consequences for their research results. If there are consistent dissimilarities among the measures, however, our alarm bells should start ringing. If every measure tells a different story and if statistical inferences are not generalisable across measures, selecting one measure or another becomes of crucial importance. Furthermore, if measures lead systematically to different conclusions, it becomes well-founded to question the external validity of extant findings on the topic. For these reasons, ultimately, the chapter at hand provides essential guidance for future quantitative research on state capacity.

Last, I want to emphasise that the prime purpose of this chapter is not to contribute to the conceptual literature on state capacity. The reasons of such a choice are twofold. First, I follow the advice of Adcock and Collier (2001: 533), according to whom "arguments about the background concept and those about validity can be addressed adequately only when each is engaged on its own terms". To make sure that the two issues are not conflated, I desist from discussing the "background concept" in itself. Second, assessing commonly used measures of state capacity is critical because our knowledge on state capacity is affected by how it is quantified. Concepts are important, but so are

numbers. Yet, the literature on these numbers is scarce. As well put by Jerven (2013: 119), "scholars pay great attention to defining the concepts ad devote great effort to theorizing the existence of the phenomenon and spend comparatively little time critically probing the numbers that are supposed to represent them". The primary objective of this chapter is thus to analyse the numerical information that is commonly used to represent state capacity.

Explaining Apples but Quantifying Oranges?

By reviewing a sample of measures that have been recently used to quantify state capacity in Chapter I, we have seen that there are many options to choose from. Researchers, however, do not always agree upon whether a given tool captures state capacity or not. Worldwide Governance Indicators' *Rule of law* and *Political stability and absence of violence* are good examples of indices that some researchers see as measures of state capacity, but some others do not.

For instance, according to Fortin (2010), Worldwide Governance Indicators' *Rule of law* and *Political stability and absence of violence* are related to other latent concepts than state capacity. Norris (2012), instead, argues that both *Rule of law* and *Political stability and absence of violence* are measures of state capacity. Likewise, Melville and Mironyuk (2016) use both indices to quantify state capacity. In partial agreement with the last two studies, Andrews, Pritchett, and Woolcock (2017) use only the former – and exclude the latter – as one of the constituents of their index of state capacity. Englehart (2009) suggests that all the six Worldwide Governance Indicators are measures of state capacity, whereas according to Savoia and Sen (2015) *Rule of law* should be seen as a proxy for state capacity but *Political stability and absence of violence* should not.

At least for now, no measure of state capacity has consolidated itself as the "best". As a consequence, given the wide range of available options, it is understandable that researchers have divergent ideas on what constitutes an ideal measure of state capacity. Using different measures to capture state capacity would be less of a problem if researchers scrupulously opted for a measure that closely matched their definition of state capacity and their theoretical approach. Anyhow, if researchers refrain from making such a move, the use of diverse measures for the same concept can become problematic. In particular, this is true if measures are not empirically equivalent among one another and do not lead to similar inferences in applied research.

Unfortunately, more often than not, data users are uncritical and irrational when choosing among competing measures (Mudde and Schedler 2010). Frequently, the choice of the measure of state capacity is not based on theoretical grounds nor the chosen measure represents closely the provided definition of state capacity. When this is the case, researchers might end up studying many different things under the notion of state capacity. Next, I will provide some illustrative examples on the mismatch between how state capacity has been defined and how it has been operationalised. I want to stress that my intention is not to discredit the authors of the following examples in any way, but simply to evidence cases of actual or potential mismatch between definition and measurement.

Some studies do not discuss at all why some measures of state capacity are selected instead of others nor provide any clear definition of state capacity. When this is the case, needless to say, it is impossible to assess if a chosen measure represents closely the chosen definition. For instance, DeRouen and Bercovitch (2008) measure state capacity with life expectancy but do not provide any conceptualisation of state capacity. Likewise, Hiilamo

and Glantz (2015) do not define state capacity at all nor provide any explanation on why they use Centre for Systemic Peace's *State fragility index* to quantify state capacity.

Melville and Mironyuk (2016) seem to consider state capacity, quality of institutions, and good governance as more or less equivalent concepts, but refrain from defining any of them. They quantify state capacity with five (out of six) Worldwide Governance Indicators and a broad composite index that synthesises the five indices. Hence, their composite index is based on sub-components such as *Financial freedom*, *Trade policy*, and the *Ease of setting up a subsidiary for a foreign firm*. Given that the authors do not provide any clear definition of state capacity, it is not possible to evaluate whether these sub-components are relevant in any way to their conception of state capacity. It must be said, though, that an index based on the above sub-components deviates drastically from most common understandings of state capacity reviewed in Chapter I.

Other studies do define state capacity but do not use measures that represent closely the chosen definition. A classic example is Fearon and Laitin's (2003: 80) use of GDP/capita as a proxy of "a state's overall financial, administrative, police, and military capabilities". GDP/capita might capture such capabilities, but for sure it captures many other things as well. A more recent example of a mismatch between concept and measurement is Bustikova and Corduneanu-Huci's (2017) use of infant mortality as a proxy of state capacity. The authors' definition of state capacity encompasses "meritocracy, territorial reach and distribution, and evidence-based policymaking ability with respect to public goods delivery" (Bustikova and Corduneanu-Huci 2017: 281). Yet, infant mortality does not represent closely such an understanding of state capacity, or for that matter, any other common understanding of the concept.

An example of an unreasonably broad measure is found in Dincecco (2015). He defines state capacity as "the extractive and productive power of states" (Dincecco 2015: 901) but operationalises the chosen definition with Rice and Patrick's (2008) *State weakness index*, which is based on sub-components such as *income inequality*, *inflation*, *democracy*, *undernourishment*, and *child mortality*. Besides the weak analytical usefulness of such a broad index of state capacity, the chosen measure seems to capture a much broader concept than the one described by the author. At least, it would be important to explain why this particular index should represent the extractive and productive power of states more adequately than the other available measures of state capacity.

Rossignoli (2016), instead, does the opposite by selecting a measure that captures at best a part of his conceptualisation of state capacity. He defines state capacity as "the effectiveness of state sovereignty over its territory and population" (Rossignoli 2016: 427) but measures state capacity with Varieties of Democracy's *Sovereignty over population*, which is an estimate of the share of population that the state has control over (Coppedge et al. 2017). The chosen measure, thus, does capture state sovereignty over its population, but does not tell us much about the sovereignty of the state across its territory.

Similarly, Andersen et al. (2014: 1310) define state capacity as "the extent to which the state possesses the coercive and administrative capacities to penetrate and regulate society and extract and appropriate resources" but operationalise the multidimensional concept with *tax revenues/GDP*. There is no doubt that the chosen measure falls short of representing closely the chosen definition, but at least the authors are aware of the problem, discuss it satisfactorily, suggest caution in the interpretation of their findings, and complement their analysis with two indicators capturing specific dimensions of state

capacity (Andersen et al. 2014). Taking these steps should be the norm when a chosen measure does not represent closely a chosen definition.

In fact, it is not uncommon to use multiple measures of state capacity to capture different dimensions of the concept and/or to ensure the robustness of the results. While the use of multiple measures can help indeed to generate more robust results, it does not guarantee that the chosen measures match closely the chosen conceptualisation. If the concept-measurement mismatch remains large, using multiple measures is only of limited help also in robustness testing.

To give an example of the problem, Englehart (2009: 167) defines state capacity as "the willingness and capability of the state apparatus to carry out government policy" and measures it with *Law and order* (ICRG), *Corruption perceptions index* (TI), and *Tax revenues/GDP* (WDI). Even if according to Englehart (2009: 176) these are "three of the most promising indicators" of state capacity, it is self-evident that alone none of them captures the author's conceptualisation. Although "each taps a different aspect of state capacity" (Englehart 2009: 167), it remains unclear whether using all the three measures covers the author's conceptualisation of state capacity in its entirety or not. In another similar example, Gibler and Miller (2014: 635) define state capacity as "the ability of the government to enforce its sovereignty across all its lands" but measure it with data on military personnel and government consumption. It is unrealistic to think that the provided definition would be completely captured by the two measures.

If a measure should fully cover a given definition because otherwise our understanding of a concept can be distorted (Bollen 1989), the aforementioned examples have shown that too often a chosen tool of measurement does not represent well a given definition of state capacity. In the worst cases, the operational definition of state capacity

remains completely unclear. As a consequence, it becomes difficult to understand what these studies are actually analysing in the name of state capacity.

This major limitation common to many studies on state capacity deteriorates the quality of the empirical literature on the topic and undermines the accumulation of knowledge. At the minimum, researchers should justify why a particular indicator has been selected instead of others and discuss the validity of the selected indicator. Before acquiring a deeper understanding of extant cross-national research on state capacity, thus, it is crucial to analyse and compare comprehensively the statistical instruments that have been used to measure state capacity. This is precisely what the next part of the study at hand aims to do.

As already discussed, if common measures of state capacity lead to similar interpretations, researchers do not need to worry too much about selecting one measure instead of another. If common measures of state capacity instead are in disagreement with one another and if these disagreements affect research results, choosing carefully the most appropriate measure or testing the robustness of the results with multiple measures becomes imperative.

Data and Methods

Selecting Relevant Measures of State Capacity

As we have seen, a variety of measures have been used to quantify state capacity in cross-national comparative social science literature. Since it is impossible to analyse comprehensively all these measures in a single study, I select some of the most established ones for further analysis according to four criteria. First, the selected measures must have academic relevance. The first criterion is met if a given measure has been used to quantify state capacity in several recent (>= 2010) cross-national comparative social

science studies by different authors. Second, I focus here only on subjective measures of state capacity. The second criterion is met if a given measure is at least partly based on perception-based data from expert surveys or assessments. Third, I select exclusively measures that rate countries over time on a yearly basis and cover most of the countries in the world. Fourth and last, only measures that are freely accessible are included.

Before proceeding with the analysis, I want to underline that even if all the chosen measures are commonly used to quantify state capacity, many of them were not created in the first place to capture state capacity. In fact, it can be useful to recall that most of the measures that researchers use to quantify state capacity were created for other purposes than for the measurement of state capacity. Given that the chosen measures, however, have been frequently used to quantify state capacity, the intended purpose of the measures is not of much interest for my study. The aim of this chapter is not to assess the intended purpose of the measures in relation to state capacity. On the contrary, the aim of this chapter is to analyse, compare, and evaluate measures that researchers commonly use to quantify state capacity, regardless of their intended purpose. Ultimately, my work will point out whether these measures are empirically similar or not and whether they lead to similar inferences or not. Selected measures, their sources, and their main characteristics are presented in Table 2.1.

The Quality of Government Institute of the University of Gothenburg publishes the well-known *Quality of Government Index* (QOG) (Teorell et al. 2019), which has been widely used in studies on state capacity and quality of government. The index is based on three separate sub-indicators – bureaucracy quality, corruption, and law and order – and captures "corruption within the political system", "the strength and impartiality of the legal system", "popular observance of the law", and "the institutional strength and quality of the bureaucracy" (Teorell et al. 2019: 385). QOG is computed as the simple average

Table 2.1. Main characteristics of selected measures of state capacity

Measure	Producer	Years	Countries	Scale	Source(s) of underlying data	Underlying variables	Type of data	Used in (e.g.)
Quality of Government Index (QOG)	Quality of Government Institute	From 1984	147	0 to 1 (low to high)	PRS Group's ICRG	3	Subjective	Charron and Lapuente (2010, 2011); Rothstein, Samanni, and Teorell (2012); Berliner et al. (2015); Povitkina and Bolkvadze (2019); Walther, Hellström, and Bergman (2019).
Capacity (HSI)	J. Hanson and R. Sigman	1960-2009	163	Z-score (low to high)	Multiple sources	24	Subjective and objective	Grassi and Memoli (2016a, 2016b); Van Ham and Seim (2018); Kim and Kroeger (2018); Bizzarro et al. (2018).
Government Effectiveness (WGI)	World Bank Institute	From 1996	193	Z-score (low to high)	Multiple sources	48	Subjective	Charron and Lapuente (2010, 2011); Halleröd et al. (2013); Böhmelt, Bove, and Gleditsch (2019).
State Fragility Index (SFI)	Center for Systemic Peace	1995-2018	167	0 to 25 (high to low)	Multiple sources	14	Subjective and objective	Besley and Persson (2011); Cingolani, Thomsson, and De Crombrugghe (2015); Hiilamo and Glantz (2015).
Failed States Index (FSI)	Fund for Peace	From 2005	178	0 to 120 (high to low)	In-house coders	100+	Subjective and objective	Møller and Skaaning (2011b); Lee and Zhang (2017); D'Arcy and Nistotskaya (2017).
Impartial Public Officials (VDEM)	Varieties of Democracy	From 1789	179	Z-score (low to high)	Survey experts	1	Subjective	Gjerlow et al. (2018); Bizzarro et al. (2018); Grundholm and Thorsen (2019).
Corruption Perceptions Index (CPI)	Transparency International	From 1995	180	0 to 10 until 2011; 0 to 100 since 2012 (low to high)	Multiple sources	14	Subjective	Joshi, Hughes, and Sisk (2015); Cingolani, Thomsson, and De Crombrugghe (2015); Lin (2015).

Note: Number of countries refers to year 2009 for HSI but to year 2018 for all other measures.

of these three sub-indicators, which are all taken from PRS Group's International Country Risk Guide (ICRG) and coded by PRS Group's country experts. The index is scaled to range from 0 (low) to 1 (high) and provides data for almost 150 countries in the world since 1984.

Hanson and Sigman's (2013) *Capacity* (HSI) has quickly become one of the most popular indices of state capacity in comparative social science literature. By drawing on the state capabilities identified by Skocpol (1985), the index focuses on the extractive, coercive, and administrative dimensions of the concept. These three dimensions are not quantified separately but instead captured by 24 sub-indicators, which are synthesised into a single latent variable through Bayesian factor analysis. HSI runs from low to high and follows a standardised (z-score) scale with a mean of 0 and a standard deviation of 1. It provides annual data for up to 163 countries in 50 years (1960-2009) and its 2013 version can be retrieved freely from several replication datasets. An updated version of the data (from 1960 to 2015) has been published in late 2020 (Hanson and Sigman 2020). Since the updated version comes with some changes in the underlying indicators and the country scores, the study at hand concerns only the 2013 version, which is the one that has been extensively used in the literature.

Government Effectiveness (WGI) is one of the six famous World Bank's Worldwide Governance Indicators. The index "captures perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies" (Kaufmann, Kraay, and Mastruzzi 2011: 4). WGI is a composite index based on multiple sub-indicators (48 in 2018). The index runs from low to high on a standardised (z-score) scale with a mean of 0 and a standard

deviation of 1. It covers virtually all countries in the world, and it is available biannually from 1996 to 2002 and annually from 2003 onwards.

The *State Fragility Index* (SFI) is produced and published by the Center for Systemic Peace. The index captures the state's "capacity to manage conflict, make and implement public policy, and deliver essential services, and its systemic resilience in maintaining system coherence, cohesion, and quality of life responding effectively to challenges and crises, and sustaining progressive development" (Marshall and Elzinga-Marshall 2017: 51). SFI is based on two dimensions, state effectiveness and state legitimacy, which are additively aggregated into the final index. These two dimensions, in turn, synthesise 14 sub-indicators related to four aspects (political, social, economic, security) of state performance. SFI ranges from 0 to 25, where a higher score indicates more fragility, and provides annual scores for all countries in the world with a population of at least 500,000 since 1995.

The *Failed States Index* (FSI), produced by the US-based non-profit research organisation Fund for Peace, is conceived to provide an entry point "into deeper interpretative analysis by civil society, government, businesses and practitioners alike – to understand more about a state's capacities and pressures which contribute to levels of fragility and resilience" (Fund for Peace 2019: 33). FSI scores are based on expert coding, content analysis of articles and reports, and quantitative secondary data concerning 12 dimensions such as security, rule of law, and public services. More than 100 sub-indicators capturing these dimensions are synthesised to get the final composite index but no precise information about these sub-indicators is provided. FSI has been published annually since 2005 and it ranked in its 2019 report 178 countries in the world. Its overall score ranges from 0 to 120, where a lower score indicates more capacity.

Transparency International's *Corruption Perceptions Index* (CPI) aggregates some of the most important existing measures of corruption and closely related issues, such as transparency, accountability, and professionalisation of the bureaucracy. "The CPI scores and ranks countries/territories based on how corrupt a public sector is perceived to be by experts and business executives" (Transparency International 2019: 1). Since "measures of corruption may provide another way of measuring state capacity" (Englehart 2009: 46), the index has been used to quantify state capacity in many cross-national studies. CPI has been published annually since 1995, it is based on secondary expert survey data from multiple sources, and the 2019 edition covers 180 countries in the world. It ranges from 0 (low) to 10 (high) until 2011 and from 0 (low) to 100 (high) from 2012 onwards.

Varieties of Democracy Institute's *Rigorous and Impartial Public Administration* (VDEM) provides information about "the extent to which public officials generally abide by the law and treat like cases alike, or conversely, the extent to which public administration is characterized by arbitrariness and biases" (Coppedge et al. 2019: 162). It has been used in several recent studies as a proxy of state capacity, because as we have previously seen, many researchers have argued that the functioning of a state's bureaucratic apparatus is the most critical aspect of state capacity. VDEM is based on evaluations by multiple country experts – mainly academics – and provides annual data from 1789 onwards for nearly all countries in the world. It runs from low to high and follows approximately the scale of a normal z-score (mean of 0; standard deviation of 1).

To sum up, the seven measures of state capacity that are selected for further analysis are Teorell et al.'s (2019) *Quality of Government Index* (QOG), Hanson and Sigman's (2013) *Capacity* (HSI), World Bank's (2019) *Government Effectiveness* (WGI), Centre for Systemic Peace's (2019) *State Fragility Index* (SFI), Fund for Peace's (2019) *Failed States Index* (FSI), Transparency International's (2019) *Corruption Perceptions Index*

(CPI), and Varieties of Democracy Institute's (Pemstein et al. 2019) *Rigorous and Impartial Public Administration* (VDEM).

Research Strategy

Now that I have selected some of the most established measures of state capacity, it is time to proceed to their comparison and evaluation. First, I examine qualitatively the components and the content of the chosen measures. The content of the measures is explored vis à vis the functional dimensions of state capacity identified in Chapter I. Besides a simple inspection of our seven measures of state capacity my objective is to answer the question *how useful is a given instrument to a given research agenda?* This means that throughout the exercise content validity is understood in terms of what Fitzpatrick (1983: 8) calls "content relevance". In particular, according to Fitzpatrick (1983: 7), "a content valid test must cover important aspects of the content universe that a user wishes to assess" (Fitzpatrick 1983: 7). Even if it is impossible to know all the universes that researchers might want to access with measures of state capacity, we can assess the relevance of the content of the measures in relation to different dimensions of state capacity.

After qualitatively analysing the content of the selected measures of state capacity, I shift my attention to a quantitative assessment of the measures. Unless otherwise stated, FSI and SFI are reversed so that a higher score indicates a higher level of state capacity. First, I explore some of the basic statistical properties of the measures through violin plots. For instance, more or less normally distributed variables are preferable to completely non-normal variables not because we expect real world state capacity to be distributed Gaussian but because many common statistical tests and analyses assume that variables follow approximately a "bell-shaped" curve.

Acquiring information on basic statistical features and distributional characteristics of measures of state capacity is important in itself. Nevertheless, violin plots allow us to discover also some less apparent measurement issues. To be more specific, following the evaluation criteria used in Vaccaro's (2021) evaluation of democracy measures (see also Knutsen 2010), I take the assumption that the inability of measures of state capacity to differentiate between countries can be considered a weakness; especially so if it occurs at extreme values. We would prefer to have measures that are informative about the differences among countries and do not clump at the extremes of a scale (Goertz 2020).

Second, after analysing basic statistical properties of the chosen measures of state capacity, I carry on with the exercise by assessing the strength of the associations among the measures and the convergent validity of the measures through correlations. Correlation analysis is a conventional tool for assessing the convergent validity of instruments measuring the same construct (e.g., Bollen 1989; Seawright and Collier 2014). To ensure the robustness of the results, correlations are computed with both Pearson's and Spearman's methods.

Since correlations provide us only information about bivariate association, I use also principal component analysis (PCA) as an exploratory tool to analyse the multivariate association of the measures and the structure of the data. "The main objective of a PCA is to reduce the dimensionality of a set of data" (Jolliffe 2002: 87). Thus, if the bulk of the total variance in the data is best explained by one single component, we are induced to conclude that common measures of state capacity are closely related among one another and represent statistically one single unidimensional concept.

Third, after evaluating the convergence of the chosen measures of state capacity, I proceed in my analysis by assessing the interchangeability of the measures. To be considered as *interchangeable*, "equivalent measures should produce similar causal

inferences" (Seawright and Collier 2014: 124). To give a practical example, in assessing the interchangeability of seven measures of rule of law, Møller and Skaaning (2011a: 384) "test whether the results are relatively similar or dissimilar when the seven measures are used interchangeably as dependent variables in multiple OLS regression analyses". Following their approach, I evaluate the interchangeability of common measures of state capacity by regressing the measures on the same set of external predictors in a multivariate setting.

Gaining knowledge not only on the convergence of the measures but also on their interchangeability is essential because previous studies show that even highly correlated measures can lead to completely different statistical inferences (e.g., Casper and Tufis 2003; Högström 2013; Vaccaro 2021). Some scholars argue that since measures of well-functioning states are highly correlated, "how the quality of government is measured is not so important, because many different policy indicators portray a very similar picture" (Tabellini 2008: 264). Others suggest instead that in fact if the scholar's aim is feasible statistical inference, assessing interchangeability might be even more important than assessing convergence (Seawright and Collier 2014). If my results indicate substantial divergence between convergence and interchangeability, researchers who use these measures for inferential purposes should not rely only on correlations in evaluating the empirical equivalence of the measures.

I assess the interchangeability of the selected measures of state capacity by replicating a set of influential studies on the relationship between state capacity and democracy. As previously said, the primary aim of these regressions is to assess whether different measures of state capacity lead to similar empirical findings. Anyhow, as a consequence, we are also able to evaluate the external validity of the replicated studies. If the choice of the measure of state capacity matters substantially for the conclusions to

be drawn, we can conclude not only that the interchangeability of the measures is weak but also that the replicated studies have weak external validity, and thus, their findings on the relationship between state capacity and democracy cannot be generalised across measures of state capacity.

Finally, in the last empirical part of this chapter, to understand more in detail the similarities and differences of measures of state capacity, I shift my focus on individual country scores. As we shall see, different measures lead to different interpretations (i.e., are not interchangeable), and thus, measures must disagree substantially at least on the ratings of some of the countries. First, country scores are analysed in bivariate settings. Then, by creating an indicator of multivariate country-specific rating discrepancy, I determine which countries have particularly similar or dissimilar scores across all the measures and shed light on the possible causes of rating discrepancy. Country scores of some of the most discrepantly rated countries are further analysed against the previously explored content captured by individual measures. I focus on distances between cases in terms of scores. Yet, I acknowledge that an alternative approach could be to analyse disagreements in rank.

Results

Content Validity

Table 2.2 presents an overview the content of the selected measures of state capacity. We already know that the measures under scrutiny are based on slightly different underpinnings. Hence, even if all the measures are frequently used to quantify state capacity, we should not be too surprised to find that there are some differences from one measure to another, in terms of content. Here, I will explore the content of the measures in relation to the dimensions of state capacity identified by Savoia and Sen (2015) and

Table 2.2. Content of selected measures of state capacity

Measure	Dimensions	Components/indicators/attributes	Further explanations
QOG	Administrative	Institutional strength and quality of the bureaucracy	Autonomy from political pressures, established mechanisms for recruitment and training, and bureaucratic strength and expertise to govern without drastic changes in policy or interruptions in government services
		Corruption within the political system	Mainly concerned with excessive patronage, nepotism, job reservations, favouritism, secret party funding, and suspiciously close ties between politics and business
	Legal	Strength and impartiality of the legal system	
	Legal/coercive	Popular observance of the law	
HSI	Extractive	Efficiency of revenue mobilization	
		Relative political capacity	The difference between actual and predicted levels of tax extraction
		Tax evasion not damaging	The amount of damage to public finances caused by tax evasion
		Taxes on income (as % of revenue)	
		Taxes on international trade (as % of revenue)	
		Total tax revenue (as % of GDP)	
	Coercive	Military personnel (per 1,000 in population)	
		Military spending (per million in population)	
		Monopoly on the use of force	The extent to which the state's monopoly on the use of force covers the entire territory of the country
		The level of state sanctioned or state perpetrated violence within the state's own territorial borders	
	Administrative	Administration and civil service	Political autonomy and conflicts of interest safeguards in the civil service, whistleblower protections, and transparency, fairness, and conflicts of interest safeguards in public procurement and privatization
		Administrative efficiency	
		Institutional strength and quality of the bureaucracy	
		Confidence in the civil service	
		Effective implementation of government decisions	
		Quality of budgetary and financial management	The extent to which there is a comprehensive and credible budget linked to policy priorities, effective financial management systems, and timely and accurate accounting and fiscal reporting
		Quality of public administration	The extent to which there is a comprehensive and credible budget linked to policy priorities, effective financial management systems, and timely and accurate accounting and fiscal reporting
		Weberianness scale	Degree of meritocratic recruitment and predictable, long-term career rewards in bureaucracy

	Infrastructural	Statistical capacity	Level of statistical capacity based on an assessment of methodology, source data, periodicity, and timeliness
		Census frequency	
	Legal	Contract intensive money	Ratio of non-currency money to the total money supply
	Other	Stability of political institutions	
		Artificiality of state borders	
		Degree of mountainous terrain	
WGI	Administrative	Quality of bureaucracy and its ability to carry out government policy	
		Institutional strength and quality of the bureaucracy	
		Quality of public administration	The extent to which civilian central government staff is structured to design and implement government policy and deliver services effectively
		Quality of budgetary and financial management	The extent to which there is a comprehensive and credible budget linked to policy priorities, effective financial management systems, and timely and accurate accounting and fiscal reporting
		Efficiency of revenue mobilization	
		Public management	
		Civil service integrity	
		Health and education	
		Welfare	
		Business environment and infrastructure	
		How problematic is electricity of the growth of your business?	
		How problematic is transportation for the growth of your business?	
		Steering capability	The ability of the government to manage reforms effectively and to achieve its policy priorities
		Resource efficiency	The ability of the government to use available resources optimally
		The allocation and management of public resources for rural development	
		Bureaucracy does not hinder business activity	
		Efficiency of distribution infrastructure of goods and services	
		Excessive bureaucracy and red tape	
		Quality of primary education	
		Satisfaction with public transportation system	
		Satisfaction with roads and highways	
		Satisfaction with education system	

		Quality of overall infrastructure	
	Infrastructural	Territorial coverage of public schools	
		Territorial coverage of basic health services	
		Territorial coverage of drinking water and sanitation networks	
		Territorial coverage of electricity grid	
		Territorial coverage of transport infrastructure	
		Territorial coverage of maintenance and solid waste disposal	
	Administrative/ infrastructural	Infrastructure disruption	Likelihood of disruption to and/or inadequacy of infrastructure for transport
	Legal/coercive	State failure	The ability of the state to ensure law and order, the supply of basic goods, and to respond or manage emergencies
	Other	Policy instability	The risk that the government's broad policy framework shifts over the next year
		Amount of trust in government	
		Consensus-building	The ability of the political leadership to establish a broad consensus on reform with other actors in society without sacrificing its reform goals
		The degree of the adaptability of government policy to changes in the economy	
SFI	Coercive	Total residual war	Total summed magnitudes of all societal and interstate major episodes of political violence
		The level of state sanctioned or state perpetrated violence within the state's own territorial borders	
		Polity fragmentation	The amount of territory that is effectively under control of the local authority and actively separated from the central authority of the regime
	Other	Regime durability	The number of years since the most recent regime change
		Total number of coup events	
		Current leader's years in office	
		GDP/capita	
		Human Development Index	
		Factionalism	The extent to which alternative preferences for policy and leadership can be pursued in the political arena
		The degree of ethnic group political discrimination	
		Political salience of elite ethnicity	
		Exclusionary ideology of running elite	
		Share of export trade in manufactured goods	
		Infant mortality	
FSI	Coercive	Security apparatus	Monopoly on the use of force

		Relationship between security and citizenry
		Proper use of force
		Availability of weapons and plan to disarm former combatants
	External intervention	Existence of external support for factions opposed to the government
		Amount of external military assistance and foreign troops or attacks
		Dependence on economic intervention or aid
	Group grievance	Post-conflict response
		Equitable and efficient distribution of resources
		Divisions between ethnic and/or religious groups
		Vigilante justice and mass violence
Administrative/ infrastructural	Public services	General condition of public services and equality of access
		Level of health services
		Level of education
		Burden of housing costs
		Quality of infrastructure
Adminstrative/ extractive/coercive/ other	State legitimacy	Confidence of the people in the government
		Violence of political opposition
		Corruption of public officials
		Openness and fairness of the political process
		Political violence and politically motivated attacks
Legal/other	Human rights and rule of law	Protection of civil rights, political rights, and freedoms
		Abuse of rights
		Freedom of media and equal access to information
		Legal rights and justice
		The level of political power sharing encouraged by the system
	Uneven development	Economic and legal equality
		Opportunities to improve economic status
		Existence of ghettos and slums
Other	Human flight and brain drain	Retention of technical and intellectual capital
		Amount of remittances
		Growth of diasporas and its importance for home state economy
	Economic decline	The level of government debt
		The level of economic conditions
		Favourable business climate and consumer confidence
		The level of economic diversification
	Demographic pressures	Pressures related to population growth, density, and infant mortality
		Ability to control the spreading of diseases
		Adequacy of food supply
		Existence of sound environmental policies

			Existence of land/resource competition and access to potable water
		Refugees and internally displaced persons	Likeliness of refugee influx and provision of adequate conditions to refugees
			Amount and predicted growth of internally displaced persons
			Plans and resources for refugees and internally displaced persons
		Factionalized elites	Representativeness of leadership
			National identity and nationalism
			Fair distribution and control of resources
			Reasonable laws and representative leadership
VDEM	Administrative	Rigorousness and impartiality of public officials in the performance of their duties	The extent to which public administration is characterised by arbitrariness and biases
CPI	Administrative	Perceptions of corruption in the public sector	The level of accountability of the executive to oversight institutions and of public employees for their performance, the access of civil society to information on public affairs, and the capture of the state by narrow vested interests
			To what extent are public officeholders prevented from abusing their position for private interests?
			To what extent are public officeholders who abuse their positions prosecuted or penalized?
			To what extent does the government successfully contain corruption?
			Are there clear procedures and accountability governing the allocation and use of public funds?
			Are public funds misappropriated by ministers/public officials for private or party-political purposes?
			Are there special funds for which there is no accountability?
			Are there general abuses of public resources?
			Is there a professional civil service or are large numbers of officials directly appointed by the government?
			Is there an independent body auditing the management of public finances?
			Is there a tradition of a payment of bribes to secure contracts and gain favours?
			Has the government implemented effective anti-corruption initiatives?
			Is the country's economy free of excessive state involvement?
			Is the government free from excessive bureaucratic regulations, registration requirements, and other controls that increase opportunities for corruption?
			Are there significant limitations on the participation of government officials in economic life?
			Does the government advertise jobs and contracts?
			Does the state enforce an effective legislative or administrative process—particularly one that is free of prejudice against one's political opponents—to prevent, investigate, and prosecute the corruption of government officials and civil servants?
			Are allegations of corruption given wide and extensive airing in the media?

		Does the public display a high intolerance for official corruption?
		The risk that individuals/companies will face bribery or other corrupt practices to carry out business
		Does bribery and corruption exist or not?
		How do you grade the problem of corruption in the country in which you are working?
		Corruption within the political system
		How common is diversion of public funds to companies, individuals or groups due to corruption?
		The extent to which government officials use public office for private gain
		How pervasive is political corruption?
	Legal	Is there an independent judiciary with the power to try ministers/public officials for abuses?
		Are there adequate laws requiring financial disclosure and disallowing conflict of interest?
		Do whistleblowers, anti-corruption activists, investigators, and journalists enjoy legal protections that make them feel secure about reporting cases of bribery and corruption?
		How common is it for firms to make undocumented extra payments of bribes connected with imports and exports, public utilities, annual tax payments, awarding of public contracts and licences, and obtaining favourable judicial decisions?

discussed in the first chapter of this dissertation. Moreover, I will provide some examples of potential problems in the analytical usefulness of each measure.

In terms of its "content universe", QOG captures mainly the administrative, legal, and coercive dimensions of state capacity. Its focus seems to be on the procedural aspects of the state such as corruption and impartiality. Countries with a high score on the index should have thus an autonomous and meritocratic bureaucracy, well-trained civil servants, law-abiding citizens, a strong and impartial judicial system, and a lack of clientelism, nepotism, favouritism, and too close ties between politicians and businesses. Qualitatively, the administrative dimension of state capacity accounts two thirds of QOG's overall score. Institutional strength, quality of the bureaucracy, and corruption are all usually associated with administrative capacity; the strength and impartiality of the legal system are particularly relevant for legal capacity; the popular observance of the law, based on factors such as crime rate, can be associated at least with both the legal and coercive dimensions of state capacity. QOG does not seem to capture aspects of state capacity strictly related to the extractive or infrastructural dimensions. Thus, researchers should not use QOG if they aim to quantify in particular the extractive and infrastructural aspects of state capacity.

HSI taps into all the five identified dimensions of state capacity: extractive, coercive, administrative, infrastructural, and legal.[3] In terms of content, thus, HSI seems to be the measure that best reflects the previously identified dimensions of state capacity. Yet, the relationship between some of its sub-components such as *mountainous terrain* and *fractal borders* and the final index itself remains dubious. Such doubts are even increased by the nearly absent statistical association between these two sub-components and the final

---

[3] Although the authors of the index classify their components only into three dimensions of state capacity: extractive, coercive, and administrative (Hanson and Sigman 2013).

index (Hanson and Sigman 2013). Moreover, it is also unclear whether the amount of mountainous terrain and the artificiality of state borders are attributes of state capacity that most researchers wish to assess in their research agendas. The inclusion of *anocracy* as one of the sub-components of state capacity risks also to decrease the analytical utility of HSI. Anocracy captures the consistency of political institutions and it is based on the well-known *Polity* measure of democracy. Since HSI is indirectly based on the Polity scores, scholars should be at least careful in using the index to analyse the relationship between democracy and state capacity, particularly so if democracy is measured with Polity.

WGI synthesises numerous sub-indicators related mainly to the quality of public administration and the quality of public services. Therefore, it focuses strongly on the administrative dimension of state capacity. Compared to the other measures, the content of some of its sub-components seems to be excessively business oriented. For instance, sub-indicators such as *excessive bureaucracy and red tape* and *bureaucracy does not hinder business activity* are attributes that are not really constituents of most common definitions of state capacity. Most scholars working on state capacity are not likely to be interested in assessing such attributes, unless their intention is precisely to capture a definition of state capacity that includes these pro-business attributes. An additional problem of the WGI lies in its inclusion of sub-indicators that capture the quality and reach of public services. The inclusion of attributes such as *health and education* and *welfare* restricts the analytical usefulness of WGI. Researchers should not use WGI to analyse the relationship between state capacity and the provision of public services, given that the latter is already included in WGI. The index incorporates also a few single sub-indicators related to other dimensions of state capacity, such as the territorial coverage of electricity (infrastructural capacity) and enforcement of law and order (legal/coercive

capacity), but in general, solely in terms of its content, WGI seems to be most useful for researchers who wish to capture principally the administrative dimension of state capacity.

SFI focuses especially on coercive aspects of state capacity such as political violence, security, and war. In addition to coercive capacity, SFI captures many general attributes of well-functioning states in a broad perspective. Most of these attributes do not fit well into any of the previously identified dimensions of state capacity and undermine the analytical usefulness of the index. For instance, sub-components such as *infant mortality, human development, and GDP/capita* do not seem to be relevant attributes of state capacity in most research agendas and capture expected outputs that are unlikely to be driven exclusively by state capacity. In fact, these sub-components can be even analytically counterproductive if researchers wish to analyse the relationship between state capacity and socio-economic development. If researchers instead wish to quantify a "content universe" that includes infant mortality, human development, and economic development, then SFI seems to be an adequate choice. In particular, the use of GDP/capita as a measure of state capacity has been criticised (see e.g., D'Arcy and Nistotskaya 2021). We can expect the level of economic development to be related to state capacity but if we measure state capacity with GDP/capita it is not possible to analyse the relationship between the two. Users of SFI should be informed that the index gives a particularly high importance to the level of economic development, given that it is included twice in the index: first "explicitly" through GDP/capita, and then "implicitly" as a component of the Human Development Index.

With more than 100 sub-indicators, FSI is the broadest measure of state capacity assessed in my study. The index captures many different attributes that erode its analytical usefulness in some of the most common research agendas on state capacity. To give a

few examples, the inclusion of sub-components related to economic conditions, transparency, education, health, and some aspects of the political process make the index ill-suited to examine the linkage between all these social, political, and economic phenomena and state capacity. It seems to quantify all the identified dimensions of state capacity, at least up to a certain extent. Yet, since the underlying attributes and the content of the underlying indicators are not disclosed in their entirety, we cannot be completely sure about the "content universe" captured by FSI. Naturally, this means also that FSI is less transparent compared to the other measures. We know that its sub-indicators are based on quantitative secondary data, content analysis of articles and reports, and qualitative research by social scientists but we do not know what is the secondary data used in FSI, what are the articles and reports used in FSI, or why precisely experts have coded a given country with one score instead of another.

CPI synthesises several indicators related to corruption in the public sector. It seems to focus mainly on procedural aspects related to administrative capacity. In terms of its content, CPI does not seem to capture the universe related to the coercive, extractive, and infrastructural dimensions of state capacity. Hence, researchers should probably avoid using the index if they wish to quantify any of these three dimensions of state capacity, unless they have defensible reasons to not to do so. While many of the underlying attributes reflect administrative aspects of state capacity, some of the attributes (e.g., *to what extent are public officeholders who abuse their positions prosecuted or penalised?*) capture the legal dimension of state capacity as well. Some other underlying "questions" do not really fit any of the common dimensions of state capacity. For example, questions such as *is the country's economy free of excessive state involvement?* do not seem to be relevant for the analysis of any of the previously identified dimensions of state capacity. If scholars consider the ability to extract revenues to be a core characteristic of state

capacity, pro-business perceptions of excessive state involvement in the economy should not be used to quantify state capacity, because they might value positively states that are *not* able to extract revenues. Some underlying questions such as *are allegations of corruption given wide and extensive airing in the media?* conflate freedom of the media with state capacity. This means that CPI should not be used to analyse the relationship between media freedom and state capacity.

In terms of content, VDEM has the narrowest scope among the measures of state capacity analysed in the chapter at hand. The measure captures expert perceptions regarding the *impartiality of public officials in performing their duties*. Users of VDEM should keep in mind that its content validity is high only if they wish to quantify such a precise aspect of state capacity. This is not to say by any means that VDEM should not be used to measure state capacity. In truth, some procedural definitions of state capacity consider impartiality as the most important element of the concept. Thus, even if the content of VDEM cannot reflect the entirety of the multidimensional concept of state capacity, its content is more than relevant to quantify certain core aspects of state capacity. On the one hand, using such a minimalist measure of state capacity mitigates the risk of including of theoretically irrelevant attributes the content. On the other hand, we can take for granted that many theoretically relevant attributes are not included in VDEM. Compared to other measures of state capacity, its users do not risk conflating too many attributes into one single measure. VDEM's content validity is high if it is used to quantify primarily the administrative dimension of state capacity. The other common dimensions are at best covered by the instrument indirectly through the administrative dimension. Theoretically, the content of VDEM reflects closely the content universe that scholars wish to analyse when the latter equals to the impartiality of public officials.

In this section we have explored and compared the content of some of the most frequently used measures of state capacity. We have found that each measure is based on slightly different sub-components and content. Of the surveyed measures, HSI represents most closely the dimensions of state capacity individuated by Savoia and Sen (2015). Nevertheless, evaluating the validity of the content of the measures in absolute terms is tricky, also because "there are no agreed-upon criteria for establishing whether, in fact, a measure has attained content validity" (Zeller and Carmines 1980: 79). The relevance of the content of a given measure must depend then on the theoretical and analytical objectives of its users. I have provided some examples on research agendas where each of the surveyed measures is likely or unlikely to be useful.

To sum up the main findings of this section, VDEM's content validity is the highest when the research question concerns a precise aspect of state capacity: the impartial behaviour of public officials in the exercise of their duties. QOG's content validity is highest when the research question concerns mainly administrative but also legal and coercive aspects of state capacity. While both VDEM and QOG exclude some of the identified dimensions of state capacity, their users can be confident that neither of the measures includes attributes that are irrelevant to common definitions of state capacity. On the contrary, the remaining five measures – in particular FSI, SFI, WGI, and CPI – seem to include content that weakens the analytical usefulness of the measures in several research agendas and represents attributes that do not fit well into the identified dimensions of state capacity.

When choosing one measure over another, researchers should take into account the different aspects of state capacity covered by the measures and select a measure that closely represents the selected theoretical perspective. Things become trickier, however, if no measure represents well the proposed approach or if researchers are not willing to

select the measure that best represents their theory. In such cases, the statistical equivalence of the measures becomes of critical importance. Next, we will turn our attention to the statistical similarities and differences among the measures of state capacity.

Basic Statistical Properties

Violin plots (Figure 2.1) illustrate the basic statistical properties of the selected measures in all common country-years. The grey outlines of the "violins" show the frequency distribution of each measure. The black-bordered box in the middle of each violin stretches out from the first to the third quartile of each variable. The whiskers stretch out to the lowest and highest observations that are not considered unusual in the data. Single observations that do not fall inside this range of the data (i.e., outliers) are represented by dots above or below the whiskers. The small black rectangle inside the box represents the median of each variable.

Figure 2.1. Violin plots of measures of state capacity (2005-2009)
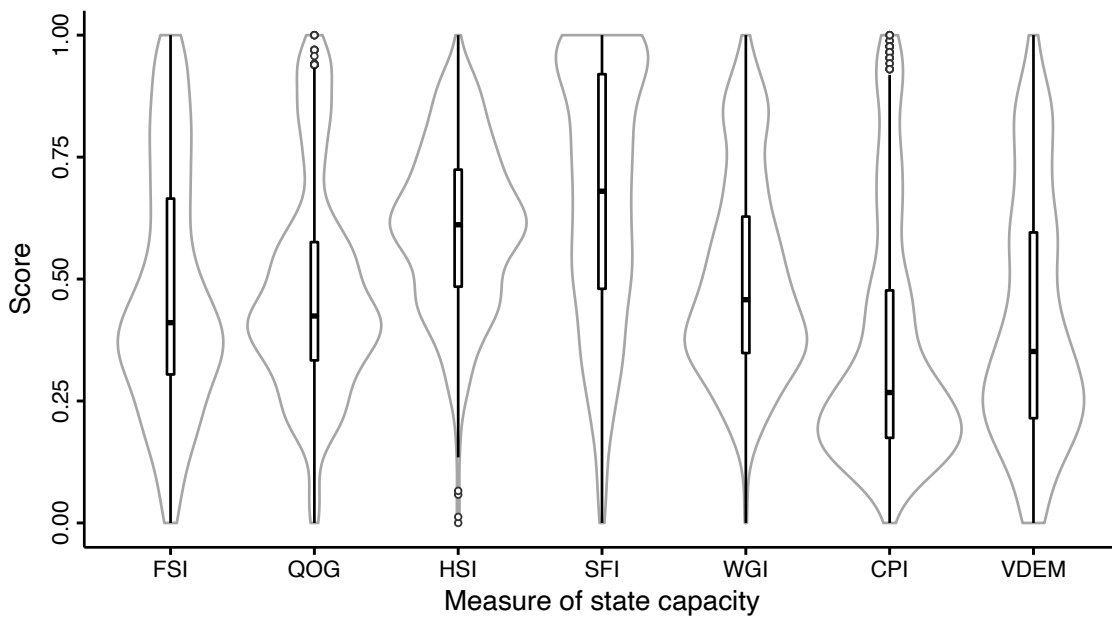
Note: Missing data handled with listwise deletion. Scores are normalised to range from 0 to 1.

First and foremost, a visual inspection of the violin plots shows immediately that some of the measures have a fairly normal distribution, but others are far from being normally distributed. In particular, CPI and SFI seem to be the "least Gaussian" measures. CPI is heavily skewed to the right and has a particularly low mode and median. Nearly half of its observations are in the bottom quarter of the scale and more than 75% of its observations are below the mid-point of the scale. In practice, this means that generally countries receive less generous scores with CPI than with the other measures. SFI, instead, suffers from the opposing problem: it has a heavily left-skewed distribution and a comparatively high mode and median. More than 35% of its country-years are in the top 20% of the scale, but only around 5% of its country-years are in the bottom 20% of the scale. Hence, in general, countries are more "capable" with SFI than with the other measures. The remaining five measures of state capacity (FSI, QOG, HSI, WGI, and VDEM) are more normally distributed than CPI and SFI. FSI's, QOG's, WGI's, and VDEM's main peaks and medians are closer to the lower end of the scale, whereas HSI's main peak and median is closer to the upper end of the scale. QOG and CPI have outliers at high levels of state capacity, whereas HSI has outliers at low levels of state capacity.

In addition to these general distributional features, violin plots are useful in revealing also less obvious empirical shortcomings of some of the measures. SFI seems to compress too many observations at the upper extreme of the state capacity scale. From 2005 to 2009, SFI rates as many as 83 country-years with the maximum level of state capacity. Since the other measures of state capacity are able to distinguish between these countries nearly without any exceptions, we are induced to conclude that SFI has severe limitations in its ability to distinguish high state capacity countries among one another. Moreover, given that SFI assigns the maximum possible score to such a large number of countries, SFI cannot capture any potential increase in the level of state capacity in many countries.

The clumping at the upper extreme of the scale and the inability to distinguish among high capacity countries are serious weaknesses of SFI.

The other six measures of state capacity do not have such an agglomeration of equivalently rated observations at either of the two extremes of the scale. Nevertheless, some of the other measures do have some problems in differentiating countries one from another. CPI is not finely grained enough to distinguish between many countries at low levels of state capacity. For instance, in 2009 – the most recent year of common observations – Sierra Leone, Ecuador, Ukraine, Kenya, Zimbabwe, Cameroon, and Russia have exactly the same score with CPI. QOG, instead, is not finely grained enough to distinguish between many countries at intermediate levels of state capacity. In 2009, Pakistan, Cameroon, Romania, Lebanon, South Africa, Brazil, Panama, Guyana, Nicaragua, Jamaica, Madagascar, and Colombia have exactly the same score with QOG. In both cases, despite the empirical equivalence, it is unlikely that there were no actual differences in the level of state capacity of all these countries. Otherwise the remaining measures would not have been able to capture the differences between these countries virtually without exceptions.

FSI, HSI, WGI, and VDEM are able to distinguish well between different observations. In fact, WGI is able to distinguish between all the common observations in 2009. With FSI, HSI, and VDEM instead no more than two countries have an identical score in 2009. Therefore, if we hold on to the assumption that the ability to differentiate between countries is an asset of any measure of state capacity, from this specific perspective, FSI, HSI, WGI, and VDEM are more informative measures than SFI, CPI, and QOG. Overall, SFI seems to have less desirable statistical properties than the other surveyed measures.

Convergent Validity

Now we have acquired interesting information about the qualitative content and basic statistical properties of the measures. There are both similarities and differences among frequently used measures of state capacity, but we do not know yet if these measures are statistically associated among one another. Correlation coefficients in years of common coverage (Table 2.3) show that the measures are highly correlated among one another. Correlations are computed with both Pearson's and Spearman's methods, to ensure the robustness of the results to both estimators. Pearson's method measures the strength of the linear association between two variables, whereas Spearman's method measures the strength of the monotonic association between two variables. The interpretation of the results is not significantly affected by the chosen method. In general, the findings show that measures of state capacity are highly associated among one another.

Table 2.3. Pairwise correlation coefficients of measures of state capacity (2005-2009)

	FSI	QOG	HSI	SFI	WGI	CPI	VDEM
FSI	1.00	0.80	0.83	0.89	0.86	0.85	0.75
		(672)	(785)	(797)	(849)	(804)	(821)
QOG	0.87	1.00	0.85	0.79	0.90	0.85	0.74
	(672)		(665)	(675)	(695)	(675)	(685)
HSI	0.85	0.86	1.00	0.84	0.90	0.85	0.70
	(785)	(665)		(809)	(809)	(782)	(809)
SFI	0.87	0.76	0.83	1.00	0.83	0.82	0.70
	(797)	(675)	(809)		(826)	(795)	(826)
WGI	0.90	0.93	0.90	0.81	1.00	0.92	0.79
	(849)	(695)	(809)	(826)		(850)	(869)
CPI	0.89	0.91	0.84	0.75	0.94	1.00	0.79
	(804)	(675)	(782)	(795)	(850)		(824)
VDEM	0.81	0.80	0.72	0.70	0.83	0.83	1.00
	(821)	(685)	(809)	(826)	(869)	(824)	

Note: Pearson's correlation coefficients (bottom-left quadrant) and Spearman's correlation coefficients (upper-right quadrant) in common years of coverage (2005-2009). Number of observations in parentheses; all coefficients significant at the $p < 0.001$ level.

With Pearson's method, the weakest correlations are between SFI and VDEM (0.70) and HSI and VDEM (0.72), whereas the strongest correlations are between CPI and WGI (0.94) and QOG and WGI (0.93). Likewise, with Spearman's method, the weakest correlations are between SFI and VDEM (0.70) and HSI and VDEM (0.70). The strongest correlations, instead, are with CPI and WGI (0.92), QOG and WGI (0.90), and HSI and WGI (0.90). VDEM is somewhat less strongly associated to the other measures, because its correlation coefficient never exceeds 0.83, regardless of the method. Nonetheless, overall, these findings indicate a high convergence among frequently used measures of state capacity.

So far, we have examined the convergent validity of measures of state capacity in years of common coverage and we have found that measures of state capacity are strongly related to each other from 2005 to 2009. Yet, strong correlations hold also over a longer time period[4] (Tables A1-A7, Appendix A). In fact, the over-time consistence of the bivariate relationships among measures of state capacity is astonishing. Only with two pairs of measures the strength of the correlation varies more than 0.10 over the examined period of time: the correlation between QOG and SFI ranges from 0.73 (in year 2014) to 0.84 (in years 1997 and 1998), and the correlation between QOG and CPI ranges from 0.82 (in year 1995) to 0.93 (in years 1997, 2015, 2016, and 2017). Interestingly, the bivariate correlations between CPI and the other measures take a relatively pronounced leap from 1995 to 1996, suggesting that there might be something anomalous in the CPI scores of 1995. For instance, the comparatively small number of countries (40) rated by CPI in 1995 could play an important role in affecting its relationship with other measures.

---

[4] 1995-2017; data before 1995 is not analysed because most of the measures do not cover earlier years.

Bivariate correlations provide information about the relationship between two given variables. Anyhow, we can analyse the relationship among our measures of state capacity with multivariate methods as well. Principal component analysis (PCA) is commonly used as a variable-reduction technique but it can also help to understand better the association among multiple variables and the structure of a set of data.

The results of the PCA (Table 2.4) show that around 87% of the total variance can be attributed to one single component. The second component explains less than 5% of the total variance. According to the Kaiser criterion, components with eigenvalues under 1.0 should not be retained. Thus, the PCA indicates that measures of state capacity are best represented by one single component and suggests that statistically all the measures capture the same concept. Only if we would have found the second component to explain a substantial amount of total variance, we could have questioned whether indicators of state capacity are measuring the same concept at all. Robustness tests with extended year coverage (Tables A8-A9, Appendix A) do not change our conclusions and the bottom line remains the same: the instruments are highly convergent and seem to measure the one and same concept of state capacity.

Table 2.4. Principal component analysis of measures of state capacity (2005-2009)

Component	Eigenvalue	% of explained variance	Cumulative % of explained variance
1	6.083	86.90	86.90
2	0.332	4.74	91.64
3	0.268	3.82	95.46
4	0.113	1.61	97.08
5	0.095	1.36	98.44
6	0.067	0.95	99.39
7	0.043	0.61	100.00

Interchangeability and External Validity of Previous Studies

So far, we have explored the content of some of the most established measures of state capacity, we have evaluated their basic statistical properties, and we have analysed their convergence. We have found that the measures have high convergent validity, are strongly related to each other, and seem to quantify statistically one single concept. As a handful of previous studies on measurement validity have evidenced, however, high correlations do not necessarily translate into high interchangeability. Hence, to assess more thoroughly the empirical consequences of choosing one measure instead of another in hypothesis-testing, I replicate four regression models published in four studies on the effect of democracy on state capacity.

The choice of replicating studies on this specific topic is not casual. First of all, it is determined by the fact the state-democracy literature constitutes one of the largest strains of research in which state capacity is examined as an outcome. Second, as already mentioned, important parts of the dissertation at hand are framed around the relationship between state capacity and democracy. As a by-product of our replication analysis, besides acquiring information about the interchangeability of the measures, we are also able to gain knowledge on the external validity of the results of the original models.

Next, I replicate three longitudinal models and one cross-section model from influential studies on the state-democracy nexus. The three longitudinal models are taken from Bäck and Hadenius' (2008), Carbone and Memoli's (2015), and Grassi and Memoli's (2016a) studies. The single cross-section model is taken from Charron and Lapuente (2010). Actual replication data is available for Carbone and Memoli's (2015) and Grassi and Memoli's (2016a) studies. Bäck and Hadenius' (2008) and Charron and Lapuente's (2010) studies are replicated to the best of my ability by following

scrupulously the procedure described by the authors. Summary statistics of the data used in the replications are reported in Tables A10-A13 (Appendix A).

I want to stress that with these replications I do not intend to criticise in any way the concerned studies. In fact, I believe that all the four replicated studies have made an impressive contribution to the literature on the nexus between state capacity and democracy. Even if the replications provide us also interesting information about the robustness of the previous findings in relation to the chosen set of measures of state capacity, the primary objective of the replication analysis remains to evaluate the interchangeability of our common measures of state capacity. To ease the comparability of the estimations, measures of state capacity are normalised to range from 0 (low) to 1 (high). First, I replicate the three longitudinal models in chronological order. Then, I replicate Charron and Lapuente's (2010) cross-section model.

I begin with Bäck and Hadenius' (2008) study, where the authors find evidence of a curvilinear relationship between democracy and state capacity: at low levels of democracy the relationship is negative, while at high levels of democracy the relationship is positive. To operationalise state capacity the authors aggregate *Bureaucracy Efficiency* and *Corruption* from PRS Group's ICRG into an additive index that covers the period of time from 1984 to 2002. Only three (QOG, HSI, VDEM) of our seven frequently used measures of state capacity cover the entire period of Bäck and Hadenius's study, and thus, the original study is replicated only with three "alternative" models. A summary of the regression results is presented in Table 2.5.

The "original" Model 1 confirms that the relationship between democracy and state capacity is curvilinear. As claimed by Bäck and Hadenius (2008), at low levels of democracy the "effect" of democracy is negative, whereas at high levels of democracy it is positive. In Model 2 state capacity is measured with QOG. Now the predicted effect is

Table 2.5. Replication of Bäck and Hadenius (2008) and re-analysis with alternative measures of state capacity

	Replication (1)	QOG (2)	HSI (3)	VDEM (4)
Democracy	-0.004*	-0.004	0.00003	0.003*
	(0.002)	(0.002)	(0.001)	(0.001)
Democracy^2	0.001**	0.001*	0.00005	-0.0002
	(0.0002)	(0.0002)	(0.0001)	(0.0001)
Ln(GDP/capita)	0.003	0.004	0.008***	0.004***
	(0.002)	(0.002)	(0.002)	(0.001)
Trade	0.00001	0.00003	0.00001	-0.00002*
	(0.00003)	(0.00003)	(0.00001)	(0.00001)
British colony	0.005	0.005	0.001	-0.002
	(0.003)	(0.003)	(0.001)	(0.002)
Lagged dependent variable	0.936***	0.935***	0.918***	0.961***
	(0.018)	(0.017)	(0.023)	(0.011)
Constant	0.004	0.003	-0.018*	-0.013*
	(0.011)	(0.012)	(0.008)	(0.006)
R^2	0.96	0.97	0.97	0.97
Countries	132	132	150	158
Observations	1979	1979	2317	2443

Note: Pooled OLS models with panel-corrected standard errors in parentheses; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Model 1 is a replication of Model 3 in Table 1 in Bäck and Hadenius (2008); dependent variable in Model 1 is an additive index of two ICRG indicators (PRS Group): *Corruption* and *Bureaucracy Quality*. Models reproduce the specification and estimation methods of the original study. Independent variables are taken from the QoG Standard Dataset (Teorell et al. 2019). Democracy = fh_ipolity2. GDP/capita = wdi_gdpcapcon2010. Trade = wdi_trade. British colony = ht_colonial.

similar, but the main democracy term is significant only at a lower (90%) level of confidence. The strong equivalence between the two models is not surprising because the original measure of state capacity is based on almost the same sub-indicators than QOG. On the contrary, however, the curvilinear relationship does not hold even closely in Model 3, in which state capacity is measured with HSI. Model 4 provides even more astonishing results compared to Bäck and Hadenius' (2008) original results. Model 4, in which state capacity is measured with VDEM, provides evidence of a curvilinear relationship between democracy and state capacity (although the squared democracy term is significant only at the 90% level) but the curvilinear relationship is completely opposite than in the original model. Weakly, but still, Model 4 suggests that democracy and state capacity are positively related at low levels of democracy but negatively related at high levels of democracy.

Average marginal effect plots (Figure 2.2) paint a more detailed picture of the consequences of choosing one measure over another. With Bäck and Hadenius' (2008) measure of state capacity the effect of democracy is likely to be negative in countries with a complete absence of democracy (e.g., North Korea from 1994 to 2002 and Saudi Arabia from 1992 to 2002). The effect of democracy is nonsignificant in countries with a low level of democracy but becomes positive and statistically significant in countries with an intermediate or high level of democracy ($>= 5$). Considering the levels of democracy in 2002, this means that already in countries such as Russia, Nigeria, and Burkina Faso the relationship between democracy and state capacity is statistically significant and positive.

Figure 2.2. Average marginal effects of democracy on state capacity, conditional on the level of democracy: replications of Bäck and Hadenius (2008)
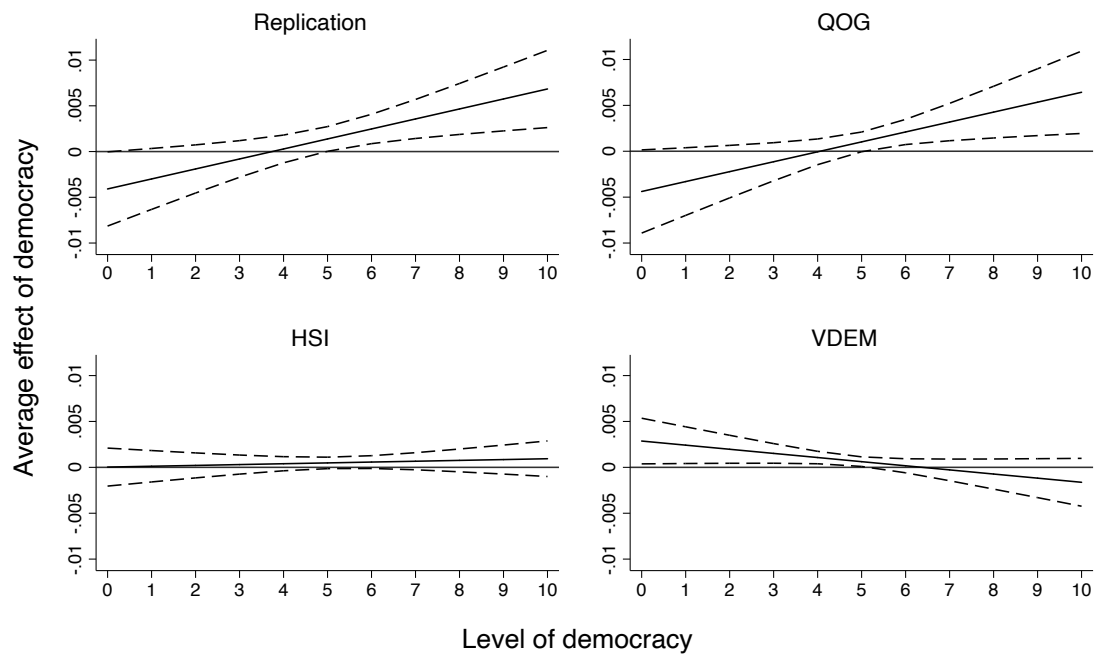
The results are more or less equivalent when state capacity is measured with QOG. When state capacity is measured with VDEM instead the results point towards the opposite direction. From low to intermediate levels of democracy (<6) the relationship

between democracy and state capacity is positive and statistically significant. Considering again the levels of democracy in 2002, this means that the effect of democracy is positive both in completely undemocratic countries such as North Korea, Saudi Arabia, and Iraq and partially democratic countries like Russia, Nigeria, and Burkina Faso. With VDEM, however, the relationship between democracy and state capacity becomes nonsignificant as the level of democracy increases. When state capacity is measured with HSI, instead, the results provide no evidence of any statistically significant curvilinear association between democracy and state capacity.

Second, I test whether Carbone and Memoli's (2015) findings are sensitive to the choice of the measure of state capacity (Table 2.6). Model 1 is replicated with the original measure used in Carbone and Memoli's study, where *Monopoly on the Use of Force* and *Basic Administration* from Bertelsmann Stiftung are multiplicatively aggregated. The original model provides strong evidence about a curvilinear relationship between democracy and state capacity. At extremely low levels of democracy the relationship is inverse, but it turns positive after a certain level of democracy has been reached. In Models 2-8 the original measure is replaced, one by one, with our alternative measures of state capacity. Even if both the main democracy and the democracy^2 terms share the same positive signs across models, choosing one measure over another can lead to completely different inferences. Among the alternative models, the strong curvilinear association between democracy and state capacity holds only with FSI and SFI. When state capacity is measured with CPI, neither of the democracy terms is statistically significant. When state capacity is measured with WGI and VDEM only the main democracy term is statistically significant. With QOG and HSI both democracy terms are statistically significant at conventional levels, but the curvilinear relationship seems to be much weaker than with FSI and SFI.

Table 2.6. Replication of Carbone and Memoli (2015) and re-analysis with alternative measures of state capacity

	Replication (1)	FSI (2)	QOG (3)	HSI (4)	SFI (5)	WGI (6)	CPI (7)	VDEM (8)
Democracy	0.089***	0.086***	0.038*	0.036**	0.087***	0.031*	0.029	0.067**
	(0.022)	(0.014)	(0.016)	(0.011)	(0.014)	(0.015)	(0.018)	(0.022)
Democracy^2	0.069***	0.056***	0.020*	0.025*	0.052***	0.015	0.013	0.018
	(0.019)	(0.014)	(0.010)	(0.011)	(0.012)	(0.013)	(0.014)	(0.015)
Duration of democracy	-0.003	-0.001	-0.0004	-0.0001	0.001	-0.0001	0.00003	0.0002
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Duration of democracy^2	-0.00001	-0.00001	0.00001	0.000003	-0.00001	0.00001	-0.00001	-0.000004
	(0.00003)	(0.00001)	(0.00002)	(0.00001)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
Democracy × Duration of democracy	0.003*	0.001	-0.0003	-0.001	-0.001	0.00001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ethnic fractionalization	-0.093	-0.120**	-0.061	-0.100**	-0.171***	-0.073	0.021	0.064
	(0.069)	(0.045)	(0.049)	(0.036)	(0.043)	(0.040)	(0.039)	(0.040)
Log(GDP/capita)	0.099***	0.032***	0.031***	0.054***	0.087***	0.040***	0.045***	0.030***
	(0.012)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.008)	(0.007)
Log(Area)	-0.025**	-0.017*	-0.007	-0.005	-0.025***	-0.017*	-0.026**	-0.013*
	(0.008)	(0.007)	(0.009)	(0.006)	(0.005)	(0.007)	(0.009)	(0.007)
Constant	0.063	0.356***	0.265*	0.214**	0.276**	0.352***	0.206*	0.247*
	(0.152)	(0.098)	(0.121)	(0.080)	(0.094)	(0.089)	(0.096)	(0.101)
Sigma_u	0.15	0.08	0.11	0.07	0.10	0.09	0.10	0.11
Sigma_e	0.07	0.02	0.02	0.02	0.03	0.02	0.03	0.03
Rho	0.82	0.93	0.97	0.90	0.90	0.95	0.90	0.93
Between R^2	0.64	0.71	0.39	0.72	0.81	0.58	0.55	0.49
Wald chi-square (8)	248.69	207.60	72.37	284.75	733.60	175.38	98.66	91.02
Prob > chi-square	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Countries	122	121	99	121	122	122	122	122
Observations	344	344	284	344	345	345	343	345

Note: Random effect models with robust standard errors in parentheses; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Model 1 is a replication of Model 4 in Table 1 in Carbone and Memoli (2015); dependent variable in Model 1 is a multiplicative index of two Bertelsmann Stiftung indicators: *Monopoly on the Use of Force* and *Basic Administration*. Models reproduce the specification and estimation methods of the original study. Model 1 is run with original data provided by the authors; all independent variables are taken from this original dataset; see information about sources in Carbone and Memoli (2015). HSI provides data only until 2009. Thus, I have coded its scores in 2010 equal to its scores in 2009.

As before, average marginal effect plots (Figure 2.3) can guide us in a more exhaustive interpretation of the results. The plots show that the main finding of the original model holds only in two of the alternative models. Using FSI or SFI leads to relatively similar findings compared to the original model. Using QOG, HSI, or VDEM instead leads to the interpretation that the positive effect of democracy on state capacity begins only after a country has reached an intermediate level of democracy. With none of the three measures, anyhow, democracy has a statistically significant negative effect on state capacity in undemocratic countries. On the contrary, models with WGI and CPI do not support any of the above findings. With WGI the statistically significant effect of democracy on state capacity holds only at intermediate levels of democracy. With CPI instead the relationship between democracy and state capacity does not seem to be dependent on the level of democracy at all.

Figure 2.3. Average marginal effects of democracy on state capacity, conditional on the level of democracy: replications of Carbone and Memoli (2015)
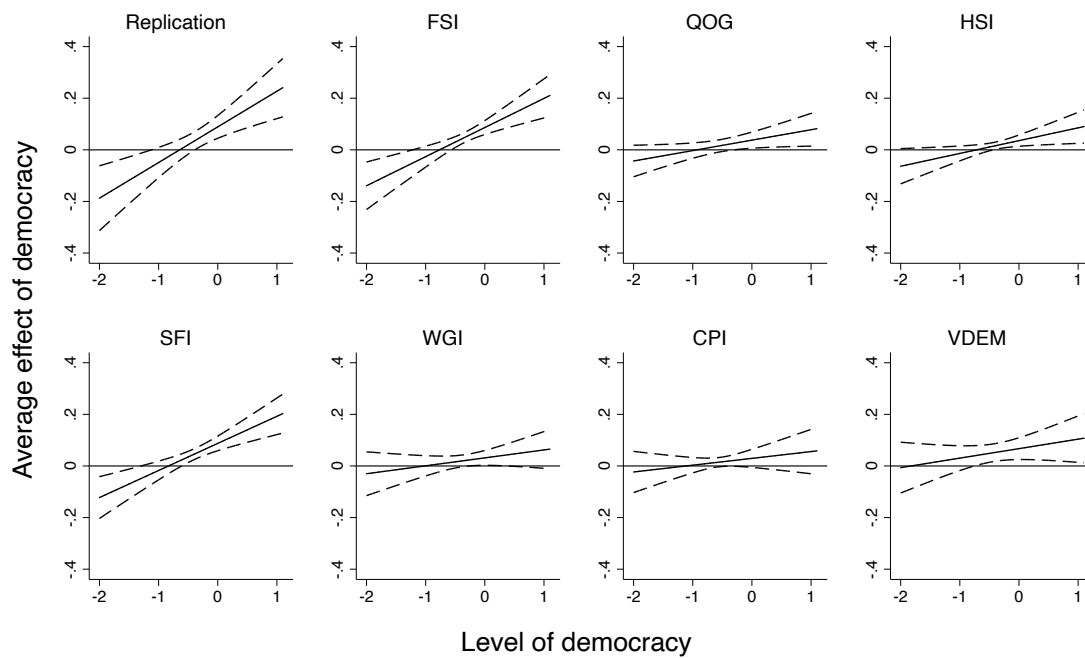
The original study investigates also the role of democratic duration in the relationship between democracy and state capacity. Model 1 confirms that "democratic duration becomes a crucial [*positive*] factor when combined with the degree of democracy" (Carbone and Memoli 2015: 18).[5] Interestingly, however, such interpretation is not supported by any of the alternative models.

Third, I replicate Grassi and Memoli's (2016) study and assess the external validity of its findings to different measures of state capacity (Table 2.7). The discrepancies between the original model – where state capacity is quantified with HSI – and the alternative models are even more pronounced than in the two previously discussed sets of regressions. The original model indicates a significant non-linear relationship between democracy and state capacity: this relationship is negative in autocratic countries but fades out once a country reaches a certain level of democratisation. Moreover, the original model suggests that left-wing executives have fostered the creation of state capacity. The former finding is not supported by any of the alternative models. The latter finding is confirmed only by one of the alternative models.

In the original model both the main democracy term and its quadratic term are statistically significant at conventional levels. The main democracy term has a positive coefficient, whereas the quadratic terms has a negative coefficient. With QOG the main term is significant at the 99.9% level and the quadratic term is very close to conventional significance levels (significant at the 90% level), but the point estimates suggest a completely opposite story compared to the original model. With QOG the main term is positive and the quadratic term is negative. With WGI or VDEM neither of the democracy

---

[5] Text in square brackets added by author.

Table 2.7. Replication of Grassi and Memoli (2016) and re-analysis with alternative measures of state capacity

	Replication (1)	QOG (2)	SFI (3)	WGI (4)	CPI (5)	VDEM (6)
Lagged stateness	0.169***	-0.040	-0.016	0.011	0.017	0.018
	(0.021)	(0.049)	(0.035)	(0.015)	(0.026)	(0.015)
Democracy	-0.034***	0.046***	-0.009	-0.002	0.012*	-0.005
	(0.007)	(0.010)	(0.006)	(0.008)	(0.005)	(0.012)
Democracy^2	0.004*	-0.005	0.005**	0.0004	-0.001	0.005
	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
Executive partisan balance	0.133**	-0.070	0.116***	-0.054	-0.108*	0.069
	(0.042)	(0.039)	(0.031)	(0.083)	(0.050)	(0.076)
Level of economic development (1994)	0.095	0.075	0.286***	0.218***	0.147*	0.185**
	(0.059)	(0.067)	(0.048)	(0.059)	(0.074)	(0.069)
Land size (km^2)	0.018	0.008	-0.029	-0.007	-0.002	-0.019
	(0.020)	(0.023)	(0.016)	(0.021)	(0.025)	(0.022)
Oil rents	-0.130	-0.072	-0.142	-0.290**	-0.146*	-0.090
	(0.163)	(0.127)	(0.146)	(0.103)	(0.068)	(0.162)
Log(Ethnic fractionalization)	-0.068	-0.054	-0.155*	-0.021	-0.103	-0.082
	(0.128)	(0.156)	(0.071)	(0.141)	(0.155)	(0.090)
Constant	-0.681	-0.376	-1.616***	-1.232*	-0.902	-1.091
	(0.467)	(0.381)	(0.400)	(0.498)	(0.625)	(0.563)
Sigma_u	0.17	0.19	0.15	0.16	0.17	0.19
Sigma_e	0.05	0.09	0.08	0.05	0.06	0.05
Rho	0.91	0.80	0.79	0.92	0.90	0.94
Between R^2	0.30	0.15	0.76	0.59	0.38	0.49
Wald chi-square (8)	1393.54	228.57	170.67	52.05	78.00	25.47
Prob > chi-square	0.000	0.000	0.000	0.000	0.000	0.001
Countries	18	18	18	18	18	18
Observations	197	197	197	161	186	197

Note: Random effects models with robust standard error in parentheses; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Model 1 is a replication of Model 3 in Table 1 in Grassi and Memoli (2016); dependent variable in Model 1 is HSI. Models reproduce the specification and estimation methods of the original study. Model 1 is run with original data provided by the authors; all independent variables are taken from this original dataset; see information about sources in Grassi and Memoli (2016). FSI is excluded because it provides data only from 2005 onwards.

terms are significant. With SFI only the squared term is significant whereas with CPI only the main term is significant.

With the help of average marginal effect plots (Figure 2.4) we can examine more in detail how the predicted impact of democracy is sensitive to the chosen measurement. In the original model, in which state capacity is measured with HSI, the initially negative marginal effect of democracy gradually fades out when the level of democracy increases. On the contrary, with QOG, democracy is positively related to state capacity in autocracies, but once a certain level of democratisation has been reached, the effect becomes not different from zero. With CPI and VDEM, the relationship between democracy and state capacity is statistically significant and positive only at intermediate levels of democracy, whereas with WGI, the average effect of democracy on state capacity does not seem to depend on the level of democracy. The model with SFI suggests

Figure 2.4. Average marginal effects of democracy on state capacity, conditional on the level of democracy: replications of Grassi and Memoli (2016)
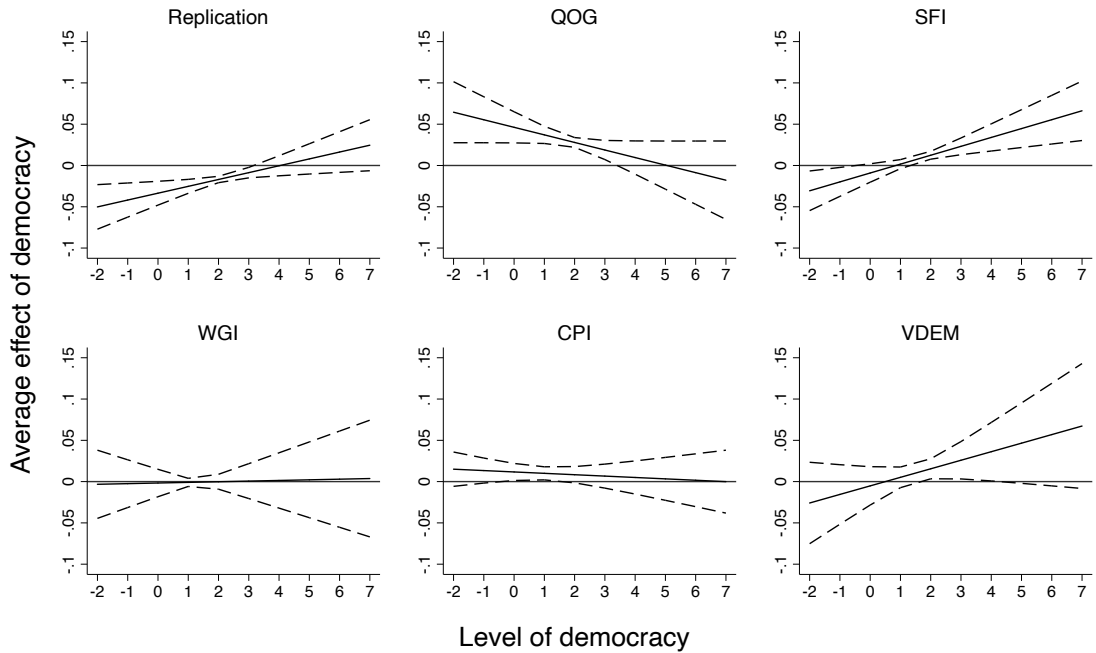
that democracy is positively related to state capacity only after a certain level of democracy has been reached. With SFI, democracy is negatively and significantly related to state capacity in the most undemocratic countries.

As to the partisan balance of the executive – another main explanatory variable in the original study – the models with HSI and SFI find a positive relationship between left-wing executives and state capacity. Conversely, with CPI, it turns out that right-wing executives have a statistically significant and positive relationship with state capacity. The model with QOG supports the latter finding, although only at a weaker (90%) level of statistical significance. According to the models in which WGI or VDEM are the dependent variables, state capacity is not related to the partisan balance of the executive at conventional levels of statistical significance.

Last, I replicate one of the cross-section models from Charron and Lapuente's (2010) study (Table 2.8). The authors of the original study find that the interaction between democracy and economic development is positive and significant, but that the evidence of a curvilinear association between democracy and state capacity is weak, because the squared democracy term does not reach conventional levels of statistical significance. First, Model 1 is replicated with QOG, which is the measure used in the original study. Then, one by one, the original model is re-estimated with the alternative measures of state capacity (Models 2-7). Last, in the "supplementary" Model 8, I complement the interchangeability tests with Carbone and Memoli's (2015) index of state capacity (CMI).

Let us start with the results concerning the interactive effect between democracy and economic development. The original results are confirmed by Model 1 but are not robust to some of the re-estimated models. The interaction term between democracy and GDP/capita is positive and statistically significant at least at the 99% level with FSI, HSI, WGI, CPI, and VDEM, but nonsignificant at conventional levels with SFI and CMI.

Table 2.8. Replication of Charron and Lapuente (2010) and re-analysis with alternative measures of state capacity

	Replication (1)	FSI (2)	HSI (3)	SFI (4)	WGI (5)	CPI (6)	VDEM (7)	CMI (8)
Democracy	-0.284***	-0.176***	-0.115***	-0.030	-0.199***	-0.355***	-0.192***	-0.168*
	(0.041)	(0.033)	(0.026)	(0.035)	(0.030)	(0.054)	(0.031)	(0.078)
Democracy^2	0.005	0.008**	0.004*	0.009***	0.003	0.007	0.006**	0.019***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)	(0.004)
Ln(GDP/capita)	-0.085*	-0.014	0.061**	0.158***	-0.044	-0.096	-0.094***	0.077
	(0.037)	(0.028)	(0.022)	(0.029)	(0.025)	(0.052)	(0.026)	(0.059)
Democracy × Ln(GDP/capita)	0.028***	0.015***	0.010**	-0.004	0.022***	0.034***	0.020***	-0.002
	(0.005)	(0.004)	(0.003)	(0.004)	(0.004)	(0.007)	(0.004)	(0.010)
Trade openness	0.0002	0.0002	-0.0004*	0.0004	-0.0002	-0.0001	-0.0002	0.001
	(0.0003)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0002)	(0.001)
British colony	0.040	0.002	0.021	0.001	0.027	0.057	-0.023	-0.006
	(0.032)	(0.027)	(0.020)	(0.027)	(0.021)	(0.038)	(0.024)	(0.048)
Constant	1.142***	0.452	0.007	-0.775**	0.726***	1.160**	1.141***	0.110
	(0.303)	(0.229)	(0.181)	(0.242)	(0.206)	(0.425)	(0.218)	(0.482)
R^2	0.75	0.85	0.83	0.84	0.83	0.79	0.79	0.56
Countries/Observations	108	113	120	122	133	93	126	89

Note: Cross-sectional OLS models with standard errors in parentheses; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Model 1 is a replication of Model 2 in Table 2 in Charron and Lapuente (2010); dependent variable in Model 1 is QOG. Models reproduce the specification and estimation methods of the original study. Independent variables are taken from the QoG Standard Dataset (Teorell, Holmberg, and Rothstein 2008): Democracy = fh_ipolity2; GDP/capita = gle_rgdp; Trade = pwt_openk; British colony = ht_colonial; CMI measured in 2006, because coded only from 2006 onwards. FSI measured in 2005, because coded only from 2005 onwards.

Additionally, the choice of the measure seems to cause substantial divergences in the main GDP/capita term. When democracy is 0, the effect of economic development on state capacity is statistically significant and negative in models with QOG and VDEM, statistically significant and positive in models with HSI and SFI, and nonsignificant in the remaining four models. As regards to the conditioning effect of the level of democracy, the beta coefficient of the main democracy term is negative and statistically significant in all models except the one with SFI. The estimated beta coefficient of the quadratic democracy term is positive across models and reaches statistical significance at least at the 95% level in all models except the ones with QOG, WGI, and CPI.

Average marginal effect plots (Figure 2.5) of the relationship between democracy and state capacity, conditional on the level of economic development, portray a more detailed picture of the differences across models. In the original model, in which state capacity is measured with QOG, the relationship between democracy and state capacity is negative in extremely poor countries but becomes positive after a certain level of economic development is reached. Such result is confirmed if state capacity is measured with WGI or CPI, albeit with some differences in the estimated level of economic development at which the relationship between democracy and state capacity becomes positive. With WGI, the relationship is already positive in countries with a GDP/capita of around 2,500$ or more such as Cambodia. With CPI, instead, the relationship is only positive in countries with a GDP/capita of around 5,500$ or more such as Romania. Analogously, with FSI, HSI, and VDEM democracy is positively related to state capacity only after a certain level of economic development. Yet, with these three measures, there is no statistically significant relationship between democracy and state capacity in extremely poor countries. With SFI the relationship between democracy and state

capacity is positive regardless of the level of economic development. With CMI the results are similar, albeit nonsignificant at extremely low levels of GDP/capita.

Figure 2.5. Average marginal effects of democracy on state capacity, conditional on economic development: replications of Charron and Lapuente (2010)
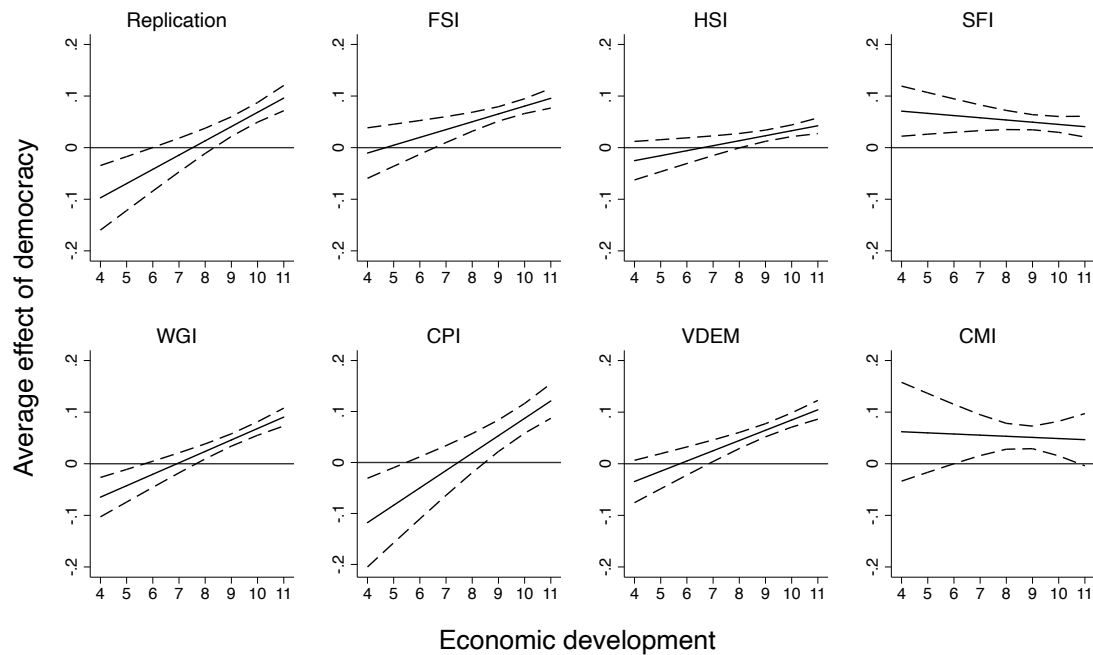
As regards to the curvilinear relationship between democracy and state capacity (i.e., conditional on the level of democracy), average marginal effect plots (Figure 2.6) show relatively similar patterns across models. Regardless of the chosen measure, the average effect of democracy on state capacity is positive at intermediate and high levels of democracy. The effect is statistically significant at conventional levels with no exceptions. The magnitude of the effect, however, is sensitive to the chosen measure of state capacity. With CMI, the magnitude of the predicted positive effect is particularly large. At low levels of democracy, the relationship between democracy and state capacity is negative and statistically significant at least at the 95% level with CMI and SFI, and at a lower (90%) level of confidence also with FSI. With the remaining measures of state

capacity, there is no clear evidence that democracy affects state capacity at low levels of democracy.

Figure 2.6. Average marginal effects of democracy on state capacity, conditional on the level of democracy: replications of Charron and Lapuente (2010)
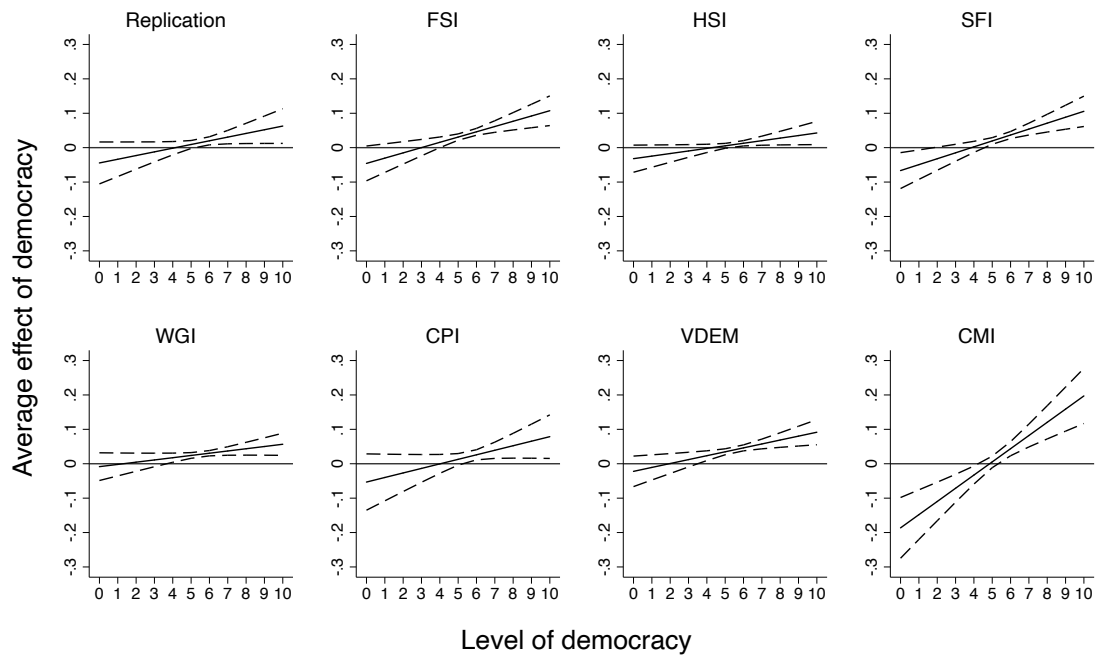
The extensive statistical exercise of replicating four studies with seven[6] common measures of state capacity has shown convincingly that the choice of the measure plays a key role in affecting the conclusions drawn from the replicated studies, undermining both the interchangeability of the measures and the external validity of these studies to alternative measures of state capacity. Since measures do not always cover the same sample of countries, my findings could be driven by different samples rather than different measures. To rule out so called "selection bias", I run the previous sets of models with a common sample of observations. The results (Tables A14-A17, Appendix A) are

---

[6] Eight measures, if the "supplementary" CMI is taken into account.

not substantially affected by restricting the models within each set of replications to the same sample and the main conclusions are not altered by using a set of common observations. Selection bias does not influence the interpretation of our findings. Different conclusions are driven by heterogeneity in measures, not samples.

We have found strong evidence that some of the most established measures of state capacity are highly correlated among one another and represent statistically the same unidimensional concept. Anyhow, our findings have shown as well that highly correlated measures can lead to completely opposing conclusions, even if regressed on exactly the same set of predictors with the same estimation methods. These findings suggest that researchers should not take for granted that strongly correlated measures lead to similar inferences. Furthermore, these findings indicate that overall the interchangeability of measures of state capacity is weak and the external validity of the replicated studies is weak. Not even one single pair of measures produces consistently similar results across the replicated models, although WGI and CPI seem to be the most interchangeable pair of measures.

It is worrisome that previous findings on the linkage between democracy and state capacity can be so sensitive to the chosen measure. For instance, my findings show that when Charron and Lapuente's (2010) and Carbone and Memoli's (2015) original models are cross validated by switching the two originally used measures of state capacity (respectively QOG and CIM), the original conclusions of the two studies do not hold. It would be reckless, however, to deduce that the original studies are necessarily erroneous. Conversely, my findings simply indicate that the conclusions drawn from the replicated models are not generalisable across measures of state capacity. Next, I shift the attention towards a more detailed analysis of the root cause of the lack of interchangeability, which is the disagreement in individual country scores.

Country-Specific Rating Discrepancy

In the previous section, we have found robust evidence that the interchangeability of common measures of state capacity is weak. Research results, thus, are affected by the particular measure used. This is true, even if measures are strongly correlated among one another. These somewhat contradictory findings require further investigation, and it is likely that we will better understand what causes our contradictory findings if we turn our attention at country-level scores.

With bivariate scatter plots of state capacity measures (Figure A1, Appendix A) we can understand better how individual countries are rated by each measure. Country-scores are analysed in year 2009, for the simple reason that 2009 is the most recent year of common observations. Overall, we can notice that many countries are rated with high consistency by each pair of measures, as suggested previously by the correlation analysis. For example, Somalia has an extremely low score in all measures, whereas the Nordic Countries, Switzerland, and New Zealand have extremely high scores in all measures. Yet, it becomes evident that there are also countries that are rated with substantial disagreements by our measures. Keeping in mind that the measures are normalised to range from 0 to 1, some of the bivariate rating divergences are astonishingly large (Table A18, Appendix A).

We already know from our analysis of the statistical properties of the measures that SFI tends to give countries higher and CPI tends to give countries lower scores than the other measures. Thus, it is not surprising to find that there are considerable country-level disagreements between SFI and CPI. In fact, there are as many as 45 countries that SFI rates at least 0.40 units higher than CPI. In seven of these, the discrepancy between the two ratings is more than 0.60 units: Argentina (0.70), Belarus (0.68), Jamaica (0.65), Albania (0.63), Ukraine (0.63), Greece (0.63), and Italy (0.61). Likewise, differences

between SFI and VDEM are substantial: SFI rates Belarus 0.71 units higher than VDEM, and in total there are 34 countries that SFI rates at least 0.40 units higher than VDEM.

In general, SFI rates countries with particularly large disagreements compared to the other measures. SFI rates Belarus 0.60 units higher than WGI, and there are six other countries that are rated at least 0.40 units higher by SFI than by WGI. SFI rates Albania 0.55 units higher than QOG, and there are 20 countries that are rated at least 0.40 units higher by SFI than by QOG. SFI rates Belarus 0.47 units higher than FSI, and there are five other countries that are rated with a discrepancy of at least 0.40 units between SFI and FSI. Country-specific disagreements between SFI and HSI instead are comparatively small. The most discrepantly rated country is Argentina, which is rated 0.36 units higher by SFI than by HSI.

Differences in country ratings between HSI and the other five measures are relatively large. HSI rates 21 countries at least 0.40 units higher than CPI. Six of these countries are rated at least 0.50 units higher: Iran (0.59), Russia (0.54), Venezuela (0.53), Belarus (0.52), Armenia (0.50), and Kazakhstan (0.50). For example, Iran has a score of 0.67 with HSI but a score of only 0.08 with CPI. HSI rates seven countries at least 0.50 units higher than VDEM: Egypt (0.66), Belarus (0.55), Kuwait (0.55), Malaysia (0.53), Tunisia (0.50), Azerbaijan (0.50), and Kazakhstan (0.50). Iran is the country with the largest rating discrepancy between HSI and FSI. HSI rates Iran 0.44 units higher than FSI. Venezuela instead is the country with the largest discrepancy between HSI and QOG. HSI rates it 0.50 units higher than QOG. As to disagreements between HSI and WGI, Belarus is rated 0.44 units higher by HSI; it is the only country rated by the two indices with a discrepancy larger than 0.40 units.

Due to CPI's comparatively low scores, it is not surprising to find that there are nine countries rated at least 0.30 units higher by FSI than by CPI, but no countries rated at

least 0.30 units higher by CPI than by FSI. The country with the largest rating discrepancy between the two measures is Argentina, rated 0.50 units higher by FSI. A somewhat similar pattern can be found when comparing the country ratings of CPI and VDEM. Four countries are rated at least 0.30 units higher by VDEM, and only one country is rated at least 0.30 units higher by CPI. As to the ratings in CPI and QOG, Iran is the country with the highest discrepancy. QOG rates Iran 0.45 while CPI rates Iran 0.08, meaning that its score is 0.37 units higher with QOG. WGI and CPI rate countries in a relatively similar way. This might explain also why WGI and CPI appeared to be the most interchangeable pair of measures in the previous section. Philippines is the country with the rating discrepancy: its score is 0.33 units higher with WGI than with CPI.

The country with the largest discrepancy between WGI and FSI is Cyprus, which is rated 0.81 by WGI and 0.48 by FSI. WGI and QOG tend to rate countries relatively similarly: there are no country scores with a discrepancy of more than 0.30. Differences in country scores between WGI and VDEM are more pronounced. There are three countries with a discrepancy of more than 0.40 units between the two measures: Tunisia (0.45), Malaysia (0.45), and Egypt (0.42). In each of the three cases WGI assigns a higher score than VDEM. As to QOG and VDEM, only Egypt is rated with a difference of more than 0.40 units between the two measures. With VDEM its score is 0.01 whereas with QOG its score is 0.42. Differences in country scores between QOG and FSI are relatively small too. Only two countries are rated with a discrepancy of more than 0.30 units. The largest rating discrepancy between FSI and VDEM is about Libya, which is rated 0.05 by VDEM and 0.47 by FSI. Hence, the level of state capacity in Libya is 0.42 units higher with FSI than VDEM. There are no other countries that FSI rates more than 0.40 units higher than VDEM, or vice versa.

The above analysis proves that measures of state capacity do not rate countries similarly. Given that bivariate differences in country scores between measures can be so large, it becomes more understandable that the interchangeability of measures is weak. To sum up, overall, the single observations with the largest discrepancies (Table 2.9) have relatively high scores with SFI or HSI and relatively low scores with CPI and VDEM.

Table 2.9. Observations with the highest (> 0.50 units) bivariate discrepancy between measures of state capacity in 2009

Observation (country)	Discrepancy (in units)	Higher score	Lower score
Belarus	0.71	0.84 with SFI	0.13 with VDEM
Argentina	0.70	0.92 with SFI	0.22 with CPI
Belarus	0.68	0.84 with SFI	0.16 with CPI
Libya	0.67	0.72 with SFI	0.05 with VDEM
Egypt	0.66	0.67 with HSI	0.01 with VDEM
Cuba	0.65	0.76 with SFI	0.11 with VDEM
Jamaica	0.65	0.88 with SFI	0.23 with CPI
Tunisia	0.63	0.76 with SFI	0.13 with VDEM
Albania	0.63	0.88 with SFI	0.25 with CPI
Greece	0.63	0.96 with SFI	0.33 with CPI
Ukraine	0.63	0.76 with SFI	0.13 with CPI
Italy	0.61	1.00 with SFI	0.39 with CPI
Kuwait	0.61	0.88 with SFI	0.27 with VDEM
Belarus	0.60	0.84 with SFI	0.24 with WGI
Iran	0.59	0.67 with HSI	0.08 with CPI
Mexico	0.57	0.84 with SFI	0.27 with CPI
Bahrain	0.56	0.84 with SFI	0.28 with VDEM
Albania	0.55	0.88 with SFI	0.33 with QOG
Belarus	0.55	0.68 with HSI	0.13 with VDEM
Bulgaria	0.55	0.88 with SFI	0.33 with CPI
Kuwait	0.55	0.82 with HSI	0.27 with VDEM
Latvia	0.55	0.96 with SFI	0.41 with CPI
Libya	0.55	0.72 with SFI	0.17 with CPI
Russia	0.55	0.68 with SFI	0.13 with CPI
Slovakia	0.55	0.96 with SFI	0.41 with CPI
Albania	0.54	0.88 with SFI	0.34 with VDEM
Costa Rica	0.54	0.96 with SFI	0.42 with QOG
Russia	0.54	0.67 with HSI	0.13 with CPI
Trinidad-Tobago	0.54	0.84 with SFI	0.30 with CPI
Armenia	0.53	0.72 with SFI	0.19 with CPI
Dominican Rep.	0.53	0.76 with SFI	0.23 with CPI
Malaysia	0.53	0.79 with HSI	0.26 with VDEM
Mongolia	0.53	0.72 with SFI	0.19 with CPI
Poland	0.53	1.00 with SFI	0.47 with CPI
Venezuela	0.53	0.63 with HSI	0.10 with CPI
Vietnam	0.53	0.72 with SFI	0.19 with CPI
Belarus	0.52	0.68 with HSI	0.16 with CPI
Bulgaria	0.52	0.88 with SFI	0.36 with QOG
Croatia	0.52	0.88 with SFI	0.36 with CPI

Hungary	0.52	1.00 with SFI	0.48 with CPI
Kuwait	0.52	0.88 with SFI	0.36 with CPI
Paraguay	0.52	0.64 with SFI	0.12 with CPI
Lebanon	0.51	0.68 with SFI	0.17 with CPI
Morocco	0.51	0.76 with SFI	0.25 with VDEM
Serbia	0.51	0.80 with SFI	0.29 with CPI
Vietnam	0.51	0.72 with SFI	0.21 with VDEM

Note: Scores are normalised to range from 0 to 1.

It is likely that countries such as Belarus, Albania, and Kuwait that are repeatedly among the most divergently rated ones in bivariate terms, stand out also in terms of multivariate discrepancy. To determine multivariate rating discrepancy for a particular country, I calculate the standard deviation of all the country scores (across measures) for a given country in 2009. A higher standard deviation indicates that the ratings of a given country are more spread out across measures, and a lower standard deviation indicates the opposite.

In line with our expectations, Belarus, Albania, and Kuwait are among the countries with the largest multivariate rating discrepancy (Figure 2.7, left panel). This group of countries seems to have fairly heterogeneous characteristics. There are both developed and developing countries and there are both democratic and authoritarian countries; there are no Western liberal democracies besides Italy and Greece. Interestingly, not even one of the countries in the left panel has full civil liberties (i.e., "best" possible score) according to the Freedom House's civil liberties ratings[7] in 2009. Politico-geographically, most of these countries are located in Eastern Europe, the Middle East/North Africa, or Latin America/the Caribbean. Sub-Saharan African countries, instead, are completely absent from the left panel. Nearly half of the 20 most discrepantly rated countries have a Muslim-majority population.

---

[7] Freedom House's Civil Liberties score ranges from 1 to 7; a lower score indicates more civil liberties.

Figure 2.7. Countries with largest and smallest multivariate rating discrepancy in 2009
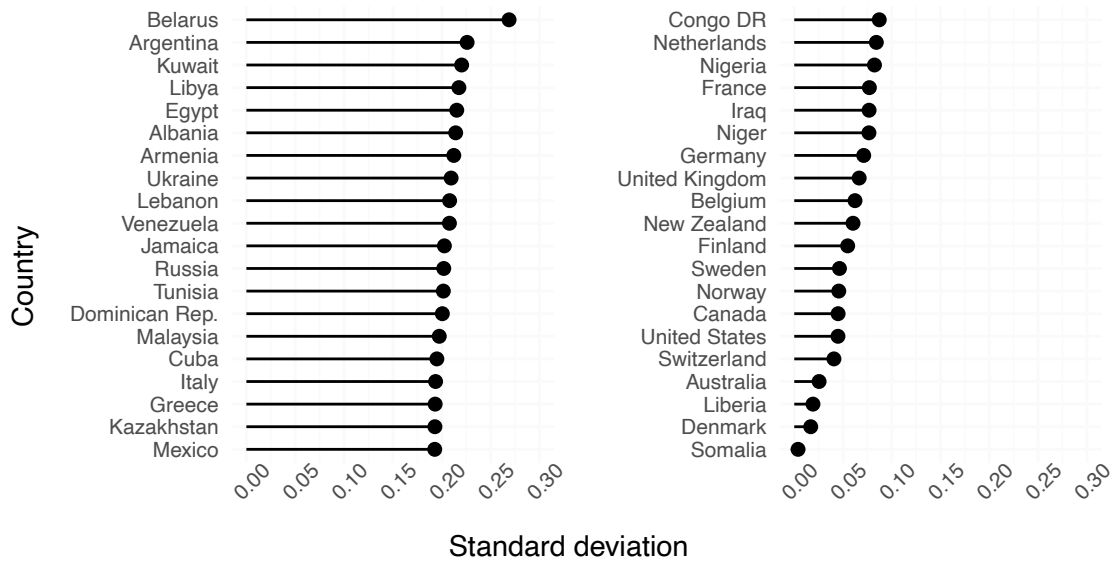
Note: Countries with largest discrepancy in the left panel. Countries with smallest discrepancy in the right panel.

Countries with small rating discrepancy (Figure 2.7, right panel) can be more straightforwardly categorised into two distinct groups: highly dysfunctional states (e.g., Somalia, Iraq, Liberia) and Western liberal democracies (e.g., Denmark, France, Australia). These countries have either very low or very high capacity, but they share in common the characteristic that their scores are more or less equivalent across measures.

Figure 2.8 provides illustrative multivariate information about country scores in the most discrepantly rated countries and confirm a pattern that was previously suggested by bivariate comparisons: many of these countries have relatively high scores with SFI and HSI but relatively low scores with CPI and VDEM. Countries with similar-looking "nets" have also fairly similar scores across different measures (i.e., are multivariately equivalent). For instance, the shapes of the nets of Italy, Greece, Albania, and Ukraine match quite closely: all the four countries have higher ratings with SFI, HSI, and FSI, but lower ratings with the other four measures. Russia, Belarus, and Kazakhstan seem to share

103

Figure 2.8. Spider charts of countries with largest multivariate rating discrepancy in 2009
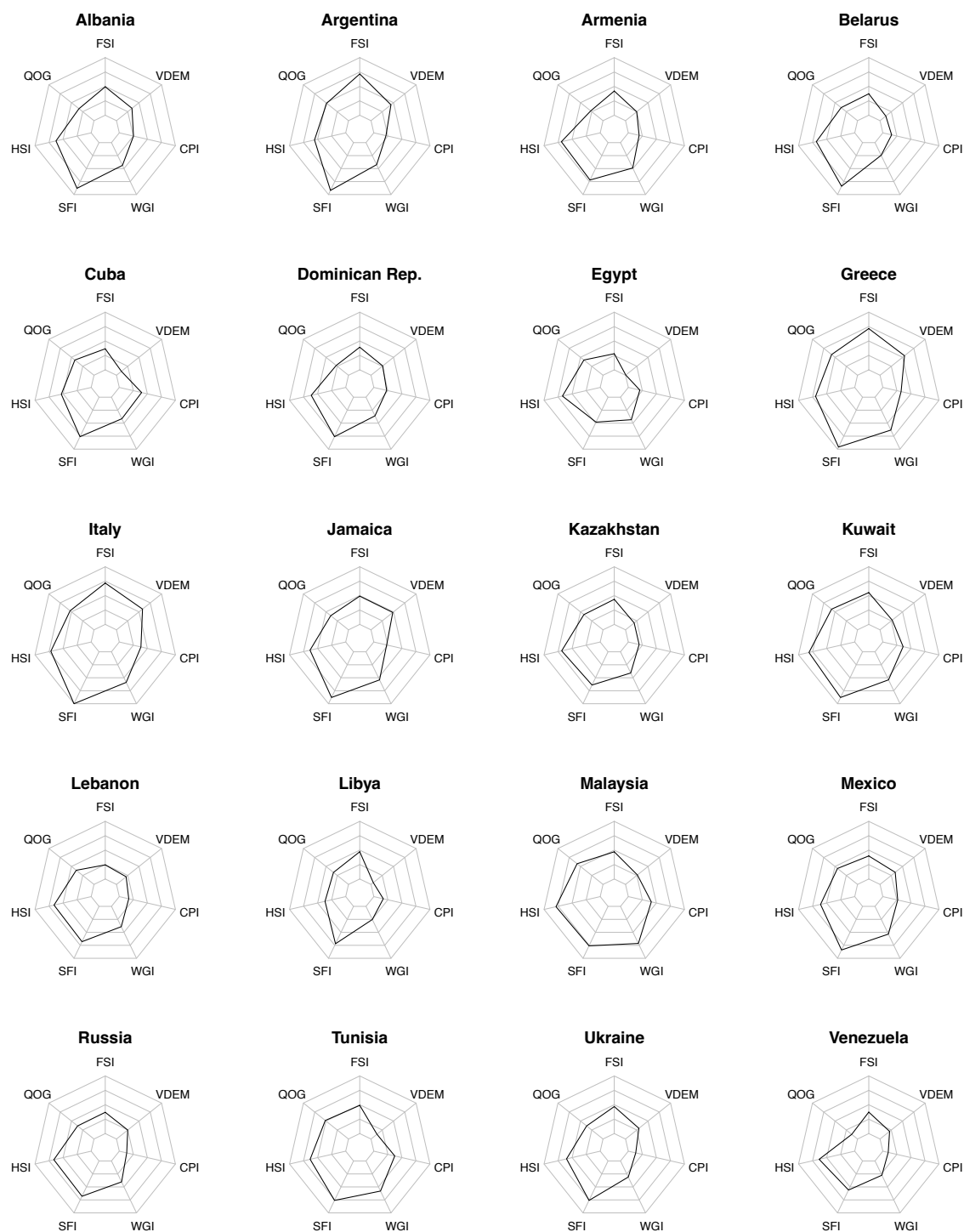
Note: Scores are normalised to range from 0 (low) to 1 (high).

some interesting analogies as well: comparatively high ratings with SFI and HSI, intermediate levels of state capacity with QOG, FSI, and WGI, but relatively low scores with CPI and VDEM. Tunisia, Egypt, Libya, and Cuba instead are rated particularly low with VDEM.

Even if we have previously found robust statistical evidence that all the surveyed measures are related to the same unidimensional construct, our case-based analysis leaves no doubts that measures disagree on many countries. In some cases, the amount of disagreement is impressive. When exploring qualitatively the content of the chosen measures of state capacity, we found out that each measure builds upon slightly different "ingredients". We would expect to find that some of the rating discrepancies reflect at least up to a certain extent these differences in the content and components of the measures.

To recall, VDEM and CPI focus respectively on impartiality and corruption, two strictly procedural attributes of state capacity. SFI and HSI instead capture a broader set of dimensions, but in both indices, the coercive dimension of state capacity plays a relatively important role; particularly so in SFI. For instance, both SFI and HSI contain sub-indicators related to violence, security, and stability. WGI and QOG focus in one way or another on the quality of the bureaucracy, even though the former emphasises also the quality of public services, whereas the latter gives importance to corruption and rule of law. FSI takes into consideration various aspects related to state capacity, such as the provision of public services, the influence of external actors, the ability to collect taxes, rule of law, environmental pressures, structural inequality, economic development, and public finances. With FSI state capacity is understood more broadly than with the other measures.

If we examine the ratings in relation to the defining attributes of each measure in individual countries, we can indeed understand better some of the causes of the country-level inconsistencies in ratings. It seems that many of the countries with high rating divergency are deficient in some of the procedural aspects of state capacity, such as corruption, but exert a strong control on the society (e.g., Belarus, Russia, Kazakhstan, Cuba, Venezuela, Malaysia, Egypt, Kuwait). All these countries tend to have comparatively high scores with SFI and HSI, but lower scores with the other measures. To give an example, Belarus is a stable and ethnically homogeneous country with powerful state institutions (Way 2005) and a high capacity to control the society (Silitski 2005). Its public apparatus is characterised by low autonomy and a lack of impartiality, and its public services are broad ranging but qualitatively deficient (Dimitrova et al. 2020). Given these characteristics, it cannot be a coincidence that Belarus has very high scores with SFI and HSI but much lower scores with the other measures. SFI and HSI focus on some of the areas in which Belarus performs well, but neither of the two measures is focused on corruption or rule of law, which instead, play a bigger role in the other five measures.

The comparative analysis of country ratings and the analysis of rating discrepancy have shown that measures disagree considerably about the level of state capacity in certain countries. Some of these disagreements can be explained through the different aspects of state capacity quantified by each instrument, which is positive news. It would have been worrying to find that the scores do not reflect the content at all. By rigorously choosing a measure that accurately represents a particular theory and by perspicuously defending the choice, scholars can push forward the research agenda on state capacity. Given that the country-level disagreements seem to reflect at least up to a certain extent the different attributes of the surveyed measures, the lack of interchangeability among

these measures is less of a problem if researchers scrupulously opt for a measure that closely matches their theory and conceptualisation of state capacity.

In spite of that, it is less promising to find that rating discrepancy depends systematically on the level state capacity (Figure 2.9). Regardless of the measure, there

Figure 2.9. Rating discrepancy and level of state capacity in 2009
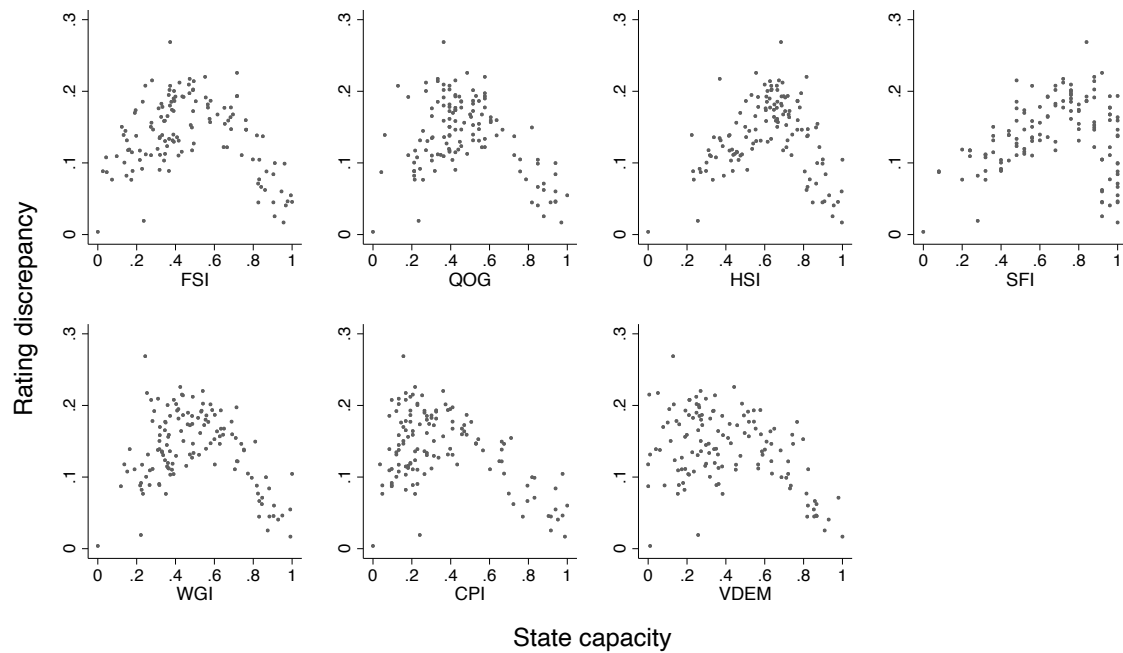
Note: State capacity scores are normalised to range from 0 (low) to 1 (high).

is a non-linear relationship between the level of state capacity and rating discrepancy. Measures tend to agree about countries with extreme levels of state capacity, but the largest rating divergences are systematically at intermediate levels of state capacity. Since all the surveyed measures are based at least partly on subjective perception-based data, the finding is understandable. Survey experts and coders are more likely to agree about clear-cut cases on the extreme ends of the spectrum, whereas less clear cases are simply harder to code, and experts can be expected to have diverging perceptions about state capacity in these countries. Systematic discrepancy, thus, can be credibly explained by

the subjective nature of our measures. Having said that, unfortunately, it seems to affect our knowledge on state capacity even when a given working definition matches perfectly with the selected measure.

Conclusions

In this chapter, I have comprehensively analysed and compared some of the currently most established measures of state capacity. I have evaluated the content, the basic statistical properties, the convergence, the interchangeability, and the country-level rating disagreement of these measures. The analysis has been predominantly statistical but not only. Qualitatively, I have explored the content of the measures and assessed the possible causes of rating discrepancy vis à vis the differences in the content and components of the measures.

The main findings of this chapter are manifold. First and foremost, the chapter at hand provides one of the first systematic comparative statistical assessments of measures that have been widely used to quantify state capacity in the social sciences. We have found that the convergence of the seven analysed measures of state capacity is high. All measures are positively correlated among each other, and the correlations are strong and consistent over time. The multivariate association and unidimensionality of the measures are confirmed by principal component analysis. Qualitatively each measure focuses on slightly different aspects of state capacity but statistically they seem to capture one single concept.

Despite a strong association among measures of state capacity, the set of replicated regression models has revealed that the interchangeability among these measures is weak and that the choice of the measure can have a substantial impact on the interpretation of the findings. In the most worrisome cases, we have found that two measures can even

lead to completely opposing interpretations. Scholars working on state capacity must be aware that results on the topic are unlikely to be generalisable across common measures, unless explicit evidence that results are robust to multiple measures is provided. Furthermore, the results of the replications cast some doubts on the extant knowledge about the relationship between democracy and state capacity. How solid is our knowledge on the topic, if the replicated and re-analysed models are highly sensitive to the chosen measure? Further studies on the state-democracy nexus are needed.

To get a clearer view of the somewhat contradictory findings about strongly correlated but weakly interchangeable measures, we shifted the level of analysis to the country-level and found striking disagreements in individual country scores among measures. By creating an indicator of multivariate rating discrepancy, we determined the countries that the seven measurements of state capacity most agree or disagree upon. The countries with the highest rating discrepancy were further analysed against each measure. The findings suggest that large rating discrepancies can be generally attributed to at least two factors: differences in the content of the measures and the systematic disagreement at intermediate levels of state capacity.

Despite high convergent validity, our findings have shown that the measures are not equivalent. This means that even if two measures of the same concept are highly correlated, we should not expect them to be the same. In terms of basic statistical properties, SFI is by far the most problematic measure. First, it is not able to distinguish among countries with a high level of state capacity. Second, it is not able to capture any possible increase in the level of state capacity in many of these countries. In terms of content, even if my objective was to explore the content more than evaluate it, there are no doubts that FSI covers such a broad understanding of state capacity that it seriously undermines its utility in causal research. For instance, in FSI, also levels of democracy

and socio-economic development are reflected in the final state capacity score. To avoid redundancy, it should not be used to explain or to be explained by any of the factors that are included in the index in the first place.

In general, SFI and HSI rate countries comparatively high and CPI and VDEM rate countries comparatively low. As we have seen, in some cases, the disagreements in country scores are overwhelming. Here, I have focused on differences in scores in 2009. Yet, an analysis of the country scores over time could provide interesting additional information on the disagreements among the measures. For instance, between 1999 and 2013[8], under Chávez's rule, Venezuela's level of state capacity increases slightly with HSI and FSI, stays more or less the same with SFI, and decreases with WGI, CPI, VDEM, and QOG. In particular, with QOG the decrease is substantial. If one's theoretical approach would presume that the level of state capacity in Venezuela decreased under Chávez's rule, then probably HSI, FSI, and SFI should not be used, since these three measures claim empirically the opposite.

Likewise, an analysis of country-specific rankings could provide further details on the disagreements among the measures. For instance, in 2009, Egypt ranks 57th in the world (more or less like China and Russia) with HSI but 169th in the world (worse than countries such as Somalia and Madagascar) with VDEM. Scholars must be aware about these differences and the practical consequences of choosing one measure instead of another. In general, it is important that the selected instrument matches the working definition of state capacity and that researchers know what the chosen instrument is actually measuring.

---

[8] Some of the measures are available only for some of the years between 1999 and 2013.

Last, the findings of this study provide two methods-related implications. First, strong correlations should not be taken as a sufficient proof of equivalency or high interchangeability between measures. Even if it is a common practice to assess the validity of measures with correlations, the unit-level analysis of individual observations has shown that highly correlated measures can have substantial discrepancies. Thus, highly correlated variables do not necessarily portray the same picture. Second, the findings remind the importance of replication studies in our field of research. Replications are fundamental to evaluate the robustness of previous findings and foster our understanding on any given topic.

Chapter III

Disaggregating State Capacity

On the Relationship between Public Officials and Civic Participation

Introduction

The previous chapter, in which we comprehensively discussed some of the advantages and disadvantages of frequently used measures of state capacity, has shown us that quantifying the multidimensional concept of state capacity is not a straightforward task by any means. A perfect measure of state capacity may never exist, but we do have now a more detailed understanding of the surveyed measures. We have seen that different measures of state capacity capture different dimensions of state capacity, that some measures capture even attributes that should probably not be included in measures of state capacity, that measures disagree on the scores of many countries, and that these disagreements, ultimately, have a significant impact on research results, undermining the accumulation of knowledge. It is hard to prove that a particular measure is better than the others in absolute terms, but there is no doubt that one measure can be more suitable than the others in a given context.

Since it has been argued that more specific definitions can facilitate causal research (e.g., Adcock and Collier 2001), if we want to engage in more meaningful cross-national research on the topic, one good option is to start disaggregating the concept of state capacity, both in terms of theory and in terms of measurement. Therefore, in the chapter at hand, I propose to shift the analysis towards more specific aspects of state capacity. As to theory, this means focusing on narrower attributes of state capacity and fitting a

theoretical framework to the selected attribute(s). As to tools of measurement, it means quantifying precise characteristics of state capacity. Specific measures enable to comprehend without hesitation what is being discussed and found, reducing the margins for conceptual disagreement and ensuring high content validity. To give a practical example, if we think about the relationship between democracy and state capacity that we examined in the previous chapter, we could start thinking about the relationship between some specific aspects of these two complex concepts and select measures that match closely with the chosen aspects.

In this chapter, by digging deeper into two core attributes of state capacity and democracy – namely the impartiality of public officials and civic participation – I show persuasively that disaggregating the complex concept of state capacity is a viable solution to push forward the research agenda on the topic. First, a disaggregation simply helps to understand what happens underneath the broader conceptual level. Second, as already mentioned, quantifying a specific attribute of state capacity reduces drastically the ambiguity regarding what is being measured, which is a common problem of many frequently used measures of state capacity, as discussed in the previous chapter. Third, an analysis of the relationship between impartial public officials and civic participation provides an important contribution to the so called "sequencing" literature on the relationship between state capacity and democracy by generating novel knowledge on the driving forces of the mechanisms of the sequence. Hence, before the empirical part of this chapter, the literature related to the sequencing debate is briefly reviewed and a theoretical framework related to these two specific aspects within the sequencing debate is developed.

In terms of measurement, the downside of the disaggregation proposed in this chapter is, obviously, that it does not provide much help to researchers who would like to

analyse state capacity more broadly as a multidimensional concept. Therefore, since the need of multidimensional tools of measurement has not disappeared nor will disappear, in the next chapter (Chapter IV), I will take action to develop a novel multidimensional measure of state capacity.

On the Sequencing Debate

In the last decade or so, a sizable body of empirical literature addressing the so-called "sequencing" debate on the relationship between democracy and state capacity has emerged in the field of comparative social sciences. Despite many large-N cross-national studies on the topic (e.g., Bratton and Chang 2006; Bäck and Hadenius 2008; Charron and Lapuente 2010; Fortin 2012; Andersen et al. 2014; Carbone and Memoli 2015; Grassi and Memoli 2016a, 2016b; D'Arcy and Nistotskaya 2017; Wang and Xu 2018), whether capable state institutions or democratic political institutions need to be created first, to sustain the subsequent development of the other, remains contested both empirically and theoretically[9].

Traditionally, the state-first argument has dominated the research agenda on the state-democracy nexus. The state-first argument, rooted in Huntington's (1968) ideas on effective governments that must control the governed before controlling themselves, has been advocated in several studies (e.g., Linz and Stepan 1996; Fukuyama 2005; D'Arcy and Nistotskaya 2017). While some kind of state institutions need to be put in place before democratisation by definition, because "democracy is a form of governance of a state" (Linz and Stepan 1996: 7), and some kind of political order needs to exist prior to democracy because political participation would be impossible in complete insecurity

---

[9] See Mazzuca and Munck (2014) for a comprehensive review of both arguments.

(Fukuyama 2014), the state-first theories fail to explain important real-world evidence. There are many recent examples in which countries with capable state institutions have not become democracies (e.g., Belarus, Singapore, United Arab Emirates) or in which democratic political institutions have been established before state capacity (e.g., India, Estonia, East-Timor). Hence, more recently, another body of literature has emerged to analyse the opposite causal direction of the sequence: the democracy-first argument, which is arguably "a better guide to action than the state first thesis" (Mazzucca and Munck 2014: 1223).

Theoretically, the democracy-first argument remains much less developed than its state-first counterpart, but consists in at least three state-consolidating mechanisms (Carbone 2015): the increase in political order through democratic participation, the increase in impartial procedures through political competition and accountability, and the increase in administrative capacity and territorial presence through the processes of organising and carrying out elections. Empirically, the democracy-first argument seems to be supported at least in part. Overall, democracy fosters state-building (Wang and Xu 2018), particularly in countries that are economically wealthy (Charron and Lapuente 2010) and already well set on the path to full democratisation (Bäck and Hadenius 2008; Carbone and Memoli 2015). Nonetheless, real-world evidence does not fully support the democracy-first theories, either. As put by Fukuyama (2014), there are many autocratic governments in history that have been able to create well-functioning states (e.g., Prussia, Japan) and many democratic ones that have not been able to build strong states with effective bureaucracies (e.g., India, Mexico).

Theoretical arguments of both causal directions are appealing but fail to explain why there have been so many exceptions throughout the history. Before proceeding to a further analysis of the relationship between state capacity and democracy, by disaggregating the

concepts and going into the specific, it is essential to understand with clarity the difference between state capacity and democracy. While both are crucial for development (Norris 2012; Cronert and Hadenius 2020), the two phenomena are distinct and do not go necessarily hand in hand (even if many of us would hope so). As we already know by now, state capacity has to do with the state's ability to implement its intended policies, regardless of the type political regime (Fukuyama 2013). On the contrary, democracy is a form of governance and more precisely a political system that is "completely or almost completely responsive to all its citizens" (Dahl 1971: 2). The core dimensions of the state, such as security and the administrative apparatus are completely different from the core dimensions of democracy, such as political participation and public contestation.

The overly broad concept of democratic governance, which has been advocated especially by international organisations and the international development policy agenda, has blurred the differences between the two concepts. To generate more knowledge on the nexus between state capacity and democracy, upon which we have surprisingly little conclusive information about, researchers are forced to distinguish one from the other. To generate more specific knowledge on the state-democracy nexus, researchers might be obliged to start disentangling the various sub-dimensions of these two complex concepts.

A Disaggregated Sequencing: On Impartial Bureaucrats and Civic Participation

The findings of the previous chapter suggest that the literature on the relationship between democracy and state capacity might have remained inconclusive because different studies have actually measured different things in the name of state capacity. Since we know now that different operationalisations of state capacity do not necessarily measure the same – in terms of qualitative content as well as numerical information – our cross-national

quantitative knowledge on the topic is very much dependent on the chosen operationalisation. Choosing a measure that closely represent the selected approach becomes fundamental.

Only very recently, some pioneering studies (e.g., Andersen and Doucette 2020) have called for a disaggregated approach in conducting research on the state-democracy nexus, in terms of theories, concepts, and measurement. This chapter can be seen also as a timely response to these recent appeals. By building a theoretical framework upon Putnam's (1993) and Banfield's (1958) studies on the civic community and institutions in Italy, I examine the relationship between two specific attributes of state capacity and democracy: impartiality of the public administration and civic participation.

Participation is generally considered as one of the two main dimensions of democracy (Dahl 1971). While it is sometimes conceived merely as political participation in elections in minimalist definitions of democracy, in a good democracy citizens do not participate only through elections but also "in the life of political parties and civil society organisations, in the discussion of public policy issues, in communicating with and demanding accountability from elected representatives, in monitoring official conduct, and in direct engagement with public issues at the local level" (Diamond and Morlino 2004: 23–24).

Hence, in a broader definition of democracy, in which the principal actors are not only formal political institutions, but also single citizens, territorial communities, and various forms of civic associations (Morlino 2012), civic participation becomes one of its core attributes. Furthermore, obviously, a vibrant associational life requires at least freedom of association, which is one of Dahl's (1971) institutional guarantees of a democratic society. A Weberian bureaucratic organisational structure, instead, is one of the main pillars of modern capable states (e.g., Hendrix 2010; Norris 2012), and

impartiality (e.g., Norris 2012; Cornell, Knutsen, and Teorell 2020) is one of the central principles of a well-functioning Weberian bureaucracy.

According to Putnam (1993) differences in civic participation are crucial in explaining well-functioning state institutions, because when a society has a vibrant civic community, its citizens see the public field as much more than only a playground where to achieve personal interests. Similar ideas had been put forward earlier by Banfield (1958) and his theory of "amoral familism" in a case study in a small village in Southern Italy. According to Banfield (1958), in a society of amoral familists, citizens do not take interest in public issues and official positions are used by their possessors to gain private advantage. Citizens of such a society are unable to act in the public interest, because civic participation and cooperation among citizens is inexistent (Banfield 1958). Both Putnam's and Banfield's theories suggest that when a society has a vibrant civic community, it is unlikely that public officials engage in partial actions and pursue private gains. Citizens in societies with a high level of civic participation "demand more effective public service, and they are prepared to act collectively to achieve their shared goals" (Putnam 1993: 182).

A broad implication of both studies is that a relationship between democracy and state capacity exists. More specifically, however, both studies claim that a strong civic society leads to more impartial and less disinterested behaviour by public officials. On the contrary, the opposite happens when a vibrant civic community is lacking. In a society where citizens have no civic virtues, both ordinary citizens and public officials see the public sphere only as an arena for pursuing and achieving private interests. Nevertheless, the findings of both studies are confined to the Italian case: it remains to be seen whether the relationship implied by both Putnam and Banfield holds not only in Italy, but also in a broader cross-national setting.

At least according to some more recent theoretical propositions, we can suspect that a strong civil society is likely to play a key role in affecting the behaviour of public servants also outside Italy. Historically, civil society has been a fundamental actor in the fight against the unequal treatment of citizens by the state (Mungiu-Pippidi 2006). According to Carothers (2007), some specific aspects of democracy can contribute to the development of an effective and impartial public administration. In particular, "creating space for independent civil society permits advocacy groups to monitor and critique state performance and work together with the state to offer new policy ideas" (Carothers 2007: 20). Nevertheless, it is good to remember that also opposing theories on the association between civic communities and impartial bureaucrats exist: for instance, an overly powerful civic society could also generate bureaucrats whose interests coincide more with civic society organisations than the state, and thus, decrease impartiality and effective implementation of policies (Migdal 1988).

Next, I discuss the chosen data and my empirical strategy to test the validity of the hypothesised relationship. As we shall see, my findings provide robust evidence of a positive effect of civic participation on the impartiality of public officials, increasing our understanding about the relationship between these two core features of democracy and state capacity, and as a consequence, generating new knowledge also on the broader relationship between democracy and the state. Furthermore, the findings support persuasively the argument that the research agenda on state capacity (and democracy) can be pushed forward by disaggregating the concept(s).

Data and Empirical Strategy

We have seen in Chapter II that researchers do not always rely upon measures that represent well their concepts and theories. Anyhow, we have seen as well that it is crucial

that the chosen measure matches the chosen definition as accurately as possible. Hence, to quantify specific aspects of democracy and state capacity, we should use specific measures that represent closely the concepts we are interested in. The key variables of my analysis are taken from the Varieties of Democracy (V-Dem) dataset (Coppedge et al. 2020a), which provides two indicators that capture closely civic participation and one that captures the impartiality of public officials.

Civic participation is operationalised with the *Core civil society* index (CCSI), which aggregates three expert assessments on the topic. These assessments quantify (1) the amount of control of the government on civil society organisations, (2) the amount of repression pursued by the government to repress civil society organisations, and (3) the amount popular involvement in civil society organisations. Robustness tests are provided with the *Civil society participation* index (CSPART), which aggregates four expert assessments on as many questions on the topic: (1) How large is the involvement of people in civil society organisations? (2) Are major civil society organisations routinely consulted by policymakers on policies relevant to their members? (3) Are women prevented from participating in civil society organisations? (4) How centralised is legislative candidate selection within parties? Both civic participation indices run from 0 (low) to 1 (high) on an interval scale.

The impartiality of public officials, instead, is operationalised with the *Rigorous and impartial public administration* indicator. We know well this measure from the previous chapter. Just to recall, the indicator is an expert assessment that answers to the question "are public officials rigorous and impartial in the performance of their duties?" (Coppedge et al. 2020a: 164). A robustness check is performed with another specific indicator measuring the impartiality of public officials. This indicator is taken from the QoG Expert Survey II dataset (Dahlström et al. 2015) and measures "to what extent

government institutions exercise their power impartially". Since the QoG Expert Survey measure is not available over time, however, it can be used only in a cross-section setting. Both measures of impartiality run on an interval scale. V-Dem's measure follows approximately a normal z-score scale with a mean of 0 for all available country-years in V-Dem's dataset (Coppedge et al. 2020b). QoG Expert Survey's measure ranges from 1 (low) to 7 (high).

As to the control variables, I consider several frequently used predictors of well-functioning bureaucracies. In the baseline models I control for *GDP/capita*, for the simple reason that "rich nations have better governments than poor ones" (La Porta et al. 1999: 266). More complex specifications include controls for *education* (e.g., Charron and Lapuente 2010; Wang and Xu 2018), *natural resources* (e.g., Charron and Lapuente 2011; Ricciuti, Savoia, and Sen 2019), *total population* (e.g., Stockemer 2014; Wang and Xu 2018), and *ethnic fractionalisation* (e.g., Charron and Lapuente 2010; Carbone and Memoli 2015). Furthermore, to ensure that other democratic aspects are accounted for, I add controls for regime type according to V-Dem's *Regimes of the World* fourfold typology. Full models control for all the above-mentioned factors. The sources of the selected data are presented in Table B1 (Appendix B). Summary statistics of the longitudinal data are reported in Table B2 (Appendix B). Summary statistics of the data used in cross-section robustness tests are reported in Table B3 (Appendix B).

If civic participation affects the impartiality of public officials, it is reasonable to assume that its impact may occur not only in the short run but also in the long run. As a consequence, we should be interested not only in assessing the short run impact of civic participation but also in understanding whether its effect on impartial public officials persists for a longer period of time. As long as our variables are stationary, we can estimate both short-term and long-term relationships with simple adjustments to "static"

specifications in ordinary least squares (OLS). For instance, one straightforward way to take into account dynamics is to add one or more lagged dependent variables on the right-hand side of the regression equation (e.g., Keele and Kelly 2006; De Boef and Keele 2008; Beck and Katz 2011). Before further proceeding with the statistical analysis, thus, I explore the stationary properties of my key variables of interest.

Unit root tests (Table B4, Appendix B) indicate that I can reject the null hypothesis that all panels contain unit roots. Put simply, I do not find statistically significant evidence of nonstationarity in my key variables. This means also that I can confidently estimate the relationship between civic participation and impartial public officials in a dynamic panel regression setting. Besides adding dynamics into the models, lagged dependent variables purge also the serial correlation in the error terms.

All time-variant independent variables are lagged by one year and contemporaneous effects of independent variables are restricted to zero. It seems theoretically implausible to believe that variation in civic participation would affect the impartiality of public officials instantaneously. On the contrary, I assume that the short-term relationship occurs with a one-year lag. Furthermore, lagging time-variant independent variables helps alleviating possible concerns of reverse causality, because the impartiality of public officials at time t is unlikely to affect any of the independent variables at time $t–1$.

The main regression models include also two-way fixed effects[10] to account for time-invariant factors and common time trends. Although it is well known that combining lagged dependent variables with fixed effects causes biased parameter estimates (Nickell 1981), such bias is negligible when the time dimension is 20 or more (Beck and Katz 2011). Since the average time dimension in my main models ranges between 41 and 61,

---

[10] Two-way fixed effects stand for both unit fixed effects and time fixed effects.

Nickell bias should be sufficiently small to not to worry about. Leaving out unit fixed effects, instead, would lead to severe omitted variable bias. Since the theoretical argument is that variation in civic participation affects variation in the impartiality of public officials within a country, the use of country-level fixed effect models seems to be not only statistically but also theoretically justified. The selected measure of ethnic fractionalisation does not vary over time, so it cannot be included (nor there is any need to include it) in dynamic panel models with fixed effects, but it is included in cross-section robustness tests.

In addition to the main set of models, I perform a battery of robustness tests. First, I test the robustness of the results to an alternative measure of civic participation. Second, I restrict the sample of the models to a common sample of country-years. Third, I test whether the results vary across macro-regions in the world. The regional classification is taken from Coppedge et al. (2020a). Fourth, I re-estimate the full models with several alternative estimators. Fifth, I test the robustness of the results to different lags of y. Sixth, I regress an alternative measure of impartial public officials on both measures of civic participation. All models are estimated with country-clustered heteroskedasticity consistent standard errors.

Besides testing my main hypothesis, rooted in culturalist approaches on the issue, I test also the empirical validity of two well-known institutionalist theories that have been proposed in the literature on democracy and state capacity: the "curvilinear relationship" (Bäck and Hadenius 2008) and "conditional on modernisation" (Charron and Lapuente 2010) theories. These studies were already replicated and re-analysed with common measures of state capacity in Chapter II. Here, I test their theoretical validity in our disaggregated approach to the topic.

Bäck and Hadenius' (2008: 13) argue that civic participation, among other factors, facilitates "steering and control from below", and thus, enhances administrative capacity through the control by citizens. However, this happens only after a certain level of democratisation has been reached. When political institutions are highly autocratic, they argue that an increase in such an influence from below actually weakens the state. Charron and Lapuente (2010), instead, posit that the relationship between democracy and state capacity is conditional on the level of modernisation of a society. They argue that citizens in less modern societies undervalue the impartiality of public officials, and as a consequence, are less ready to build efficient and impartial state institutions.

Results and Discussion

A preliminary descriptive analysis of the key variables gives support to my main hypothesis and encourages its further analysis. A scatter plot (Figure 3.1) of the country-specific average values from 1945 to 2017 confirms that on average the relationship between civic participation and impartial bureaucrats is strong, positive, and linear (Pearson's R = 0.71). Civic participation, alone, explains as much as 50% of the variation in impartial public officials. Yet, it is interesting to notice that in some countries the relationship between civic participation and impartial public officials seems to be at odds with the general trend. For instance, countries such as Nigeria, Guatemala, and Bosnia-Herzegovina have considerably less impartial bureaucrats than one would expect from their level of civic participation, whereas the opposite is true in Bhutan, Singapore, and South Yemen. Despite preliminary support for the hypothesised relationship, since we are interested primarily in within-country over time variation and since potential confounders should be controlled for, a comprehensive regression analysis could paint a completely different picture on the matter.

Figure 3.1. Civic participation and impartial bureaucrats (1945-2017 average)
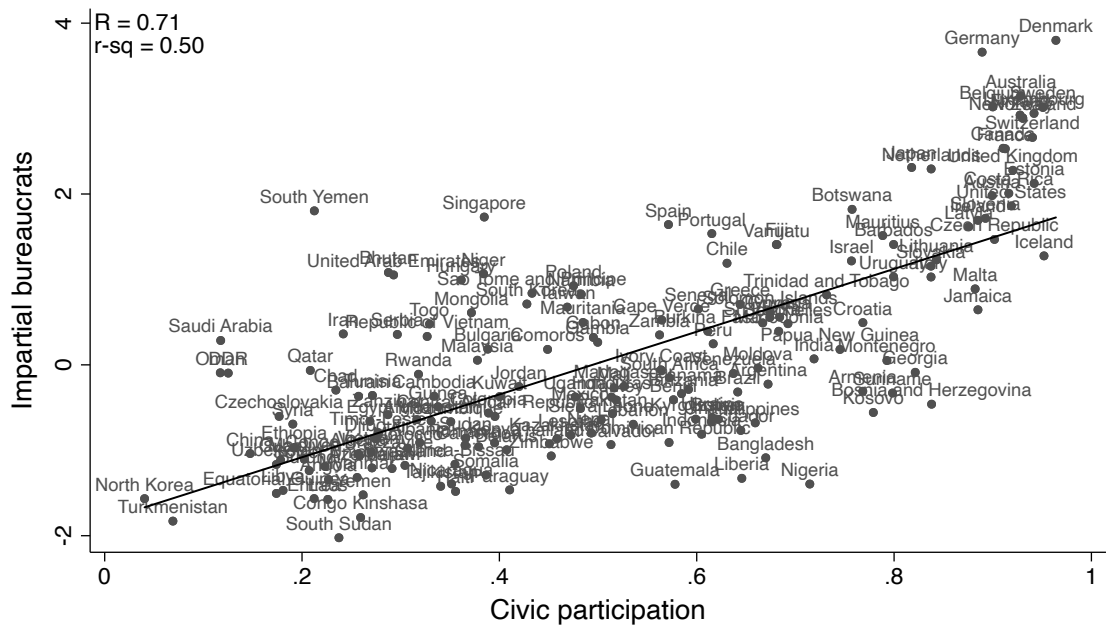
Note: Civic participation is measured with CCSI.

Table 3.1 shows a summary of the results of the main set of dynamic panel regressions. Model 1 includes two-way fixed effects, a lagged dependent variable, and a control for GDP/capita. In Model 2, additional controls for natural resource rents, total population, and years of education are included in the specification. To control for differences in formal political institutions, Model 3 adds regime type dummies to the baseline specification. Model 4 adds to the baseline specification both regime type dummies and the aforementioned additional controls. Model 5 tests the validity of the theory of a curvilinear association between civic participation and impartial public officials, whereas Model 6 assesses whether the relationship between civic participation and impartial public officials is contingent on the level of economic wealth.

In all the first four models the association between civic participation and the impartiality of public officials is positive and significant, supporting the main hypothesis of my study. According to these models, an increase in civic participation is related to an

Table 3.1. Civic participation and impartial public officials: main regression results

	Dependent variable: Impartial public officials $_t$					
	Baseline model	Additional controls	Regime types	Full model	Squared particip.	Depends on wealth
	(1)	(2)	(3)	(4)	(5)	(6)
Civic participation $_{t-1}$	0.068***	0.133***	0.092***	0.154***	0.046	0.115
	(0.020)	(0.036)	(0.028)	(0.044)	(0.065)	(0.097)
Ln(GDP/capita) $_{t-1}$	0.005	0.024	0.005	0.024	0.004	0.008
	(0.008)	(0.017)	(0.005)	(0.017)	(0.008)	(0.011)
Civic participation^2 $_{t-1}$					0.023	
					(0.062)	
Civic participation × GDP/capita $_{t-1}$						-0.006
						(0.011)
Ln(Natural resources) $_{t-1}$		0.003		0.004		
		(0.005)		(0.006)		
Ln(Population) $_{t-1}$		-0.006		-0.004		
		(0.034)		(0.035)		
Years of education $_{t-1}$		-0.016		-0.016		
		(0.010)		(0.010)		
Impartial pub. officials $_{t-1}$	0.924***	0.898***	0.925***	0.899***	0.924***	0.925***
	(0.005)	(0.008)	(0.005)	(0.008)	(0.005)	(0.005)
Long-run multipliers						
Civic participation $_{t-1}$	0.898***	1.308***	1.216***	1.532***		
	(0.231)	(0.302)	(0.344)	(0.395)		
Civic participation^2 $_{t-1}$					0.306	
					(0.817)	
Civic participation × GDP/capita $_{t-1}$						-0.079
						(0.151)
Within R^2	0.89	0.86	0.89	0.86	0.89	0.89
N	9935	5447	9913	5433	9935	9935
Sample	1946-2017	1971-2017	1946-2017	1971-2017	1946-2017	1946-2017
Countries	162	133	162	133	162	162
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Regime type dummies $_{t-1}$	No	No	Yes	Yes	No	No

Note: Robust standard errors in parentheses; *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. Constant coefficient measured but not reported. Sample of years refers to dependent variable. Civic participation is measured with CCSI.

increase in the impartiality of public officials both in the short and in the long run. The result is statistically significant at the highest level across models. Even if the short-term point estimates of civic participation are relatively small in magnitude, varying from a minimum of 0.068 in model 1 to a maximum of 0.154 in model 4, they are all statistically significant from zero at the highest level.

As already mentioned, these dynamic specifications allow us to estimate not only the abovementioned short-term effects of civic participation on impartial bureaucrats, but also the long-term effects. Unsurprisingly, the long-run effect ($\frac{\beta_{x_{t-1}}}{1-\beta_{y_{t-1}}}$) of civic participation on impartial bureaucrats is larger than the short-run effect ($\beta_{x_{t-1}}$). In models with additional controls (2 and 4), the long-run effect of civic participation is nearly 10 times its immediate effect. In models with no additional controls (1 and 3), the long-run effect is around 13 times the immediate effect. In Model 4, a one unit increase in civic participation is associated with a 1.532 unit increase in impartial bureaucracy in the long run.

This long-term effect does not occur all at once, but instead, dissipates relatively slowly over time. In Model 4, the speed of adjustment ($1 - \beta_{y_{t-1}}$) of 0.101 indicates that in the case of a shock to civic participation in a particular year, 50% of the total long-term impact of civic participation on impartial public officials materialises in approximately seven years and 90% of the total long-term impact takes place in approximately 22 years. In Models 1, 2, and 3 one half of the total long-term impact on impartial public officials occurs in around nine, seven, and nine years, respectively. Since the estimated coefficients of the lagged dependent variables are below 1, we can be confident that the models are correctly specified.

Models 5 and 6 evaluate the empirical validity of the two briefly discussed alternative theories on the topic. In Model 5 I test for Bäck and Hadenius' (2008) hypothesis of a curvilinear relationship, and in Model 6 I test for Charron and Lapuente's (2010) hypothesis of an effect conditional on the level of economic wealth. The results of the regressions do not support these alternative hypotheses. Whether it is in the short-term or in the long-term, according to the estimates, neither the level of civic participation

nor the level of economic modernisation plays a key role in explaining the relationship between civic participation and impartial public officials.

To test the robustness of my findings, I run a series of alternative specifications and estimations. Dynamic panel regressions (Table 3.2) with an alternative measure of civic participation – V-Dem's civic participation index – do not alter the interpretation of the previous results. In fact, if anything, they reinforce the findings of the first set of regressions. Again, civic participation and impartial public officials are positively related to each other in the first four models, both in the short-run and in the long-run. On an average, higher levels of civic participation seem to be conducive to bureaucrats that act more impartially in their public duties, as suggested by Banfield (1958) and Putnam (1993) in their influential studies.

As before, in the first four models, short run slope coefficients of civic participation are relatively small in magnitude but statistically significant at conventional levels and range from a minimum of 0.054 in Model 1 to a maximum of 0.116 in Model 2. According to Model 4 – the full model – a one unit increase in civic participation increases the impartiality of public officials by 0.114 units in the short run. The cumulative long run effect of civic participation, instead, ranges from a minimum of 0.736 in model 1 to a maximum of 1.200 in Model 2. In Model 4 a one unit increase in civic participation increases the impartiality of public officials by 1.171 units in the long run. Considering a shift in civic participation in a particular year, 50% of its long run effect occurs in around seven years and 90% in around 23 years.

The alternative hypothesis of a curvilinear relationship between civic participation and impartial public officials is not supported by Model 5. Likewise, the alternative hypothesis that the linkage between civic participation and impartial public officials is conditional on economic development is not supported by Model 6. In both cases, the

Table 3.2. Civic participation and impartial public officials: robustness tests with alternative measure of civic participation

	Dependent variable: Impartial public officials $_t$					
	Baseline model	Additional controls	Regime types	Full model	Squared particip.	Depends on wealth
	(1)	(2)	(3)	(4)	(5)	(6)
Civic participation $_{t-1}$	0.054**	0.116***	0.063**	0.114**	0.078	0.123
	(0.023)	(0.039)	(0.029)	(0.045)	(0.072)	(0.109)
Ln(GDP/capita) $_{t-1}$	0.004	0.021	0.004	0.020	0.004	0.009
	(0.008)	(0.017)	(0.009)	(0.017)	(0.008)	(0.012)
Civic participation^2 $_{t-1}$					-0.027	
					(0.073)	
Civic participation × GDP/capita $_{t-1}$						-0.009
						(0.014)
Ln(Natural resources) $_{t-1}$		0.004		0.004		
		(0.005)		(0.005)		
Ln(Population) $_{t-1}$		-0.001		-0.001		
		(0.033)		(0.035)		
Years of education $_{t-1}$		-0.017*		-0.017		
		(0.010)		(0.010)		
Impartial pub. officials $_{t-1}$	0.927***	0.903***	0.927***	0.903***	0.927***	0. 928***
	(0.005)	(0.007)	(0.005)	(0.008)	(0.005)	(0.005)
Long-run multipliers						
Civic participation $_{t-1}$	0.736**	1.200***	0.866**	1.171***		
	(0.291)	(0.357)	(0.390)	(0.435)		
Civic participation^2 $_{t-1}$					-0.372	
					(0.998)	
Civic participation × GDP/capita $_{t-1}$						-0.121
						(0.187)
Within R^2	0.89	0.86	0.89	0.86	0.89	0.89
N	9935	5447	9913	5433	9935	9935
Sample	1946-2017	1971-2017	1946-2017	1971-2017	1946-2017	1946-2017
Countries	162	133	162	133	162	162
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Regime type dummies $_{t-1}$	No	No	Yes	Yes	No	No

Note: Robust standard errors in parentheses; *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. Constant coefficient measured but not reported. Sample of years refers to dependent variable. Civic participation is measured with CSPART.

estimated results regarding the key explanatory variables are not even closely significant at conventional levels. In the first two sets of regressions, I do not find robust evidence of a significant relationship between any of the covariates and impartial public officials.

Coefficient plots in Figures 3.2 and 3.3 provide a straightforward illustration of, respectively, the strength of the short-term and long-term relationships between civic

participation and impartial public officials. The slope coefficients of the first four models in Table 3.1 (left panel) and Table 3.2 (right panel), represented by the dots and their respective 95% confidence intervals represented by the spikes, are all completely above the horizontal line of zero, indicating that civic participation is positively related to impartial public officials. As to the short-term effects (Figure 3.2), Models 2 and 4 in both panels have sensibly higher point estimates compared to Models 1 and 3 but are also less precise. Similarly, as to the long-term effects (Figure 3.3), Models 2 and 4 in both panels have higher point estimates than Models 1 and 3, but only Model 1 is distinctly more precise than the other models, regardless of the chosen measure of civic participation.

Figure 3.2. Short-term effect of civic participation on impartial public officials
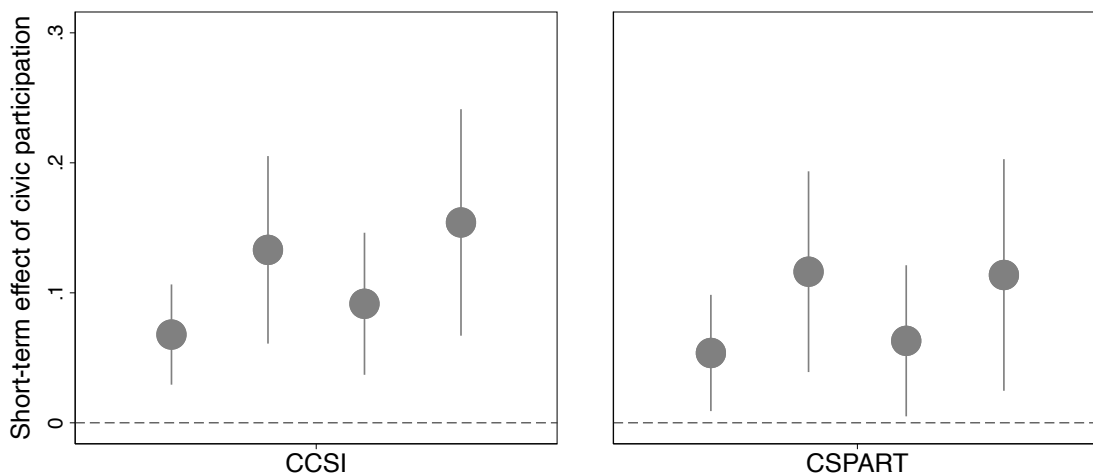
Note: Slope coefficients on the left panel refer to models 1, 2, 3, and 4 in Table 3.1 (main panel regressions) in their respective order. Slope coefficients on the right panel refer to models 1, 2, 3, and 4 in Table 3.2 (alternative panel regressions) in their respective order.

Since each model has its own set of controls, there is no doubt that one reason for different estimates across models is the specification. Nevertheless, due to data availability, there are also some differences in the samples of each model. To investigate whether different samples lead to different estimations, I run again the first four models of the first set of regressions with a common sample of country-years. The re-estimated

"common-sample" models (Table B5, Appendix B) show that differences in country-years play a certain role in differences across models in the first set of regressions. Once the estimates are based on the same sample of country-years, beta coefficients, standard errors, and confidence intervals become more convergent across models. This result holds both for short-term estimates and long-term estimates.

Figure 3.3. Long-term effect of civic participation on impartial public officials
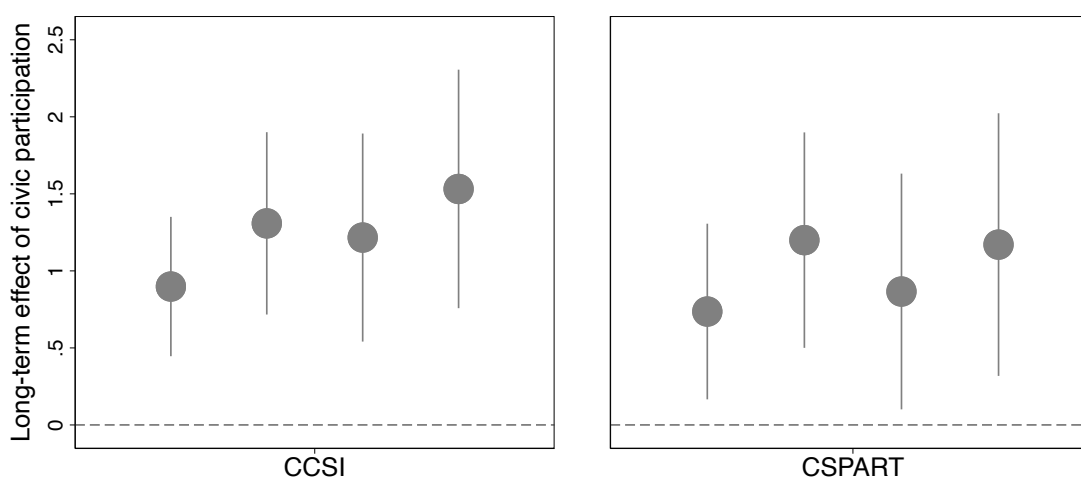
Note: Slope coefficients on the left panel refer to models 1, 2, 3, and 4 in Table 3.1 (main panel regressions) in their respective order. Slope coefficients on the right panel refer to models 1, 2, 3, and 4 in Table 3.2 (alternative panel regressions) in their respective order.

As another robustness check, I divide the common sample by macro-regions and re-estimate the full models separately for each macro-region. Even if we are looking at within-country variations, the average relationship between civic participation and impartial public officials could still differ from one macro-region to another. Interestingly, as reported in Table 3.3, this is precisely the case. On average, there seems to be a positive relationship between civic participation and impartial public officials in every region in the world. Nevertheless, the result is statistically significant both in the short run and the long run only in Western countries (Model 1), Sub-Saharan Africa (Model 2), Latin America (Model 3), and Asia (Model 5).

Table 3.3. Civic participation and impartial public officials by macro-region: re-estimations of the full models in Table 3.1

	Dependent variable: Impartial public officials t					
	Western countries	Sub-Saharan Africa	Latin America	Former Soviet Union	Asia	Middle East & N. Africa
	(1)	(2)	(3)	(4)	(5)	(6)
Civic participation $t-1$	0.879**	0.141*	0.283**	0.054	0.346***	0.075
	(0.379)	(0.077)	(0.114)	(0.176)	(0.112)	(0.143)
Ln(GDP/capita) $t-1$	-0.023	-0.011	0.021	0.074*	0.165**	0.068
	(0.037)	(0.029)	(0.039)	(0.041)	(0.061)	(0.070)
Ln(Natural resources) $t-1$	0.006	-0.001	0.006	0.029	-0.029	0.018
	(0.006)	(0.018)	(0.015)	(0.023)	(0.022)	(0.015)
Ln(Population) $t-1$	0.088	-0.080	-0.120	-0.117	0.128	0.021
	(0.111)	(0.109)	(0.162)	(0.255)	(0.157)	(0.167)
Years of education $t-1$	-0.014	-0.012	-0.002	-0.120*	-0.071*	0.008
	(0.014)	(0.020)	(0.033)	(0.068)	(0.034)	(0.052)
Impartial pub. officials $t-1$	0.853***	0.919***	0.872***	0.808***	0.842***	0.864***
	(0.030)	(0.011)	(0.021)	(0.053)	(0.036)	(0.030)
Long-run multiplier						
Civic participation $t-1$	5.988**	1.738**	2.210***	0.281	2.193***	0.550
	(2.701)	(0.840)	(0.677)	(0.877)	(0.601)	(0.965)
R^2	0.89	0.87	0.90	0.83	0.85	0.78
N	978	1642	1074	499	692	548
Sample	1971-2017	1971-2017	1971-2017	1971-2017	1971-2017	1971-2017
Countries	22	38	23	20	17	13
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Regime type dummies $t-1$	No	No	Yes	Yes	No	No

Note: Robust standard errors in parentheses; *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. Constant coefficient measured but not reported. Sample of years refers to dependent variable. Civic participation is measured with CCSI.

In Western countries the relationship is particularly strong. In the short run, a one unit increase in civic participation is associated with a 0.879 unit increase in impartial public officials, whereas in the long run, a one unit increase in civic participation increases the impartiality of public officials by 5.988 units. Since the dependent variable ranges approximately from –5 to 5, the magnitude of the association is indeed considerable. In Former Soviet Union countries and in Middle Eastern and North African countries, there seems to be no statistically significant association between civic participation and impartial public officials.

As a further robustness test, to ensure that my findings are not conditional on the chosen method, I run a set of full models with alternative estimators (Table 3.4). In Model 1, I run a "static" model without a lagged y on the right-hand side of the regression equation. In Model 2, I run a pooled OLS model where both the intercept and the slope are assumed to be the same for all countries. In Model 3, I run a generalised least squares random intercepts model, where country-specific effects are not estimated but taken into account by estimating the variance explained by the country-specific effects. In Model 4, I run a random intercepts and random slopes maximum likelihood model where not only the intercept but also the slopes are allowed to differ across countries. In Model 5, I run the full specification in first differences to sweep out unit-level fixed effects and eliminate all long-term dynamics.

Table 3.4. Civic participation and impartial public officials: alternative estimators

|	*Dependent variable: Impartial public officials $_t$*				
	Static FE	Pooled OLS	Random intercepts	Random slopes	First differences
	(1)	(2)	(3)	(4)	(5)
Civic participation $_{t-1}$	1.703***	0.107***	0.107***	0.116***	0.485***
	(0.244)	(0.027)	(0.027)	(0.028)	(0.112)
Ln(GDP/capita) $_{t-1}$	0.280***	0.011	0.011	0.013*	0.025
	(0.102)	(0.008)	(0.008)	(0.008)	(0.082)
Ln(Natural resources) $_{t-1}$	-0.020	-0.002	-0.002	-0.002	0.008
	(0.032)	(0.002)	(0.002)	(0.002)	(0.008)
Ln(Population) $_{t-1}$	-0.547**	-0.005**	-0.005**	-0.005**	-0.097
	(0.238)	(0.002)	(0.002)	(0.002)	(0.223)
Years of education $_{t-1}$	0.064	0.003	0.003	0.003	0.042
	(0.064)	(0.002)	(0.002)	(0.002)	(0.073)
Impartial pub. officials $_{t-1}$		0.961***	0.961***	0.955***	0.002
		(0.005)	(0.005)	(0.006)	(0.018)
N	5433	5433	5433	5433	5293
Sample	1971-2017	1971-2017	1971-2017	1971-2017	1972-2017
Countries	133	133	133	133	133
Country FE	Yes	No	No	No	No
Year FE	Yes	Yes	Yes	Yes	Yes
Regime type dummies $_{t-1}$	Yes	Yes	Yes	Yes	Yes

Note: Robust standard errors in parentheses; *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. Constant coefficient measured but not reported. All variables in Model 5 are first differenced. Sample of years refers to dependent variable. Civic participation is measured with CCSI.

All five alternative estimators confirm that there is a strong and statistically significant positive relationship between civic participation and impartial public officials. Since in Model 5 I difference out both the dependent variable and the full set of independent variables, it is important to recall that with first differenced equations, the estimated results are based on variation in changes on both sides of the regression equation. This means that the interpretation of Model 5 is that changes in civic participation at time $t-1$ are positively related to changes in impartiality of public officials at time t.

Considering that a substantial amount of remaining residual serial correlation in an OLS model with a lagged dependent variable would lead to biased estimates and considering that "adding additional lags of the dependent variable helps to correct residual autocorrelation" (Wilkins 2017: 17), I test the robustness of my findings to specifications with additional lags of the dependent variable (Table 3.5). The estimates of the full model in the first set of regressions (Model 1) remain largely unaltered by the inclusion of one (Model 2), two (Model 3), or three (Model 4) additional lags of y. The average short-term effect remains virtually the same (0.154 with one lag; 0.150 with two or three lags; 0.151 with four lags) and the average long run effect, which is particularly sensitive to the amount of remaining residual serial correlation (Wilkins 2017), becomes slightly weaker with the inclusion of additional lags of y (1.532 with one lag; 1.396 with two lags; 1.395 with three lags; 1.392 with four lags), yet similar enough to support more than convincingly the previous findings of this study. The robustness of the results to additional lags of the dependent variable reinforces my findings: the relationship between civic participation and impartial public officials is strong both in the short-run and in the long-run.

Table 3.5. Civic participation and impartial public officials: different lags of y

	Dependent variable (DV): Impartial public officials $_t$			
	1 lag of DV	2 lags of DV	3 lags of DV	4 lags of DV
	(1)	(2)	(3)	(4)
Civic participation $_{t-1}$	0.154***	0.150***	0.150***	0.151***
	(0.044)	(0.043)	(0.043)	(0.043)
Impartial pub. officials $_{t-1}$	0.899***	0.969***	0.969***	0.969***
	(0.008)	(0.019)	(0.018)	(0.019)
Impartial pub. officials $_{t-2}$		-0.077***	-0.077***	-0.077***
		(0.018)	(0.021)	(0.021)
Impartial pub. officials $_{t-3}$			0.0001	-0.005
			(0.013)	(0.023)
Impartial pub. officials $_{t-4}$				-0.005
				(0.018)
Long-run multipliers				
Civic participation $_{t-1}$	1.532***	1.395***	1.396***	1.392***
	(0.395)	(0.394)	(0.365)	(0.369)
R^2	0.86	0.86	0.86	0.86
N	5433	5432	5431	5421
Sample	1971-2017	1971-2017	1971-2017	1971-2017
Countries	133	133	133	133
Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Note: Robust standard errors in parentheses; *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. All models include full set of control variables. Constant coefficient measured but not reported. Sample of years refers to dependent variable. Civic participation is measured with CCSI.

So far, to quantify impartial public officials, I have used in throughout the models V-Dem's indicator of impartial public administration. Thus, finally, I test the validity of my findings with an alternative measure of impartial public officials. Table 3.6 reports the cross-section regression results with an alternative measure of impartial public officials from the QoG Expert Survey. In the first three models, civic participation is quantified with the core civic society index (CCSI), while in the last three models I use the already discussed alternative measure of civic participation (CSPART). The hypothesised positive effect of civic participation on the impartiality of public officials receives strong support in all the cross-section models.

In each of the six models, the linear association between civic participation on impartial public officials is positive and significant at conventional levels. In Models 1-

3, a one unit increase civic participation is related to an increase between 1.055 and 1.141 units in impartial public officials. In Models 4-6, where civic participation is measured with an alternative index, a one unit increase in civic participation is related to an increase between 1.345 and 1.545 units in impartial public officials. In contrast with my first set of dynamic panel regressions, some of the control variables are significantly related to impartial public officials. In particular, in cross-section models, economic wealth and education are positively related to impartial bureaucrats.

Table 3.6. Civic participation and impartial public officials: cross-section robustness tests with alternative measure of impartiality

	Dependent variable: Impartial public officials $_t$					
	Baseline	Additional controls	Full model	Baseline	Additional controls	Full model
	(1)	(2)	(3)	(4)	(5)	(6)
Civic participation (CCSI) $_{t-1}$	1.055***	1.135**	1.141**			
	(0.368)	(0.442)	(0.563)			
Civic participation (CSPART) $_{t-1}$				1.395***	1.545***	1.345**
				(1.903)	(0.551)	(0.672)
Ln(GDP/capita) $_{t-1}$	0.742***	0.471**	0.333	0.757***	0.486***	0.341*
	(0.115)	(0.183)	(0.204)	(0.189)	(0.178)	(0.199)
Ln(Natural resources) $_{t-1}$		0.034	0.020		0.028	0.016
		(0.051)	(0.048)		(0.052)	(0.049)
Ln(Population) $_{t-1}$		-0.104	-0.074		-0.100	-0.073
		(0.066)	(0.067)		(0.065)	(0.066)
Years of education $_{t-1}$		0.156***	0.160***		0.154***	0.157***
		(0.059)	(0.059)		(0.058)	(0.058)
Ethnic fractionalisation $_{t-1}$		-0.361	-0.295		-0.476	-0.378
		(0.436)	(0.406)		(0.433)	(0.415)
R^2	0.65	0.69	0.72	0.65	0.70	0.72
N	110	99	99	110	99	99
Region dummies	Yes	Yes	Yes	Yes	Yes	Yes
Regime type dummies $_{t-1}$	No	No	Yes	No	No	Yes

Note: Robust standard errors in parentheses; *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. Constant coefficient measured but not reported. Dependent variable measured in 2013.

To sum up the results, my empirical analysis provides strong evidence of an overall positive effect of civic participation on impartial public officials. Banfield's (1958) and Putnam's (1993) discoveries in the context of Italy seem to hold also over time in global

sample of countries. The finding is robust to alternative measures of both civic participation and impartiality, and to a variety of specifications that control for a wide range of factors that are often used to predict bureaucratic quality. Interestingly, however, I find substantial differences among macro-regions. In fact, if different macro-regions of the world are analysed separately, the relationship between civic participation and impartial public officials persists in Western countries, Sub-Saharan Africa, Latin America, and Asia but does not hold in Former Soviet Union countries or in the Middle East and North Africa.

My findings provide also robust empirical evidence to reject the opposing theory according to which a powerful civic society induces public officials to act more partially, by strengthening more their ties with civic society organisations than their interest in the common good, as some scholars have posited. Even if relatively small in magnitude, the average short-term effect of civic participation on impartial bureaucrats is clearly different from zero. The average long-run impact, instead, is much larger and seems to dissipate relatively slowly over time.

Differences in the estimates between the panel models and the cross-section models can be explained at least by three factors. First of all, simply, panel models are based on data not only across countries, but also over years, whereas cross-section models take into account only variation between countries, but not over years. Second, my panel models take into account dynamics. Compared to a static model where time is treated at best as a nuisance, with dynamic models we are able to capture both the short run and long run effects, although it is good to remember that recent literature suggests being careful in interpreting the long run effects (e.g., Reed and Zhu 2017; Wilkins 2017). Third, the main panel regression estimates in this study are based on within-country over time variation. In these models we assess how an increase in civic participation in a given country over

time affects variation in the impartiality of public officials over time in that particular country. In the cross-section models, instead, the estimates provide information about how much more impartial countries with high civic participation are than countries with low civic participation in 2013, when civic participation is captured in 2012. Simply put, the two types of models do not predict the same and their interpretation is not the same.

The two alternative hypotheses advanced in well-known studies on the relationship between state capacity and democracy, do not find consistent support in neither my panel nor my cross-section models. A deeper investigation of the reasons for the divergent results is out of the intended scope of my study, but it is important to recall that besides some differences in covered years and the sample of countries, both Bäck and Hadenius (2008) and Charron and Lapuente (2010) analyse the association between democracy and state capacity as a whole, without going into specific characteristics of the two concepts. Thus, even if the results of my study provide robust evidence about a positive linear effect of civic participation on the impartiality of public officials, and does not find evidence of curvilinear relationship between the two nor of a relationship that depends on the level of economic wealth, it is important to stress that the empirical analysis in the study at hand does not assess the association between democracy and state capacity as a whole. An analysis of the dynamic relationship between democracy and state capacity as a whole could portray a different story.

A word of caution must be said also about the predicted causal direction. I have interpreted the above results causally, assuming civic participation to cause impartially behaving public officials. Nonetheless, an impartial bureaucracy might also facilitate the development of a dynamic and inclusive civil society. Possible problems of reverse causation should be mitigated by the use of lagged dependent variables in the models. As mentioned in the previous section, it is unlikely that present values of impartial

bureaucracy would affect past values of civic participation. That being said, even if my regression analysis strongly supports the hypothesised relationship, I acknowledge that the chosen estimation strategy cannot entirely exclude the possibility of reverse causality.

Conclusions

The primary aim of this chapter has been to explore if the research agenda on state capacity can be pushed forward by disaggregating the concept, both in terms of theory and in terms of measures. The findings suggest compellingly that a disaggregation of the complex and multidimensional concept of state capacity is a viable solution to mitigate the problems related to the frequent mismatch between theory and measurement. We have seen throughout the chapter at hand that a focus on specific attributes of state capacity circumvents the difficulties in capturing state capacity in its whole and makes it easier to understand what is precisely being discussed.

If a measure should fully cover a given definition, as argued by Bollen (1989), focusing on clear-cut aspects of state capacity allows researchers to choose instruments that match more precisely with a particular definition. A drawback caused by the specificity of my empirical analysis is that the findings of the study at hand cannot be considered as informative about the state-democracy nexus as a whole. Nevertheless, they are more than useful in understanding how the "sequencing" might work in practice through more precise mechanisms and provide interesting insights to grasp the larger picture.

Framed by the abovementioned argument, this chapter provides also an important contribution to the topical sequencing debate by analysing the relationship between two specific characteristics of democracy and state capacity. My findings bring forth persuasive evidence of a statistically significant association between civic participation

and impartial bureaucrats. A more vibrant civic community seems to constrain public officials to act more impartially both in the short run and in the long run. My findings suggest as well that besides exploring the sequentiality between democracy and state capacity as a whole, the research agenda on the state-democracy nexus would benefit from focusing on more specific characteristics not only of state capacity but also democracy.

Some scholars have recently called for a more disaggregated approach on the state-democracy nexus. For instance, according to Andersen and Doucette (2020: 7), "future research should study whether different aspects of democracy strengthen bureaucratic quality". The study at hand can be considered as a response to these demands and one of the first attempts to analyse the relationship between democracy and state capacity with a disaggregated approach. Besides the self-evident conceptual and theoretical implications of such approach, it is important to stress that focusing on clear-cut aspects of democracy and state capacity allows scholars to choose tools of measurement that match more precisely with the selected concepts. My study has focused on the impact of civic participation on impartial bureaucracy, but I cannot exclude the existence of possible reverse causal association between the two factors. Future studies are needed to assess both causal directions of the relationship and the relationships among other disaggregated aspects within the state-democracy framework.

Last, even if the evidence of an overall positive association between civic participation and impartial public officials is strong and robust across several specifications and models, the relationship seems to be sensitive to macro-regional differences. In Western countries, Sub-Saharan Africa, Latin America, and Asia the hypothesised relationship holds at conventional levels of statistical significance both in the short run and in the long run. In Former Soviet Union countries and in the Middle

East and North Africa, instead, civic participation does not seem to be conducive to a more impartial behaviour by public officials. The causes of these macro-regional differences call for further investigation in future studies on the topic. It could be that the state-democracy nexus works through other mechanisms in Former Soviet Union countries and in the Middle East and North Africa, but it could be as well that democracy and state capacity are simply not related to each other in these two specific macro-regions.

Chapter IV

Estimating State Capacity

On the Development of a Novel Cross-National Measure of State Capacity

Introduction

Chapter II showed us that even the most established measures of state capacity can have some undesirable qualitative and quantitative properties. In particular, as to their content, some of the measures contain unnecessary components and others do not contain even the most basic components that we would expect a valid measure of state capacity to be based upon. As we have seen in Chapter III, one effective way to tackle measurement issues in the state capacity literature is to shift the level of analysis from general to very specific aspects of state capacity. Anyhow, a disaggregation of the concept into more specific components has its drawbacks as well.

State capacity is an inherently multidimensional concept, and both scholars and practitioners might want to quantify and analyse state capacity as a whole, or at least in its major dimensions, and not just one single specific aspect of it. While increasing our knowledge on specific characteristics of state capacity is one way to make progress in the field of study, the interest in the broad and complex concept of state capacity remains topical for academics and a broader audience as well. Testing theories related to state capacity as a whole and gaining robust knowledge about its relationship with other relevant social, political, and economic factors, however, is not achievable with poor measures. Research on state capacity requires sound and appropriate measures that

capture the core dimensions of the concept without overstretching common definitions of state capacity.

Having said that, measuring state capacity is not easy by any means. As many other social science concepts, state capacity cannot be directly observed, and thus, at best, it can be estimated (Lindvall and Teorell 2016). Since unobserved concepts must be necessarily defined to have any meaning (Neumayer and Plümper 2017), defining state capacity becomes of primary importance. In reviewing some of the definitions of state capacity previously in this study in Chapter I, we have seen that many scholars agree that state capacity, broadly understood, has to do with the ability of the state to implement its intended policies and thus to reach its goals. The common understanding decreases radically, however, when it comes to estimating and quantifying state capacity: countless of different measures have been used to capture state capacity[11]. Furthermore, usually, these different measures have not been developed with the intention to capture state capacity as a whole, so their content validity, understood as the matching between a given measure and its definition (Bollen 1989), remains open to debate.

Neumayer and Plümper (2017) identify three strategies that scholars use to quantify unobservable concepts: single proxies, subjective measures, or multiple proxies turned into a single variable.

In the case of state capacity, single proxies such as *income tax/GDP* or *military personnel per capita* can be useful to quantify specific aspects of capable states, but lack in content validity if the objective is to capture a broader multidimensionality of state capacity. It is self-evident that the share of income tax – or any other tax – does not fully capture the concept of state capacity. Likewise, the share of military personnel taps a very

---

[11] Besides Chapter I in the study at hand, see Hendrix (2010), Cingolani (2013), Saylor (2013), Savoia and Sen (2015), and Hanson (2018) for previous reviews of measures of state capacity.

specific component of coercive functions of capable states and cannot be considered a valid measure of state capacity as a whole.

Subjective measures of state capacity instead are easy to criticise because of their inherent subjectivity: "subjective quantifications will inevitably resort to taking more observable causes or consequences of the concept into subjective account" (Neumayer and Plümper 2017: 116), and thus, contain measurement error (Bollen 1993). Whether the information is collected through expert surveys or citizen surveys, perceptions of state capacity are likely to be affected at least by state performance in expected consequences of state capacity, such as economic growth and sustainable development. Nevertheless, there is a lack of "objective" data regarding many aspects of state capacity and often researchers have no other solution than to consider subjective measures. The administrative dimension of state capacity, in particular, suffers from a lack of objective data. Furthermore, in the case of state capacity, existing subjective measures are at best approximations of certain characteristics of the state and are not designed to capture perceptions of state capacity as a whole.

In the chapter at hand, compared to the two aforementioned strategies – single proxies and subjective measures – the aggregation of multiple proxies into a single variable is most likely to produce the desired outcome, which is a quantification of state capacity that captures its core dimensions. An abundant literature on the advantages and disadvantages of aggregate measures of governance has emerged in the last decade or so. For instance, the aggregation of individual indicators of the same construct can reduce measurement error (Kaufman, Kraay, and Mastruzzi 2011) and has the potential ability of capturing fully a multidimensional concept. Possible shortcomings, such as the amplification of common measurement errors (Arndt and Oman 2006) and the obfuscation of the interpretation of what is being measured (Andrews, Hay and Myers

2010) seem to be outplayed by the advantages of aggregating a variety of proxy measures of state capacity. These possible shortcomings are addressed also in the next parts of this chapter.

Research Strategy

In constructing a new cross-national measure of state capacity that provides data not only across countries but also over a reasonable time period – the time-series dimension is essential for the concept (Kurtz and Schrank 2012) – I follow primarily OECD's (2008) well-established framework for building composite measures and focus on its five main steps: (1) development of a theoretical framework; (2) selection of data; (3) standardisation and imputation of missing data; (4) weighting and aggregation of data; (5) validation of the novel composite measure.

The next steps of my research proceed as follows: First, the concept of state capacity is defined and anchored to a theoretical framework. Second, measures related to the core dimensions of state capacity are identified and collected from a variety of secondary sources. The content validity of any composite measure is highly affected by this second step, so it is crucial to select indicators that fit well the concept. Third, selected indicators are standardised, and missing values are estimated with multiple imputation. The multiply imputed data is further analysed, to ensure that estimated values are reasonable both within units (over time) and between units (across countries). Fourth, selected indicators are weighted and aggregated with exploratory factor analysis (EFA). This approach allows to generate knowledge about the "true" dimensionality of state capacity, base the weighting on "objective" statistical dimensions of the underlying data, and to aggregate possible sub-components of state capacity according to these objective weights. Fifth, the

new composite measure is validated with commonly used tools that bridge different traditions of measurement validation, as recommended by Seawright and Collier (2014).

Constructing the New Index

Theoretical Framework

Defining the concept is a fundamental step in developing composite measures, but often producers of governance indices skip completely this step (Gisselquist 2014). In fact, neglecting concept specification is the main problem of extant measures of state capacity (Fukuyama 2013). Thus, first of all, before quantifying state capacity, it is essential to pay attention to some definitional and conceptual aspects related to it. In the introductory chapter of my dissertation we have already reviewed several different definitions of state capacity and have reached the conclusion that state capacity has been defined in a variety of ways. Yet, some definitions are more common than others. According to Acemoglu and Robinson (2019: 12) "the capacity of a state is its ability to achieve its objectives". This is also how state capacity is understood in the chapter at hand.

Without discussing again all the different understandings of the state and state capacity, comprehensively reviewed in Chapter I, we shall briefly remind ourselves about the core dimensions of state capacity that researchers have identified and that should be captured by a multidimensional measure of state capacity to ensure high content validity. To be faithful to the choices made in Chapters I and II, we shall continue with Savoia and Sen's (2015) classification, which seems to encompass the most common functional dimensions of state capacity. To recall, these dimensions are (1) administrative capacity, (2) legal capacity, (3) infrastructural capacity, (4) fiscal capacity, and (5) military or coercive capacity. In selecting the indicators, I draw upon these five functional dimensions of state capacity.

We shall remember as well, however, that besides identifying the functional dimensions of state capacity, some scholars have highlighted the difference between the power of the state and the procedures of the state.

Bäck and Hadenius (2008: 3) identify two criteria of functioning states: "the capacity of the state organs to maintain sovereignty over a geographic territory" and the capacity of the state organs to carry out well their tasks. While according to the authors, the first criterion has to do with basic military, legal, and fiscal functions, the second criterion has to do with *how* various activities of the state are implemented. The first criterion is more related to the power of the state, whereas the second criterion has to do with sound procedures. Rothstein and Teorell (2008: 170) develop further the criterion of *how* the activities of the state are implemented through their theory of "impartiality in the exercise of public power": if a professional bureaucracy is to work well, it needs to act impartially.

Fukuyama (2013) tries to make a clear distinction between capacity and procedures but acknowledges that certain procedural aspects are indispensable for any measure of state capacity. Impartial procedures alone, however, cannot guarantee that a state gets things done as intended, because a capable state needs also resources and a professional bureaucracy (Fukuyama 2013). On similar lines of thought, Lindvall and Teorell (2016) propose to distinguish between procedural constraints and power. They suggest that measures of state capacity should focus only on the latter, providing practical examples of some of the key resources of powerful states: government funds, well-trained public officials, and information about the society it governs. These attempts to distinguish between the power dimension and the procedural dimension of state capacity are conceptually appealing, but even so, they fail to provide more specific insights on how to differentiate power and procedures one from the other in practice.

Just to give an example, we know that to get things done states need to have a certain amount of power over the society, concretised in resources and authority and embodied in the administrative apparatus of modern states. A professional administrative apparatus, however, has procedural constraints by definition, first and foremost because professionality entails at least a certain amount of impartiality: basic characteristics of a professional bureaucracy, such as meritocratic recruitment, imply the existence of procedural constraints. As put by Rothstein (2019: 26), "meritocracy, everything else being equal, increases the competence in the public sector and thereby state capacity".

A professional bureaucracy that works under some procedural constraints, such as the principle of impartiality, is likely to be better in getting things done. Since civil servants working under the principle of impartiality need to follow the law and cannot be engaged in particularistic behaviour (e.g., nepotism, clientelism, familism), the ability to get things done as intended is by logic reasoning higher than in partial and unprofessional bureaucracies, *ceteris paribus*. Obviously, this is true unless it is a policy decision to do things unprofessionally, which is a highly unimaginable option that the study at hand does not take into further consideration.

Despite these difficulties in disentangling power and procedural constraints one from another, I try to map the selected indicators also in terms of these two "alternative" dimensions of the state, which are concerned with the very essence of our definition of state capacity (i.e., the ability to reach its goals regardless of the functional dimension) rather than the different functions of capable states. Anyway, it is important to acknowledge that theoretical literature does not provide a clear-cut path on how to measure state capacity in practice.

First and foremost, in asking what a governance index aims to measure, Gisselquist (2014) raises differentiation from other concepts as one of the important issues. This is

true especially when it comes to state capacity, which suffers from somewhat blurred conceptual limits. Hence, differentiating between state capacity and other similar concepts becomes a crucial step in ensuring the conceptual validity of a new measure, but not only. It enhances also the analytical usefulness of the new index. For instance, if we want to gain knowledge on the relationship between state capacity and democracy, which has been a recurrent topic of research also in my dissertation, we cannot conflate the two concepts. Blending state capacity together with democracy would not only go against the "principle" of conceptual differentiation but also restrict the analytical usefulness of the new measure, given that it is impossible to analyse the relationship between state capacity and democracy if the two are combined to begin with.

Besides democracy, also aspects strictly related to economic wealth and socio-economic development should be excluded from a measure of state capacity. Researchers are often interested to assess whether state capacity is indeed related to different aspects of socio-economic well-being. We have also seen that achieving state capacity is one of the UN's Agenda 2030 Sustainable Development Goals. Hence, it is essential also for policymakers that a measure of state capacity allows to assess the causal relationship between state capacity and developmental outcomes. If we are to make any progress on the state capacity research agenda, concepts such as democracy and development should not be conflated with (measures of) state capacity. In general, Munck and Verkuilen's (2002) recommendations related to measures of democracy apply to measures of state capacity as well: scholars should avoid including theoretically irrelevant attributes in the measures and excluding theoretically relevant attributes from the measures they use.

Ideally, a measure of state capacity allows its users to analyse both the causes and the consequences of capable states, but also provides comparative information on state performance. Several scholars have argued that state capacity should be quantified with

output measures (e.g., Andrews 2008; Kurtz and Schrank 2012; Grävingholt, Ziaja, and Kreibaum 2015) such as access to improved water sources, child mortality, and primary school enrolment. Other scholars, instead, argue that "it might be better to leave output as a dependent variable to be explained by state quality, rather than being a measure of capacity in itself" (Fukuyama 2013: 355). From the above discussion it is obvious that quantifying state capacity with outputs is not a viable option, if we are interested in understanding the causal relationship between state capacity and any of these outputs. Hence, at least some output measures need to be left out. For instance, evaluating the relationship between state capacity and child mortality becomes impossible if the two are conflated to begin with.

Selection of Data

The new composite measure of state capacity is based on existing secondary data from several sources. Both "subjective" perceptions-based indicators and "objective" facts-based indicators are used to mitigate deficiencies of both types of data, although there are more relevant perceptions-based indicators than facts-based indicators available on the topic. Following the advice of Munck and Verkuilen (2002), many overly general facts-based indicators that have been previously used to measure state capacity, such as state antiquity, are excluded from the data collection process, because they capture also many theoretically irrelevant attributes. Indicators that we are interested to predict with or be predicted by state capacity, such as democracy, GDP/capita, and economic growth, are not taken into consideration, either. Chosen sub-indicators are presented in Table 4.1.[12]

---

[12] Besides the indicators presented in Table 4.1, a measure of information capacity from Brambor et al. (2020) was included in the preliminary analysis but dropped out after the results of EFAs which indicated that the measure is not sufficiently related to the other items.

Table 4.1. Components of the new index of state capacity

Indicator	Producer	Focus	Dimension	Years	Countries	Missing data
State authority over territory	Varieties of Democracy	Coercion/infrastructural	Power	1984-2018	158	0.02%
Internal conflict	International Country Risk Guide	Coercion	Power	1984-2016	133	21.25%
Political violence	Varieties of Democracy	Coercion	Power	1984-2018	129	22.22%
Government revenues	ICTD/UNU-WIDER	Extraction	Power	1984-2017	142	35.19%
Absolute political extraction	Fisunoglu et al. (2020)	Extraction	Power	1984-2018	152	5.53%
Tax revenues	ICTD/UNU-WIDER	Extraction	Power	1984-2017	145	25.19%
Bureaucratic remuneration	Varieties of Democracy	Administrative	Power	1984-2018	157	3.77%
Bureaucracy quality	International Country Risk Guide	Administrative	Procedures	1984-2016	133	21.25%
Bureaucratic recruitment criteria	Varieties of Democracy	Administrative	Procedures	1984-2018	157	3.45%
Access to electricity	World Bank	Infrastructural	Power	1990-2018	156	29.83%
Urban population	UN Population Division	Infrastructural	Power	1984-2018	174	1.25%
Public sector corruption	Varieties of Democracy	Administrative	Procedures	1984-2018	158	0.00%
Corruption	International Country Risk Guide	Administrative	Procedures	1984-2016	133	21.25%
Judicial corruption	Varieties of Democracy	Legal/administrative	Procedures	1984-2018	158	0.40%
Property rights V-Dem	Varieties of Democracy	Legal	Procedures	1984-2018	158	0.00%
Property rights HF	Heritage Foundation	Legal	Procedures	1995-2018	156	34.24%
Law and order	International Country Risk Guide	Coercive/legal	Procedures	1984-2016	133	21.25%
Access to justice	Varieties of Democracy	Legal	Procedures	1984-2018	158	0.00%
Contract intensive money	International Monetary Fund	Legal	Procedures	1984-2016	131	39.49%
Impartial public officials	Varieties of Democracy	Administrative	Procedures	1984-2018	158	0.00%

Note: Countries and Years indicate respectively the number of non-missing countries and the range of non-missing years in my dataset before imputation

We have seen that there is substantial room for improvement in existing measures of state capacity, especially when it comes to content validity. To be high in content validity, a given measure should capture adequately a given concept by avoiding the exclusion of key elements from the measure and the inclusion of unsuitable elements in the measure (Adcock and Collier 2001). Most of the frequently used measures have not been created for the purpose of estimating state capacity but some other related factors, and thus, lack often in content validity when used to estimate state capacity. As we have seen in Chapter II, many frequently used measures such as State Fragility Index (SFI), Failed States Index (FSI), Corruption Perceptions Index (CPI), and Government Effectiveness (WGI) capture elements that undermine their analytical usefulness and lead to conceptual overstretching. For instance, just to recall some of the arguably unwanted "ingredients", SFI and FSI capture elements related to democracy, whereas WGI and CPI seem to reflect pro-business attitudes.

The collected data captures the five individuated dimensions of state capacity. *Internal conflicts (ICRG)* is an assessment of the amount of political violence within a country. States devastated by civil wars have the lowest possible score on the indicator (PRS Group 2018). *Political violence (V-Dem)* captures the extent of the "use of physical force to achieve political objectives by non-state actors" (Coppedge et al. 2020: 211). *State authority over territory (V-Dem)* is an assessment of areas over which the state has effective control within its territorial borders. These three indicators focus mainly on the coercive dimension of the state, although the last captures its infrastructural attributes as well. Frequently used "objective" measures of coercive aspects, such as military expenditure or military personnel, are left out because the size of military should not be mixed up with coercive capacity. Just to give an example, Costa Rica and Iceland have no army at all but both countries are still able to effectively monopolise the use of force.

Government revenues/GDP (ICTD/UNU-WIDER) and *tax revenues/GDP (ICTD/UNU-WIDER)* capture the available resources of the state and its extractive capacity. *Absolute political extraction (RPC)* attempts to provide an estimate of the extractive capacity of countries through economic modelling (Fisunoglu et al. 2020). These three measures focus on extractive aspects of the state. *Access to electricity (WB)* is an estimate of the percentage of the population that has access to electricity. *Urban population (UNDESA)* measures the number of people that live in urban areas as a percentage of total population in a country. These two indicators capture the infrastructural capacity of the state. Both the distribution of infrastructure such as electricity (Koren and Sarbahi 2018) and the level of urbanisation (Cole 2015) can be seen as reflections of the territorial reach of the state.

Bureaucratic remuneration (V-Dem) is an assessment of the extent of employees of the state that are salaried. *Bureaucracy quality (ICRG)* captures the strength and the quality of a bureaucracy. Public administrations with established recruitment and training mechanisms score high on the indicator (PRS Group 2018). *Bureaucratic recruitment criteria (V-Dem)* quantifies the level of meritocracy in appointment decisions in the state administration. *Impartial public administration (V-Dem)*, captures the rigorousness and impartiality of public officials in conducting their duties. *Corruption (ICRG)* is an assessment of both actual and potential corruption through nepotism, favouritism, excessive patronage, and alike forms of partial behaviour. *Public sector corruption (V-Dem)* is an assessment of both petty corruption and public funds embezzlement by public officials. In terms of content, these indicators reflect the administrative dimension of state capacity.

Judicial corruption (V-Dem) refers to the normality of bribes and undocumented extra payments that are aimed to affect judicial decisions. *Property rights (HF)* measures

153

the degree of property rights protection and enforcement by the state. *Property rights (V-Dem)* is an assessment of the right to private property in a given country. *Law and order (ICRG)* focuses on the strength and impartiality of the judicial system, also in terms of popular observance of the law. *Contract intensive money (IMF)* is estimated as the ratio of non-currency money to total money and captures "security of contract and property rights" (Clague et al. 1999). *Access to justice (V-Dem)* is an assessment of fair trials and the extent to which individuals can take legal action if their rights are violated by public authorities. In terms of content, these indicators reflect the legal dimension of state capacity.

We have seen previously in this dissertation that different scholars suggest different solutions to the problem of quantifying state capacity in the most appropriate way. While many researchers would agree that "the capacity of a state is its ability to achieve its objectives" (Acemoglu and Robinson 2019: 12), there is no agreement on how to operationalise such a definition. Hence, in the next steps of my attempt to develop a new measure of state capacity, I undertake an exploratory approach to provide us more information on the true statistical dimensionality of state capacity.

Standardisation and Imputation of Missing Data

We have seen previously that one of the problems of many existing measures of state capacity is their limited amount of data points, not so much across countries but especially over years. For example, of the measures that we analysed and compared more thoroughly in the Chapter II, FSI is available only from 2005 onwards, CPI and SFI are available from 1995 onwards, and WGI is available from 1996 onwards. HSI is available from 1960 onwards but provides scores only until 2009. The most important exception is Impartial public officials (VDEM), which covers a consistently longer time period than

the other measures, covering all the years from 1789 onwards. Anyhow, as we already know, given its narrow scope, VDEM cannot be considered to be a valid indicator of the whole multidimensional concept of state capacity, in terms of content.

Since the time aspect is important in for researchers working on the topic of state capacity, which often varies slowly over time instead of changing abruptly from one year to another, a good measure of state capacity extends as much as possible from the present to the past. Most of the indicators that can be used to estimate state capacity, however, do not cover extensive periods of time. Furthermore, one indicator might cover more recent years whereas another indicator might lack in recent data but cover more comprehensively earlier years. This means that we have to deal with missing data.

Figure C1 (Appendix C) illustrates the patterns of missingness in the variables in our imputation dataset. Overall average missingness in our imputation dataset is 11.9%, while overall average missingness of the sub-components of the index is 14.3%. Most variables in the imputation dataset cover the entire period of time from 1984 to 2018, although, there are some exceptions to the rule. Of the indicators that are used in the subsequent factor analysis, property rights (HF) and access to electricity (WB) are not available before 1995 and 1990, respectively. ICRG indicators and contract intensive money are available only until 2016, whereas the original ICTD/UNU-WIDER indicators are available until 2017.

First of all, it is important to assess the reason of missing values. Following Little and Rubin (2002), data can be missing at random (MAR), missing completely at random (MCAR), and missing not at random (MNAR). Each of these three categories entail different preferable solutions regarding how to deal with missing values. In the study at hand it is plausible to assume that the data is MAR, which means that the missingness in the data depends on observed variables (Little and Rubin 2002). For example, we can

assume quite confidently that missing state capacity data depends on the level of development. Underdeveloped countries are likely to have more missing data than developed countries, simply because there is less information about underdeveloped countries compared to the developed ones. Because of the abovementioned reasons, in our case, we can exclude a MCAR situation. While it is not possible to exclude categorically a MNAR situation, in which the cause of missingness depends on unobserved variables, an erroneous MAR assumption has at worst a negligible impact on the estimates (Schafer and Graham 2002).

There are three main approaches that are used to account for missing data when creating indices: (1) listwise deletion, (2) single imputation, and (3) multiple imputation (OECD 2008). The common practice of simply performing a listwise deletion of missing data does not seem to be a good idea with the data at hand. First, a listwise deletion would drop from the analysis a significant number of countries, even if data points are available for most indicators, throwing out a large amount of interesting information in the best-case scenario. Since the amount of missing values in my data set is more than 5%, case deletion is not an appropriate solution (Little and Rubin 2002). Second, given the assumption of a MAR pattern in the data at hand, listwise deletion should be excluded, because it provides unbiased results only under MCAR (Blackwell, Honaker and King 2017). Some single imputation methods are unbiased under MAR condition but reduce the variance of the sample and fail to take into account the uncertainty of the estimated values (Honaker and King 2010). Multiple imputation methods instead, besides being unbiased under MAR, provide also information on the reliability of each imputed value (OECD 2008) and "final results can be interpreted as if data were not missing" (Curley et al. 2019: 597). Furthermore, since there are no any clear advantages of single

imputation over multiple imputation (Newman 2014), I proceed by filling in the missing values with multiple imputation.

The selected multiple imputation method, developed by Honaker and King (2010), is one of the most popular multiple imputation approaches in social sciences and particularly suitable for longitudinal data. It is implemented through an expectation-maximisation algorithm with bootstrapped parameters with the *Amelia II: A Program for Missing Data* (Honaker, King, and Blackwell 2011) package for *R*. The expectation-maximisation algorithm assumes that the data follows a Gaussian distribution (Honaker and King 2010), which I take into account in the imputation procedure. To make the MAR assumption as plausible as possible, the imputation dataset includes nearly 50 political, social, and economic variables. Moreover, I add into the imputation model country-level fixed effects that are interacted with second order time polynomials to take into account possible country-specific time trends. After the multiple imputation procedure, and before the next steps, I visually analyse the results to ensure that the filled-in values are reasonable alone and also in relation to the observed values. Finally, I standardise (z-score: mean of 0 and standard deviation of 1) the data within time periods.

While there are some misconceptions about conducting statistical analyses with multiply imputed data, recent studies suggest that multiple imputation should be preferred over other common methods that are used to deal with missing data unless only a few cases are missing (e.g., van Ginkel et al. 2020). It is worth mentioning that my attempt is not by any means the first one where missing data is filled in. In fact, different methods of dealing with missing data have been even used in the construction of extant state capacity indices. For instance, Fortin (2010) fills in missing values with a combination of multiple imputation and interpolation, whereas Ziaja, Grävinholt, and Kreibaum (2019) use a combination of extrapolation and interpolation to deal with missing values. Each

approach has its advantages and disadvantages, but regardless of the chosen method, it is good to remember that the scores of any aggregate index of state capacity should be interpreted with caution.

Weighting and Aggregation

Measurement theory literature makes an important distinction between reflective indicators and formative indicators (e.g., Bollen 1989; Maggino 2017). Put simply, a reflective measurement model is appropriate for instruments that are correlated among one another and are thought to be caused by latent factors. In the case of state capacity, different dimensions are likely to be mutually supporting, instead of being unrelated among one another (Hanson and Sigman 2013). On the contrary, in a formative measurement model, observed indicators do not necessarily need to be correlated among one another – actually, they should not show high intercorrelations among one another – and are thought to cause the unobserved indicator itself. Formative indicators are composed by a set observed variables "which are not assumed to have conceptual unity and form an exact weighted linear composite" (Bollen and Diamantopopulos 2017).

In theory, determining the causal direction between latent variables and observed indicators might seem relatively simple. In practice, however, it is often not possible to establish the causal direction with certainty. Thus, any measurement model must be regarded, at best, as an approximation of reality (Bollen 1989). In the study at hand, it seems that most of the chosen measures are likely to be caused by latent factors, rather than are causing them. It is possible to assume with great confidence that indicators based on surveys in which experts answer to questions related to attributes of state capacity such as corruption, property rights protection, and monopoly of violence are caused by state capacity. The other way around would imply the absurd assertion of expert perceptions

causing state capacity. With some of our observed measures, however, the causal relation becomes more blurred, which is not surprising since state-building is path dependent (Lindvall and Teorell 2016). For instance, theory suggests that state capacity affects the ability of the state to extract resources, but also that higher revenues determine the capacity-building potential of the state. The uncertainty in determining the causal linkage between latent factors and observed indicators can be considered as limitation of my study.

As we have seen, researchers have different understandings on the dimensions of state capacity. Choosing a reflective measurement model allows me to understand better the dimensional structure of the selected indicators and retain the majority of the variance common to the observed indicators. Moreover, a reflective approach assumes that the observed indicators contain measurement errors, which seems reasonable in the study at hand. Formative measurement models, instead, assume that observed indicators contain no error, which is a highly implausible belief in reality (Edwards 2011). With a latent variable approach to the problem, we are able to get essential information not only about the overall structure of the selected sub-indicators, but also about their true statistical dimensionality, measurement error, and relationship with a "latent" state capacity variable. As we have seen, researchers disagree up to a certain extent on the correct dimensionality of state capacity. As a solution, exploratory factor analysis (EFA) does not provide only statistical evidence on the dimensionality of the concept of interest, but also information about the validity of the observed indicators as measurements of state capacity and about their relationship between state capacity.

Before conducting EFA, as recommended by OECD (2008), I examine the correlation structure of the data. The selected indicators are correlated among each other and in many cases intercorrelations between indicators are considerable (Table C1,

Appendix C). The difference between the correlations before imputation and after imputation is minimal, indicating that multiple imputations have not altered in any significant way the correlations among indicators. According to the logic of the multiple imputation framework (OECD 2008; Honaker, King, and Blackwell 2011), EFA is carried out separately to each of the five imputation datasets. This analysis suggests that one indicator, namely *Information capacity* (Brambor et al. 2020), is not related to the others. Its communality is less than 0.2, and hence, it is removed from subsequent analysis following Child's (2006) recommendation. Only loadings that are 0.32 or more for one of the factors should be interpreted (Tabachnick and Fidell 2013), and in our case all the measured indicators load to one of the two factors at least by 0.32. Cross-loadings should not be high either, although a few cross-loadings can be accepted. The threshold of 0.32, which equals to approximately 10% of variance shared across factors, is valid also for cross-loadings (Costello and Osborne 2005). There are no cross-loadings higher than the common threshold throughout the separate EFAs, but it is important to notice that some of the measured indicators share an interesting amount of variance between the two factors, and that *Law and order* exceeds the cross-loading threshold in some of the EFAs.

EFA is used to analyse the underlying structure of the data and restructure the data into "latent" factors that best represent the common variation in the data. With EFA we are able to explore the underlying relations among the selected observed measures related to state capacity. Moreover, EFA is also a variable reduction technique and my aim is to explore the latent factors that are causing the majority of common variance between the observed indicators. We do not know with certainty, a priori, the structure of the data but EFA allows to circumvent one of the main problems of many previous studies regarding state capacity: the inability to agree upon the correct dimensionality of state capacity. An

EFA on the selected manifest indicators turns this problem into a strength of my analysis, since the true dimensionality is identified statistically by the measurement model.

Surprisingly few social science studies have conducted EFA with multiply imputed data.[13] The main problem that can arise when combining the two procedures, is that the loading structures related to the extracted factors could be different between imputed datasets and that there is no guarantee that the number of factors that should be retained is the same across imputed datasets (Nassiri et al. 2018). In the former case, factors could have different meanings across imputations, whereas in the latter case, it becomes difficult to choose the correct number of factors to be retained. Clearly, one could avoid the problem by pooling the data before conducting an EFA, but the conventional procedure with multiply imputed data is to pool results only as the last step of the analysis (Honaker and King 2010). Initial listwise or pairwise deletion would resolve these problems, but as we have already discussed, would lead also to biased results. Moreover, recent studies have shown that in general, despite the problems of running an EFA with multiply imputed data, common imputation methods performed better than listwise or pairwise deletion of missing observations (McNeish 2017; Goretzko, Heumann, and Bühner 2020).

As recommended by Lorenzo-Seva and van Ginkel (2016), I conduct EFA independently on each of the imputed datasets, run all the analysis separately for each EFA, including rotation and computation of factor scores, and average individual country-level scores only as a last step of the entire procedure. Following Shih, Adolph, and Liu (2012) the entire procedure is conducted after pooling data over time within each imputed dataset, instead of conducting the analysis separately for each year in each

---

[13] Fortin (2012) and Shih, Adolph, and Liu (2012) are some of the few ones.

imputed dataset. Acknowledging that this simplification reduces artificially the possibility that indicators would load to different factors over years and over imputations, the procedure can be justified by an increase in the interpretability of the latent variables.

The results of the five EFAs are extremely similar among each other (Tables C2-C6, Appendix C) and are not sensitive to individual imputed datasets. The Kaiser-Meyer-Olkin measure of sampling adequacy proves that the data at hand is "marvellous" for factor analysis (Kaiser 1974: 35). An analysis of the eigen values indicates that one or two factors should be retained. The first factor explains a large part of the variance in the data, but still, the second factor has an eigen value of around 1.00 across the individual EFAs. The Kaiser criterion suggests retaining factors with eigen values greater than 1.00, leaving us a difficult choice of selecting between a one-factor or two-factor solution. Since there is no error-free method to determine the correct number of factors (Shih, Adolph, and Liu 2012) and since comprehensibility rule (OECD 2008) and theoretical considerations support a two-factor solution, I proceed in the analysis by retaining two latent factors. After extracting the factors by minimising residuals (Harman and Jones 1966), an oblique rotation is performed because I expect the two factors to be correlated with each other.

The communalities, which represent the variance in each manifest variable explained by the retained factors, are below 0.30 for *Bureaucratic remuneration* (0.24) and *Contract intensive money* (0.24-0.26), indicating that a significant amount of variance in these indicators is not captured in the two retained factors. Likewise, the indicators capturing coercive aspects of state capacity (*State territorial control* [0.32], *Political violence* [0.30-0.31], *Internal conflicts* [0.38-0.41]) have also lower communalities compared to the other manifest indicators. On the other extreme of the spectrum, the variance in *Impartial public officials* and *Public sector corruption* is almost entirely

reflected in the two latent factors, which account for 85% and 81% of the variance, respectively. Overall, the two retained factors explain for 54%-55% of the variance in manifest variables. The first factor accounts for 62%-64% of the explained variance. The second factor accounts for 36%-38% of the explained variance.

The manifest variables load to the same factors across imputations, which eases the interpretation of the factors. *Bureaucratic quality* (0.67-0.68), *Public sector corruption* (0.90), *Corruption* (0.63-0.66), *Judicial corruption* (0.75-0.76), *Contract intensive money* (0.34-0.36), *Property rights V-Dem* (0.53), *Law and order* (0.43-0.46), *Access to justice* (0.77-0.78), *Meritocratic recruitment* (0.91-0.92), *Impartial public officials* (0.93-0.94), and *Property rights HF* (0.88) load[14] to the first latent factor in each of th*e five EFAs. Total revenues (0.59-0.62), Taxes/GDP (0.68-0.71), Absolute political extraction* (0.84-0.86), *Bureaucratic remuneration* (0.33), *Access to electricity* (0.78-0.79), *Urban population* (0.55), *State territorial control* (0.41-0.44), *Political violence* (0.35-0.39), and *Internal conflicts* (0.37-0.44) load[15] to the second latent factor in each of the five EFAs. As expected, the two factors are correlated with each other (0.71-0.72)[16], confirming the properness of an oblique rotation instead of an orthogonal one.

Before further analysing the two identified factors, their internal consistency is assessed through a Cronbach's alpha analysis (Cronbach 1951) on the items of each individual factor for each of the imputed datasets. Ranging from 0 to 1, a higher Cronbach's alpha indicates higher internal consistency, and thus, higher reliability. For the first factor the alpha coefficient ranges from 0.93 to 0.94 across the imputed datasets. For the second factor, instead, the alpha coefficient is 0.87 in each individual imputed

---

[14] The numbers in parentheses represent the range of loadings across each of the five EFAs for factor 1.
[15] The numbers in parentheses represent the range of loadings across each of the five EFAs for factor 2.
[16] The numbers in parenthesis represent the range of correlation (Pearson's R) between factors across the five EFAs.

dataset. These results show that both factors are strong and internally consistent. Furthermore, since the reliability of neither of the scales increases if one of the items is discarded, we can conclude that the two scales identified through EFA are reliable.

The results of the EFAs provide interesting information about the relationship between the manifest variables and the retained latent dimensions of state capacity. The two-factor solution supports persuasively the recently posited theoretical distinction between power and procedural constraints (e.g., Acemoglu and Robinson 2012; Fukuyama 2013; Lindvall and Teorell 2016; Johnson and Koyama 2017; D'Arcy and Nistotskaya 2021) as two related but different underpinnings of a broader concept of state capacity. Furthermore, it is supported by some of our previous findings in Chapter II, according to which frequently used measures are most discrepant about individual country-scores along two main dimensions: the power to exercise control on the society on one dimension, and procedural aspects such as corruption, impartiality, and rule of law on the other.

Indicators that are related to the legal aspects of state capacity (judicial corruption, property rights, access to justice, law and order, and contract intensive money) load to the first factor. All these measures entail procedural constraints that determine how the power of the state is exercised. Likewise, indicators that focus on procedural constraints in the bureaucracy (impartial public officials, meritocratic recruitment, bureaucratic quality, public sector corruption, and corruption) load to the first factor. The only measure of administrative aspects of the state that loads primarily to the second factor is bureaucratic remuneration. This makes sense, because procedural constraints should not affect the number of salaried state employees.

Yet, a state with only a few salaried employees is likely to be deficient in resources and expertise, two key aspects of state power (Lindvall and Teorell 2016). Other

indicators that focus on government funds and extractive aspects of the state (total revenues, taxes/GDP, and absolute political extraction) load to the second factor as well. Furthermore, all indicators related to both the infrastructural aspects and the coercive aspects of the state (access to electricity, urban population, territorial control, internal conflicts, and political violence) load mainly to the second factor, supporting the statistically discovered distinction between power and procedural constraints as the principal sub-dimensions of state capacity. For the abovementioned reasons, the two retained factors are named *Power* (Factor 1) and *Procedures* (Factor 2).

Besides providing novel data-driven answers to the theoretical questions related to the concept of state capacity and its true dimensionality, the estimated factor loadings can be used to construct the weights of a composite measure (OECD 2008). Weights are computed from the pattern matrix, which is the primary matrix of interest when oblique rotations are performed (Fabrigar and Wegener 2012; Osborne 2015). Following OECD's (2008) framework, these weights are squared and the manifest indicators with the highest factor loadings are grouped into intermediate composite measures. In our case, these intermediate composites correspond to the two latent dimensions that were discovered through factor analysis. These two intermediate composites, in turn, are aggregated additively into a multidimensional composite measure of state capacity "by assigning a weight to each one of them equal to the proportion of the explained variance in the data" (OECD 2008: 90). The final measure is normalised to range from 0 (low) to 1 (high).

As a sensitivity test, country scores are estimated also with the prediction method proposed by ten Berge et al. (1999), which preserves the correlation between factors in factor scores prediction. The results of the sensitivity test show that the two methods of weighting and aggregation lead to remarkably similar final country scores.

The data-driven approach used in the study at hand – firmly anchored to theory through the process of selecting the most appropriate content – has identified two intermingled dimensions of state capacity: power and procedural constraints on the exercise of this power. Although these two faces of the state have been well-defined in theoretical accounts on the topic, virtually all attempts to identify the fundamental dimensions of the concept and/or to develop a measure of state capacity have relied on functional approaches. Therefore, commonly used composite measures of state capacity do not provide adequate means to distinguish between power and procedural constraints, nor to estimate these two aspects of a broader concept of state capacity. The inability to portray theoretically rooted dimensions in broad measures of state capacity, has been one of the drivers of the criticism towards composite measures of the state, which have been considered as either too broad or too narrow, but in any case, theoretically ambiguous. The novel composite measure of state capacity enables to capture these two core dimensions of capable states and to represent them in accordance to their statistical importance.

From an analytical point of view, a new composite measure that can be disaggregated into its two core components, has the potential of becoming a more useful tool than many of its antecedents. We can easily imagine a situation where a given country has the power to execute intended policies but lacks in procedural constraints on the exercise of this power. We can also imagine an opposite situation where a given country has strong procedural constraints, but not the power to execute its intended policies. The novel measure of state capacity developed in the study at hand, allows to analyse the two identified core dimensions of state capacity separately as well as together.

Validation

To validate the new composite measure of state capacity (from here on CAP), I focus mainly on two tools that are frequently used to assess the validity of composite measures:[17] convergent validation, and case-based validation.

The content validity of CAP is not really discussed here, not because content validity is not important but because issues related to the content validity of CAP were already evaluated previously in this chapter, through the choice of valid sub-indicators and the reasoning behind the selection process. Therefore, in fact, the content validity of CAP can be largely assessed within the framework of our process of selecting the most appropriate underlying sub-indicators. If we recall Fitzpatrick's (1983: 7) argument that "a content valid test must cover important aspects of the content universe that a user wishes to assess", there is no doubt that through its sub-indicators CAP covers the most common functional dimensions of state capacity and even rearranges them into the two discovered sub-dimensions of the state's ability to reach its objectives: power and procedural constraints.

In terms of the analytical utility of the novel measure, compared to many other frequently used measures of state capacity, CAP enables to analyse, for example, the relationships between state capacity and democracy, state capacity and development outcomes[18], and state capacity and economic growth, because it does not conflate these concepts together. As we have seen in the second chapter, most commonly used measures of state capacity include theoretically irrelevant attributes or exclude theoretically relevant attributes. On the contrary, the content of CAP includes only sub-indicators from the "content universe" of state capacity.

---

[17] See Seawright and Collier 2014 for an overview of a variety of validation tools from different validation approaches.

[18] Except arguably urban population and access to electricity.

Before proceeding with the validation, I examine the correlations between the original observed indicators and CAP, to gain more knowledge on the individual relationships between unimputed sub-component indicators and the final composite measure. In the factor analysis framework, we have already examined how each of the sub-indicators is related to the two latent factors, but we have not seen yet how they relate to the broader composite measure of state capacity.

As presented in Table 4.2, unsurprisingly, correlation coefficients between CAP and its sub-components are relatively high, and the weightings seem to be correctly reflected into the final composite. Only four sub-components have a Pearson's r below 0.60: Bureaucratic remuneration (0.48), State territorial control (0.51), Political violence (0.58), and Contract intensive money (0.59). Three of these are loaded to the second factor (Power), which is correctly supposed to have less influence on CAP compared to the first factor (Procedures) in accordance with the employed weighting scheme. One the other extreme of the spectrum, further confirming the soundness of the final measure vis à vis its two intermediate components, there are six sub-components with a Pearson's r higher than 0.80, which are all loaded to the first factor: Bureaucratic quality (0.84), Public sector corruption (0.87), Judicial corruption (0.86), Access to justice (0.82), Impartial public officials (0.90), and Property rights HF (0.85).

We have already discussed (in Chapter II) that correlations with other measures of the same construct are a common tool to assess convergent validity (e.g., Bollen 1989; Seawright and Collier 2014). Obviously, besides high content validity, to prove that CAP really quantifies state capacity it should have high convergent validity. While correlations with other measures of state capacity do provide preliminary knowledge regarding the convergent validity of the new measure, a case-based validation approach to further evaluate the similarity of country scores vis à vis existing measures of state capacity is

Table 4.2. Correlations of state capacity index and its two sub-components with its original manifest indicators.

Sub-component	R	N
Total revenues	0.71	3422
Taxes/GDP	0.78	3950
Absolute political extraction	0.69	4988
Bureaucratic remuneration	0.48	5081
Bureaucratic quality	0.84	4158
Access to electricity	0.62	3705
Urban population	0.64	5214
State territorial control	0.51	5279
Political violence	0.58	4107
Internal conflicts	0.61	4158
Public sector corruption	0.87	5280
Corruption	0.74	4158
Judicial corruption	0.86	5259
Contract intensive money	0.59	3195
Property rights (V-Dem)	0.67	5280
Law and order	0.74	4158
Access to justice	0.82	5280
Meritocratic recruitment	0.78	5098
Impartial public officials	0.90	5280
Property rights (HF)	0.85	3472

Note: Pearson's correlation coefficients (R) computed between average values across multiply imputed datasets and original unimputed sub-components. All coefficients are significant at the $p < 0.001$ level.

fundamental. As the previous chapter of this study has shown, even highly correlated measures can rate countries with large rating discrepancy. By examining more in depth how some of the countries are rated, their evolution over time, and the relative rank order in selected years, it is possible to understand better the plausibility of individual ratings. With such a "multimethod approach" I bridge two methodologies of validation tradition to improve validity analysis of the newly constructed index, as suggested by Seawright and Collier (2014: 132).

Table 4.3 presents the correlations between the novel measure of state capacity, its two sub-dimensions, and a battery of frequently used indicators of state capacity. Overall, the novel index of state capacity is strongly related to most frequently used measures of state capacity suggesting high convergent validity. More than half of the selected

"common" measures of state capacity are strongly ($r > 0.80$) related to CAP, and only two of these measures are weakly ($r < 0.30$) related to CAP, suggesting that the *State antiquity index* ($r = 0.19$) and the *Relative political capacity* index ($r = 0.23$), might actually be measuring something else than state capacity, at least in terms of how we have understood the concept in this study.

Table 4.3. Convergent validity of CAP: correlations with frequently used measures of state capacity

| Measure | Correlation coefficient | | | N |
	Capacity	Power	Procedures	
Failed states index	0.93	0.81	0.89	2140
Quality of government index	0.89	0.73	0.89	4422
Hanson and Sigman index	0.85	0.78	0.80	3866
State fragility index	0.86	0.86	0.79	3564
Corruption perceptions index	0.89	0.70	0.91	3014
Quality of public administration	0.62	0.40	0.59	822
Relative political capacity	0.23	0.24	0.21	3901
State antiquity index	0.19	0.25	0.14	3726
State legibility index	0.56	0.65	0.46	221
Information capacity	0.56	0.57	0.53	2048
Government effectiveness	0.93	0.76	0.93	3098
Control of corruption	0.92	0.71	0.93	3098
Political stability and violence	0.75	0.67	0.72	3098
Rule of law	0.93	0.73	0.94	3098
Regulatory quality	0.90	0.70	0.91	3098
Constellations of fragility/Capacity	0.85	0.88	0.76	1694
Stateness	0.73	0.67	0.65	1002
Statistical capacity	0.61	0.58	0.53	1759

Note: Pearson's correlation coefficients computed between average values across multiply imputed datasets and original unimputed sub-components. All coefficients are significant at the $p < 0.001$ level.

The associations among some of the common measures of state capacity and the two intermediate composites, *Power* and *Procedures*, can be seen as an evidence of the conceptual consistency of these two dimensions of state capacity. Frequently used measures that capture procedural aspects of state capacity, such as corruption and impartial law enforcement, relate more strongly to Procedures than to Power. The

contrary can be said about frequently used measures that capture power-related aspects of state capacity, such as penetration into society or coercion. Just to give a few examples, *State legibility index* and *State fragility index* are more strongly related to Power, whereas *Control of corruption* and *Rule of law* are more strongly related to Procedures. Overall, the differences in correlation coefficients between common measures of capacity and the two novel intermediate composites are not large but seem to confirm steadily the abovementioned trend.

As we did with the measures surveyed in Chapter II, the basic statistical characteristics of the novel measure are evaluated through violin plots in Figure 4.1. The

Figure 4.1. Violin plots of CAP and its two intermediate sub-component indices
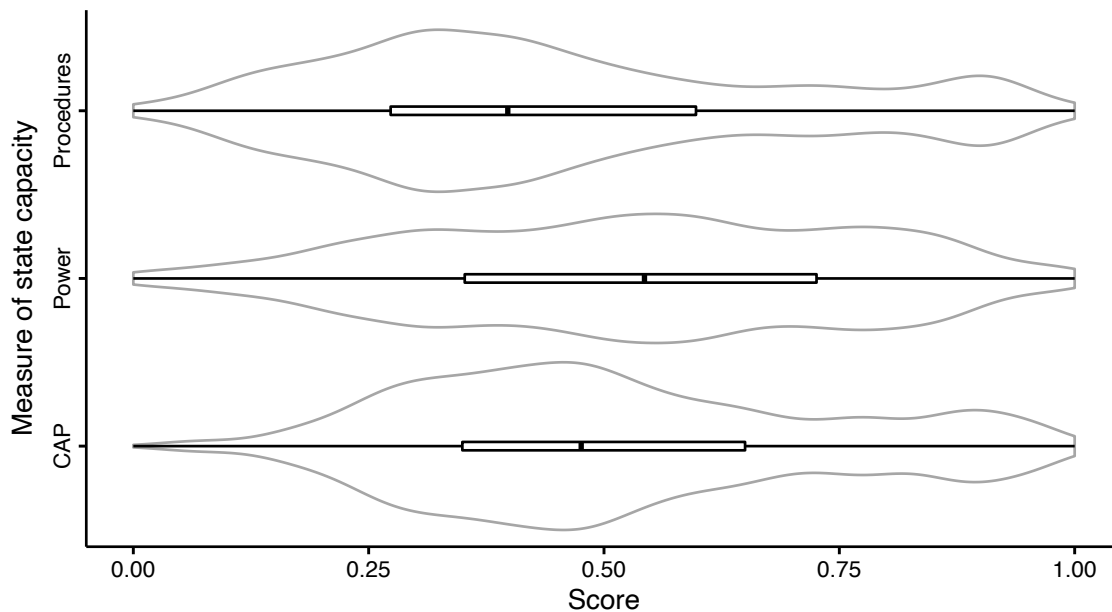
Note: Measures are normalised to range from 0 (low) to 1 (high).

general characteristics of a violin plot are not discussed here, since they were already discussed in Chapter II[19]. CAP and its two sub-components follow relatively a Gaussian

---

[19] Under the "Basic Statistical Properties" sub-section.

distribution and their medians and main modes are more or less at intermediate values of the normalised scale. The median of *Procedures* is slightly below the mid-point of the scale, whereas the median of *Power* is slightly above the mid-point of the scale. The median of CAP lies more or less in the middle of the scale. The three indices have no agglomerations of observations at extremes values nor outlier country-years. Moreover, they have the ability to distinguish between virtually all countries in the dataset. In general, we can conclude the basic statistical characteristics of CAP seem to be satisfactory, especially in comparison to the other frequently used measures of state capacity evaluated in Chapter II.

Table 4.4 shows that the country-ranking of CAP in high capacity states scale is fairly similar to the frequently used measures of state capacity that were thoroughly examined and evaluated in Chapter II. In 2009, which is the most recent year of common observations, some of the largest rank discrepancies between CAP and the other measures can be found in the cases of Hungary (ranked 21st by CAP but 56th by HSI), Norway (ranked 4th by CAP but 30.5 by SFI), and Uruguay (ranked 25th by CAP but 64.5th by QOG). Table 4.5, on the contrary, presents a similar case-based ranking comparison of low capacity states. Again, CAP's country rankings are similar to the other measures, although we can find larger discrepancies than in high capacity states. Some illustrative examples are Syria (ranked 100th by CAP but 52nd by QOG), Venezuela (ranked 101st by CAP but 65th by HSI), Sierra Leone (ranked 110th by CAP but 50th by VDEM), and Libya (ranked 108th by CAP but 54nd by QOG and 61st by SFI).

According to the case-based approach to measurement validity, a measure should reflect as convincingly as possible – through case-specific scores – relevant details of each country (Seawright and Collier 2014). While the above examples provide evidence that CAP's country scores make sense in 2009, we would like to explore also the scores

Table 4.4. Case-based convergence in high-capacity countries: a comparison of country-specific rank-orders between CAP and some of the most established measures of state capacity in 2009

Country	Measure of state capacity							
	CAP	FSI	QOG	HSI	SFI	WGI	CPI	VDEM
Denmark	1	6	2	2	8.5	2	2	1
Sweden	2	3	5.5	9	8.5	4	3.5	7
Finland	3	2	1	7	8.5	3	6.5	8
Norway	4	1	5.5	4	30.5	7	10	5
New Zealand	5	7	3	3	30.5	6	1	9
Belgium	6	13	10	23.5	30.5	13	17	4
Germany	7	19	10	15	8.5	12	11.5	2
Netherlands	8	10.5	5.5	8	8.5	9	6.5	22
Canada	9	10.5	8	12	8.5	8	8.5	14
Austria	10	8	5.5	14	8.5	11	13	20
United Kingdom	11	15	13.5	21	8.5	14	14.5	6
Switzerland	12	4	13.5	5	21.5	5	5	11
France	13	17	18	18	21.5	16	19	17
Australia	14	9	10	6	30.5	10	8.5	3
United States	15	18	16.5	17	30.5	15	16	10
Japan	16	12	19.5	25	8.5	17	14.5	23
Spain	17	25	21.5	34	21.5	27	25	13
Singapore	18	16	13.5	1	30.5	1	3.5	12
Estonia	19	34	35.5	40	21.5	24	23	16
Ireland	20	5	13.5	11	8.5	19	11.5	24
Hungary	21	33	27	56	8.5	31	31.5	19
South Korea	22	22.5	23	32	8.5	23	28.5	32
Portugal	23	14	21.5	22	8.5	21	26	27
Slovenia	24	20	24	19	8.5	22	23	31
Uruguay	25	22.5	64.5	26	38.5	35	20.5	18

Table 4.5. Case-based convergence in low-capacity countries: a comparison of country-specific rank-orders between CAP and some of the most established measures of state capacity in 2009

Country	Measure of state capacity							
	CAP	FSI	QOG	HSI	SFI	WGI	CPI	VDEM
Niger	99	110	107.5	112	113	87	75.5	58
Syria	100	96	52	97	75.5	81	87.5	101.5
Venezuela	101	71.5	120	65	81	106	114	87
Madagascar	102	85	75	102	99.5	99	71	114
Bangladesh	103	107	68	94	87	95	97.5	116
Ivory Coast	104	115	117	116	104	111	109.5	70
Cameroon	105	105	75	104	108.5	102	104	117
Iraq	106	119	113	120	120	113	120.5	112
Mali	107	76	115	100	104	97.5	80	69
Libya	108	54	93.5	108	61	110	92	118
Pakistan	109	117	75	91	108.5	97.5	97.5	113
Sierra Leone	110	103.5	112	103	118	114	104	50
Azerbaijan	111	92	100	77	93.5	88	100	111
Togo	112	97	118.5	114	93.5	118	80	103
Nigeria	113	113	115	105	116	115	92	100
Congo Rep.	114	102	109.5	119	104	116	114	98
Yemen	115	112	109.5	107	108.5	109	109.5	119
Guinea	116	118	88.5	113	118	107	118	107
Haiti	117	116	121	122	99.5	119	118	94
Zimbabwe	118	121	111	110	113	120	104	104
Guinea-Bissau	119	108	96	117	113	108	114	106
Sudan	120	122	115	121	121.5	117	120.5	115
Myanmar	121	111	105	115	118	121	122	122
Congo DR	122	120	122	118	121.5	122	114	123
Somalia	123	123	123	123	123	123	123	121

over time. One way to do so is to analyse the evolution of the country-scores of individual cases over time. Figure 4.2 provides information about the development of state capacity in selected countries around the world from 1984 to 2018. These trajectories seem to support the validity of the novel measure of state capacity.

Figure 4.2. State capacity in selected countries from 1984 to 2018
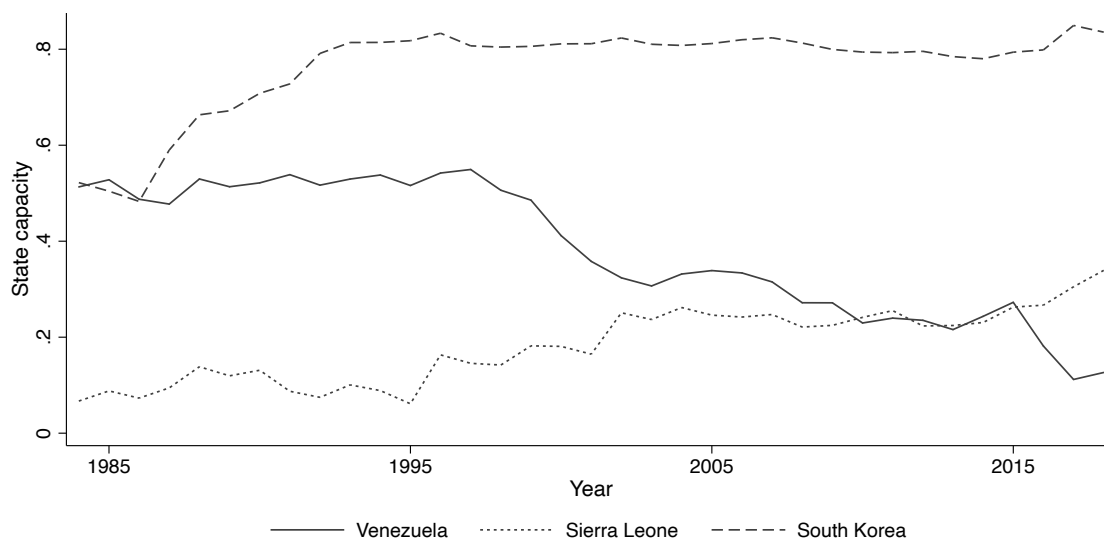
Note: State capacity is measured with CAP.

In most recent years, Venezuela has been topping the news because of its chronic state fragility, and unsurprisingly, it is one of the lowest rated countries by CAP in 2018. In 1984, however, Venezuela had more or less the same level of state capacity of South Korea according to CAP. Scholars have argued that, in the late 1980s, Venezuela managed to reduce corruption and rent seeking through economic liberalisation reforms but started to experience increasing political tensions and internal violence (Di John 2009). These two opposing effects seems to counterbalance each other in Venezuela's score on CAP until the late 1990s. Ultimately, however, downsizing the state led to a collapse in the regulatory capacity of the state in the late 1990s (Di John 2009). This coincides with the beginning of Venezuela's deep decline in state capacity, which seems

to be well-captured by CAP. With the rise of Chávez's to power in 1999, the country started renationalising strategic sectors (Di John 2009) and increased the power of the military (Gan 2020). Instead of increasing state capacity, however, the Bolivarian Revolution led to a collapse of the state's monopoly of violence over its territory, causing the loss of control of some strategic industries to armed groups (Rosales 2019), problems of internal security (Gan 2020), and the inability to protect property rights (Acemoglu and Robinson 2012).

By 1991, Sierra Leone had become a failed state with no central authority, law and order, or property rights protection (Acemoglu and Robinson 2012). Devastated by a violent civil conflict in the subsequent ten years, the country has been sometimes taken as an example of successful international peacebuilding interventions. According to CAP, Sierra Leone was one of the lowest capacity states in the world during the whole 1990s but has managed to increase its capacity since the end of the civil war. Even if corruption is still rampant and impartial legal enforcement continues to be undermined by often arbitrary forms of traditional justice (Sriram 2011), some success in building post-war state institutions has been achieved (Englebert and Tull 2008), particularly in terms of security institutions (Albrecht and Jackson 2014) and formal judicial structures of the state (Sriram 2011). This relatively small but tangible increase in Sierra Leone's state capacity is reflected in its country-score in CAP, which is still far below world average but has surpassed countries such as Venezuela, Nigeria, and Eritrea.

In the early 1980s, after two decades of rapid industrialisation, South Korea was already a comparatively well-functioning state with high fiscal capacity and a professional bureaucratic apparatus (Luiz 1999). Its radical industrial transformation, however, had also eroded the coercive and penetrative capacities of the state (Hellmann 2018), paving the way for large anti-dictatorship demonstrations that culminated to a

successful democratic transition in 1987 (You 2017). The regime change empowered civil society groups, which played an important role in further reducing clientelism and corruption, increasing meritocratic recruitment in the civil service, and strengthening the independence of the judiciary (You 2017). In CAP, these transformations are reflected as a steady increase in state capacity in late 1980s and early 1990s. Since mid-1990s South Korea's level of state capacity has remained stable and high, as evidenced by the successful state-led interventions that allowed the country to recover rapidly from the 1997-1998 East Asian financial crisis (Kalinowski 2008) and the 2008-2010 global economic crisis (Kalinowski 2016).

An evaluation of the three elucidative cases of Venezuela, Sierra Leone, and South Korea provide important evidence that CAP represents well the de facto country-level development in state capacity. These detailed accounts show that country scores have evolved as a reflection of real-world changes in state capacity. What these aggregate country scores of CAP have not been able to show, however, are the differences between the two identified sub-dimensions. Besides providing a scrupulously built composite measure of state capacity that is high in content validity, convergent validity, and case-based validity, the novel measure created in this chapter allows to capture two sub-dimensions of state capacity that no previously created measures is able to capture. Differences between power and procedural constraints, two dimensions of state capacity that have been theorised but not comprehensively quantified in literature, can be analysed with the two intermediate composites.

If the two identified intermediate composites of Power and Procedures are valid measures, in terms of the case-based tradition to measurement validity, we would expect them to represent more or less accurately the differences in power and procedures within individual countries. This means that states with notoriously high power to "control the

means of violence" (Lindvall and Teorell 2016: 16) and "penetrate civil society, and to implement political decisions throughout the realm" (Mann 1986: 113) but notoriously low procedural constraints to exercise this power impartially, or as put by Johnson and Koyama (2017: 11), that are not "governed by general rules which are applied to all citizens equally" such as Belarus, Cuba, and North Korea should have a significant discrepancy between their individual scores in the two sub-dimensions. If the two intermediate composites are measuring what they are supposed to measure, countries such as Belarus, Cuba, and North Korea should have markedly higher score on the power dimension than on the procedures dimension.

Figure 4.3 shows how the two sub-dimensions of state capacity vary from one country to another in 2018, which is the most recent year available in CAP. Providing strong support to the ability of Power and Procedures to represent the abovementioned differences, the individual country scores of CAP of Belarus, Cuba, and North Korea are primarily driven by the power dimension instead of the procedures dimension. In 2018, of all the 157 countries ranked by CAP, Belarus ranks 18th on the power sub-component but 70th on the procedures sub-component, Cuba ranks 13th on the power sub-component but 120th on the procedures sub-component, and North Korea ranks 102th on the power sub-component but 148th on the procedures sub-component. According to CAP, countries such as Ukraine, Turkey, Tajikistan, and Russia are other examples of countries with high Power but low Procedures. On the contrary, countries such as Costa Rica, Rwanda, Chile, and Singapore perform substantially better on the procedures dimension than on the power dimension. Costa Rica ranks 32nd on the procedures sub-component but 94th on the power sub-component, Rwanda ranks 34th on the procedures sub-component but 119th on the power sub-component, and Chile ranks 26th on the procedures sub-component but 69th on the power sub-component.

Figure 4.3. Sub-dimensions of state capacity in 2018
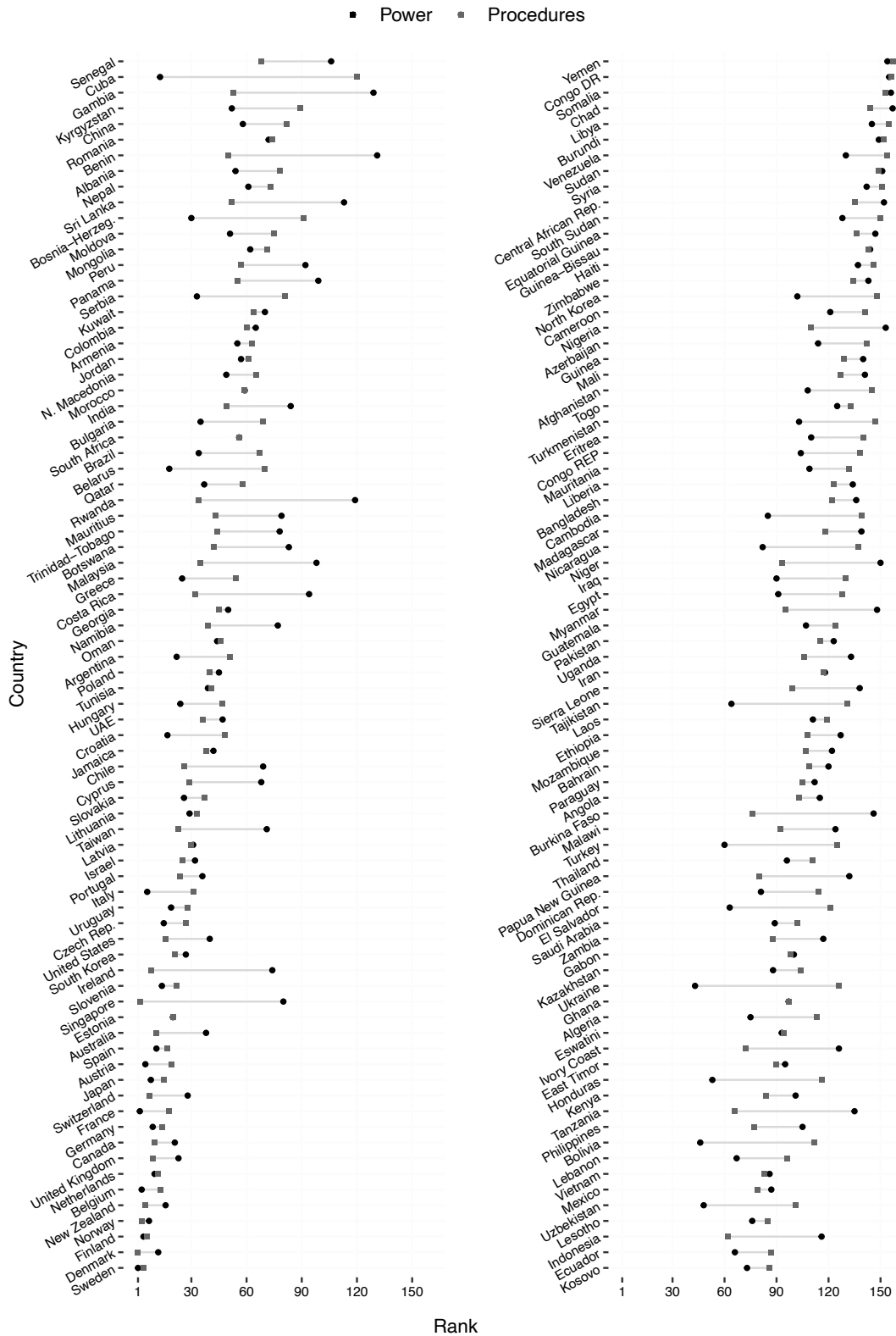
Note: Countries are ranked from top (high capacity) to bottom (low capacity) according to their level of (overall) CAP.

Several frequently used state capacity and governance rankings have rated Singapore as one of the most successful states in the world, at the same level as the Scandinavian countries, at least in the last few decades. Interestingly, the novel proposed measure of state capacity, confirms Singapore's top-level score on the procedures dimension, but ranks the South East Asian island-nation as a middle-rank country in the power dimension. This difference between the two identified sub-dimensions of state capacity can be explained by some of the arguments that have been put forward in recent case studies, where Singapore's deficiencies in its capacity to penetrate society (Woo 2020) and extract resources (Larsson 2013) have been pointed out. On the contrary, in terms of procedural constraints, Singapore continues to be "one of the least corrupt, most law-abiding countries in the world" (Rothstein 2011: 199) with high legal and administrative capacities (Larsson 2013). Despite being a broad and multidimensional measure of state capacity, through its two intermediate composites, CAP seems to be able to paint a more fine-grained picture of the level of state capacity across nations around the world than many other commonly used measures of state capacity. The largest and smallest rank discrepancies between Procedures and Power in 2018 are presented in Table C8 in Appendix C.

Explorative macro-regional comparisons of CAP and its two intermediate components in 2018 (Figure 4.4) show that in most macro-regions there are countries with both high and low capacity. The only exception is Western Europe, North America, and Australia-New Zealand, where virtually all countries have moderately high or high capacity in 2018, with the Scandinavian countries Sweden, Denmark leading the regional ranking, leaving Eastern Mediterranean countries Greece and Cyprus at the bottom. In Asia, Japan and Singapore are the highest ranked countries, whereas North Korea and Afghanistan are the lowest ranked ones. In Middle East and North Africa, Israel and

Figure 4.4. State capacity in the macro-regions of the world in 2018
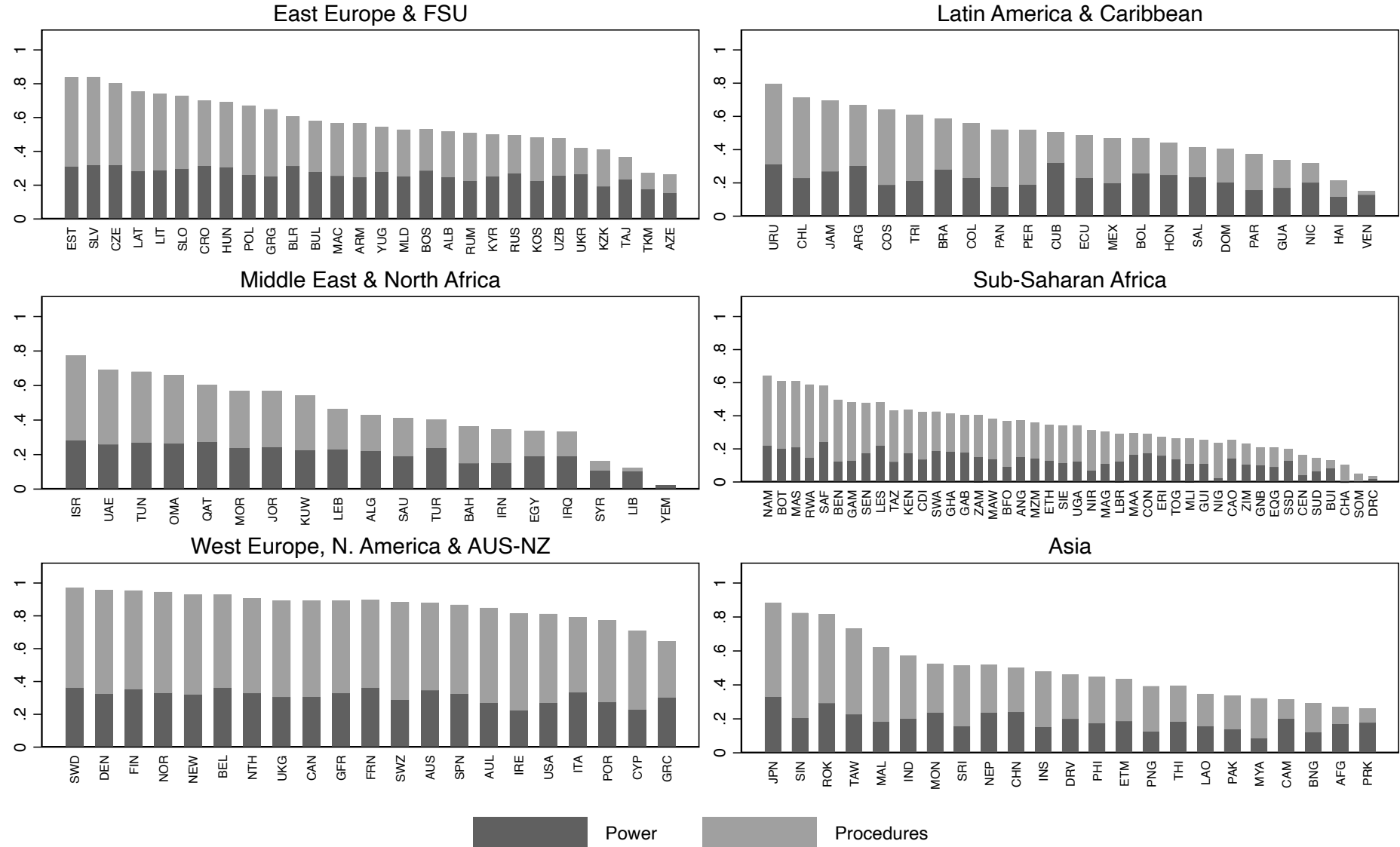
United Arab Emirates are the most capable states, while war-torn Yemen and Libya are the least capable states. In Sub-Saharan Africa the two highest ranked countries are Namibia and Botswana, while the two lowest ranked countries are Congo Dem. Rep. and Somalia. In East Europe and the Former Soviet Union Estonia leads the ranking, followed by Slovakia and Czech Republic. Azerbaijan, Turkmenistan, and Tajikistan, instead, lie at the bottom of the regional ranking. In Latin America and the Caribbean, Uruguay and Chile are the most capable states, whereas Venezuela and Haiti are the least capable ones.

As a last step of my validation procedure, I replicate and re-estimate Hanson's (2015) study on development outcomes, state capacity, and democracy, where the author finds that state capacity and democracy are primarily substitutes for each other. The results of the original study suggest that an increase in state capacity improves development outcomes at low levels of democracy, but that such effect fades out as the level of democracy increases. Vice versa, an increase in democracy improves development outcomes at low levels of state capacity, but such effect fades out as the level of state capacity increases. Similar findings of a compensatory relationship between state capacity and democracy for development outcomes have been discovered by other scholars as well (e.g., Knutsen 2013). Thus, I expect CAP to show an analogous pattern of interaction with democracy when used to predict development outcomes.

As an additional experiment, since the mechanisms that underlie the compensatory relationship have not been further analysed, I test whether such an interaction between state capacity and democracy exists regardless of the sub-dimension of state capacity. Abiding with the rules of the multiple imputation framework, the regressions are run individually with each imputation dataset, including subsequent factor analysis, factor score computation, and aggregation of scores. The results of the individual replication regressions are combined according to Rubin's rules (Rubin 1987), which take into

account both within and between variance in deriving the standard errors. A summary of these combined replication regression results is reported in Table 4.6.

Table 4.6. Replication of Hanson (2015) and re-analysis with the new measures of state capacity

	Dependent variable: ΔInfant mortality $_t$		
	CAP	Power	Procedures
	(1)	(2)	(3)
State capacity $_{t-1}$	-0.813*		
	(0.454)		
Democracy $_{t-1}$	-0.622*	-0.986**	-0.397
	(0.318)	(0.429)	(0.246)
State capacity $_{t-1}$ * Democracy $_{t-1}$	1.286**		
	(0.588)		
Power $_{t-1}$		-1.079**	
		(0.434)	
Power $_{t-1}$ * Democracy $_{t-1}$		1.670**	
		(0.683)	
Procedures $_{t-1}$			-0.380
			(0.382)
Procedures $_{t-1}$ * Democracy $_{t-1}$			0.810
			(0.498)
GDP/capita $_{t-1}$	0.081*	0.076**	0.080*
	(0.042)	(0.038)	(0.044)
ΔGDP/capita	-2.062**	-2.056**	-2.060*
	(1.042)	(1.039)	(1.043)
Population density $_{t-1}$	0.00002	-0.000002	0.00001
	(0.00003)	(0.00003)	(0.00003)
Infant mortality $_{t-1}$	-0.011***	-0.013***	-0.010***
	(0.003)	(0.003)	(0.003)
ΔInfant mortality $_{t-1}$	0.366*	0.363*	0.367*
	(0.187)	(0.188)	(0.187)
R^2	0.39	0.39	0.39
N	4964	4964	4964
Countries	158	158	158

Note: Error correction models with country-clustered heteroskedasticity robust standard errors in parentheses; *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. Constant coefficient measured but not reported. All models include time fixed effects. Measures of state capacity are normalised to range from 0 (low) to 1 (high).

I replicate and re-estimate the regressions concerning infant mortality as a dependent variable. The replications estimated with the "original" method and a similar set of covariates[20] confirm Hanson's (2015) findings. In the absence of democracy, state

---

[20] Hanson (2015) takes the 5-year mean of yearly observations to represent 5-year periods, whereas I use yearly observations without any modifications.

capacity – measured with CAP – plays a key role in reducing infant mortality. Conversely, when state capacity is low, democracy becomes essential for reducing infant mortality. As in the original study, the interaction term of state capacity and democracy is positive and statistically significant. This means in practice that the beneficial effect of state capacity becomes smaller in magnitude as the level of democracy increases, and vice versa (Model 1).

Since the novel measure enables to test separately the mutual effect of each the two individual sub-dimensions of state capacity and democracy on infant mortality, I replace overall state capacity (CAP) in the specification of Model 1 with Power in Model 2 and with Procedures in Model 3, and re-estimate the regression equations with the two intermediate composites. In Model 2, the main terms of power and democracy are both negative and statistically significant, whereas the interaction term is positive and significant. As before, the interpretation is that in the absence of democracy, power reduces infant mortality, but such effect diminishes at higher levels of democracy. In the absence of power, democracy reduces infant mortality, but such effect fades out at as the level of power increases.

In Model 3, instead, the interaction between procedural constraints and democracy is not statistically significant at the 90% confidence level, albeit positive. Likewise, the main terms of procedures and democracy are negative but not significant at the 90% level. The signs of the beta coefficients are the same as in Models 1 and 2, but there is no clear evidence that the effect of procedures on infant mortality would be conditional on the level of democracy, nor that the effect of democracy on infant mortality would be conditional on the level of procedures.

The replicated and re-estimated results suggest that the compensatory relationship between state capacity and democracy is driven by the power dimension of state capacity,

rather than its procedural dimension. State capacity and democracy seem to substitute each other for achieving successful development outcomes, but more due to state power than due to how this power is exercised.

Conclusions

The objective of this chapter has been to show that the complex and multidimensional concept of state capacity can be measured in a more meaningful way. Instead of quantifying state capacity with objective proxy indicators, specific subjective indicators, or overly broad composites, we have seen that a synthesis of several objective and subjective proxies can provide valid, reliable, easily interpretable, and analytically useful estimations of the level of state capacity in countries around the world. My approach to measurement-building has been primarily explorative, because scholars have not been completely unequivocal on the different sub-dimensions of state capacity. Therefore, besides developing a novel tool of measurement, this chapter has also contributed to the conceptual state capacity literature by identifying the core dimensions of state capacity, according to the data at hand. Despite the explorative character of my approach, however, the selected "manifest" data was not just randomly chosen from the universe of available indicators, but its association to state capacity was theoretically discussed and justified.

Through a primarily data-driven exploratory analysis of the true dimensionality of state capacity, I discovered that the selected manifest indicators are best represented by two intercorrelated latent variables named, after theoretical argumentation, Procedures and Power. These two "intermediate composites" were then aggregated into one single broad composite measure of state capacity (CAP). More specifically, the novel measure is based on twenty manifest indicators that were condensed according to patterns of shared variance first into two latent variables and the aggregated additively, following the

recommendations of OECD's (2008) well-known framework for constructing composite measures. Since overall the process of developing the proposed composite measure of state capacity has entailed both theoretical reasoning and statistical guidance, the approach in the study at hand can be considered, ultimately, not only data-driven but also theory-driven.

As a last step of my strategy to develop a new composite measure, I assessed the validity of CAP. Qualitative content validity of the measure was discussed through the initial process of selecting only relevant attributes of state capacity. Convergent validity was assessed by examining the association between CAP and the frequently used measures of state capacity that were analysed in Chapter II. Finally, since a measure is valid when its scores reflect meaningfully the ideas contained in the concept that it is supposed to measure (Adcock and Collier 2001), I evaluated the ability of individual country scores to reflect state capacity as a whole, and as a representation of the two core dimensions that were identified. Overall, CAP seems to be a valid measure of state capacity.

By distinguishing between the two sub-dimensions of state capacity, we have been able to understand also how countries perform according to these two separate aspects. In 2018, while in many countries the two dimensions are relatively in balance, some countries such as Belarus, Cuba, and Turkey have built their state capacity primarily on the power dimension, whereas other countries such as Singapore, Rwanda, and Costa Rica have built their state capacity primarily on the procedural constraints dimension. By replicating Hanson's (2015) study on the compensatory relationship between state capacity and democracy in reducing infant mortality, I have found not only that if we employ CAP instead of the original measure of state capacity the interpretation of the original results stays the same, providing further evidence of measurement

interchangeability, but also that the compensatory relationship between state capacity and democracy in explaining development outcomes is primarily driven by the power dimension instead of the procedures dimension. The distinction between the two core dimensions of state capacity allows us grasp more about the driving forces of state capacity in contemporary nations.

Needless to say, it is important to remember, that the results of this chapter contain a number of caveats and must be examined with caution. For instance, as some critics of aggregate measures of governance have argued, if subjective measures contain common measurement error, aggregating these measures would exacerbate the error instead of mitigating it. To address the potential problem, I have selected both objective and subjective measures from a variety of sources, reducing the plausibility of the occurrence of such common error. The novel measure of state capacity seems to be robust also to interpretational issues: it is clear what it is measuring, and its two sub-dimensions were not created by accident but identified through statistical "objective" analysis. Of course, conceptual dimensionality and empirical dimensionality are not necessarily the same (Bollen and Diamantopopulos 2017), but in the chapter at hand, they seem to match persuasively.

Chapter V

Conclusions

What Have We Learned?

Introduction

State capacity has become one of the most important research topics in many social science subfields, of substantial interest to the international development community, and also relevant for a broader non-expert audience outside academia and policymaking. At a first glance it seems relatively easy to understand what we mean by state capacity. We can distinguish capable states from incapable states almost intuitively. Capable states, by definition, have the capacity to put into action their intentions successfully. Incapable states, by definition, do not have such capacity. Countries without state capacity lack the ability to implement successfully whatever policies they would like to. Such dysfunctional states are widely recognised as one of the major hindrances to socially, economically, and politically flourishing societies around the world. Hence, today, one of the most pressing global challenges is to build and strengthen the capacity of states around the world.

At the time of writing, no reasonable observer would disagree that Sweden and Denmark have the ability to implement their official policies and that Congo (Dem. Rep.) and Afghanistan have serious problems in reaching their official policy objectives. In the former two countries, most citizens enjoy a relatively healthy, wealthy and safe life, whereas in the latter two countries, most citizens suffer a lack of health, wealth, and security. The former two countries are high-capacity states; the latter two countries are

low-capacity states. While the key role of state capacity for the well-being of societies is acknowledged by virtually everyone, understanding in a more detailed way the concept of state capacity, its dimensions, its causes, and its consequences remains a difficult but necessary task. To reach these objectives, the (good) measurement of state capacity becomes fundamental.

The primary objective of this dissertation has been precisely to push forward the somewhat neglected research agenda on the measurement of state capacity. Ultimately, my analysis provides researchers and experts working on the topic essential information on how state capacity has been measured, how it could be measured, and how it should be measured. It delivers past and prospective users of measures of state capacity a toolbox with several new tools that are helpful in better quantifying the concept and better understanding it through its measures. A "tangible" outcome of my study is the novel set of state capacity measures. Anyhow, by reading the study at hand, even those who prefer to quantify state capacity with other tools than the ones developed here, are more informed on some of the most established measures of state capacity, their correct use, and their advantages and disadvantages.

In the first chapter, I reviewed some of the most common definitions of state capacity and recently used measures of state capacity. In the second chapter, I selected some of the most established measures of state capacity for further evaluation. Various aspects related to the validity and interchangeability of the measures were compared and assessed. In the third chapter, I argued and showed that the research agenda on state capacity can take leaps forward by focusing on more specific attributes of the concept. In the fourth chapter, I analysed the dimensionality of state capacity and developed a new multidimensional composite index of state capacity. Finally, the fifth and last chapter draws the conclusions of my study.

Even if the conceptual and theoretical debate around state capacity is important, this dissertation has focused primarily on the measurement of state capacity. I have not stayed completely away from some of the conceptual and theoretical issues but have tried to approach them through an empirical analysis of measures of state capacity. I have assumed that if we want to know more about state capacity, we need to know more about its measurement. Reaping the harvest of recent conceptual and theoretical advances on the topic is impossible if there are no advances in the measurement of state capacity. This is particularly true for the quantitative comparative study of state capacity. Comparative considerations of measures of state capacity with a strong empirical focus remain scarce. My dissertation contributes to tackle this major deficiency in the literature from a variety of perspectives.

Summary of Main Findings

In the first chapter of this study, I reviewed common definitions of state capacity and frequently used measures of state capacity. The main objective of Chapter I was to provide the reader, in general terms, a basic understanding of the theoretical foundations of the topic of the dissertation, to identify some broader research questions related to state capacity, to discuss the importance of the subsequent chapters of the dissertation, and to position the whole study within relevant literature on state capacity. Essentially, the first chapter is an introduction to the topic of state capacity and the other parts of the dissertation.

In the second chapter of this study, I conducted a comparative assessment of some of the most established cross-national measures of state capacity. The main objective of Chapter II was to gain knowledge on the similarities and differences among the surveyed measures by focusing on aspects related to their validity as broadly understood. As to my

knowledge, at the time of writing, no such comprehensive comparative evaluation of measures of state capacity exists. Comparing measures of state capacity is particularly important because many researchers tend to assume that one measure or another will do more or less the same.

The findings of Chapter II are manifold. First of all, an analysis of the content validity of the surveyed measures provides new knowledge on the appropriateness of a particular measure to a particular research question. Some measures are more suitable than others to capture multidimensional definitions of the concept. In terms of content validity, Failed State Index (FSI), State Fragility Index (SFI), Corruption Perceptions Index (CPI), and Government Effectiveness (WGI) should be probably left to researchers focused on other concepts than state capacity, unless it is explicitly specified that state capacity is understood more broadly than we commonly do. Even in that case, researchers should be careful in using these indices, because they include some theoretically questionable sub-components, which undermine their analytical utility as measures of state capacity.

Second, an evaluation of the basic statistical properties of the measures shows that most measures have acceptable distributional characteristics and are fine-grained enough to distinguish between most observations. Overall, SFI seems to be the most problematic of the measures in terms of its basic statistical features. Third, through a traditional assessment of convergent validity, I find that all surveyed measures of state capacity are highly correlated among one another. An analysis of multivariate association with PCA suggests that all the surveyed measures are indeed quantifying the same construct. Nevertheless, an alternative approach to validity, or more precisely, an evaluation of the interchangeability of the measures shows that actually the measures are not at all

equivalent when employed in regressions. At worst, common measures of state capacity lead to opposing interpretations.

Last, the country-level scores of the measures are explored more comprehensively to gain knowledge on the differences and similarities across measures. Since we found the interchangeability of the surveyed measures to be weak, we thought that there must be some divergences in country scores across measures of state capacity. These disagreements were discovered, and they were further analysed. The findings suggest that for some countries the divergences are incredibly high, that country score divergency is strongest at intermediate levels of state capacity regardless of the measure, and that qualitatively the content of the measures reflects surprisingly well differences in country scores. If anything, since measures of state capacity are not interchangeable, the best practice seems to be to choose the measure that best represents a particular theory. The widespread idea that measures would lead to the same inferences given their strong association, however, should be forgotten.

In the third chapter of this study, I shifted the analysis towards a specific sub-component of state capacity: the impartiality of public officials. By focusing on the relationship between this specific attribute of state capacity and civic participation, my intention was to show that some of the measurement problems can be effectively resolved by digging deeper into the components of state capacity (and democracy). Thus, in terms of the measurement of state capacity, the primary objective of Chapter III was to analyse whether the possible mismatch between measures and concepts can be eliminated by focusing on specific aspects. As a by-product, however, Chapter III, had also another important objective: the actual analysis of the relationship between impartial public officials and civic participation around the world.

The findings of Chapter III do provide, indeed, persuasive evidence that the research agenda on state capacity can be pushed forward by focusing on specific aspect of the concept. Given that multidimensional measures often include components that undermine their analytical utility and/or are not relevant for common understandings of state capacity, the findings suggest that if research questions on the topic are narrowed to entail more specific aspects, it becomes clearer to understand what is being analysed. If the priority is to address the possible mismatch between definitions and measures, there seems to be no doubt that a "disaggregated" approach to state capacity leads to more fruitful results than a broad general-level approach.

As already said, the analysis of Chapter III was deliberately framed around one of the most pressing research topics related to state capacity: the state-democracy nexus, which was already analysed implicitly through the replications in Chapter II. Hence, my findings provide also novel information regarding the relationship between impartial public officials and civic participation. The findings support persuasively the hypothesised effect of civic participation on the impartiality of public officials. According to the data, analysed mainly through a battery of dynamic panel regressions, a society in which people are more actively involved in public affairs discourages public officials to behave partially. The finding is robust to several models, specifications, and estimators, to different measures, and is consistent both at the short run and at the long run, although its strength seems to depend on the macro-region.

In the fourth chapter of this study, instead, I focus on state capacity as a multidimensional concept and develop a multidimensional tool of measurement. Moreover, the selected research strategy allows me to determine the dimensionality of state capacity that best reflects the selected data. By taking into account the problems of some of the most frequently used measures of state capacity identified in Chapter II, the

primary objective of Chapter IV is to create a novel composite index of state capacity that addresses these issues. The validity of the newly developed index is evaluated with different methods from different "traditions" of measurement validation.

The findings of Chapter IV suggest that the novel tool of measurement is valid and addresses some of the problems identified in Chapter II. The content validity of the composite measure seems to be high. While the other frequently used measures of state capacity, with the exception of Hanson and Sigman's (2013) index, are not based on any theory on state capacity, my novel measure is indeed based on theory. The selected sub-components tap into different functional dimensions of state capacity that were previously identified. Anyhow, the exploratory analysis suggests as well that the dimensionality of state capacity does not follow a functional distinction, as argued by most scholars working on the topic. Instead, overall state capacity should be divided between power and procedural constraints, confirming some recent arguments put forward by some other researchers working on the topic.

Besides this major "result" of Chapter IV, the validity tests show some interesting insights regarding state capacity and its two intermediate sub-components in different countries. For instance, Singapore emerges as a country that has top-class procedural constraints but is less virtuous when it comes to the power dimension. Finally, in Chapter IV I replicate Hanson's (2015) study on state capacity, democracy, and development outcomes, to to validate CAP and to examine whether its two sub-components behave similarly or differently when interacted with democracy to predict infant mortality. The re-analysis of Hanson (2015) with Power and Procedures suggests that the "substituting" relationship between state capacity and democracy in affecting development outcomes is driven primarily by power rather than procedures.

Broader Policy Implications

The findings of my study are of crucial importance for comparative social scientists studying the state. By taking into account the results presented in this dissertation, researchers working on the topic have gained knowledge on existing measures of state capacity, have seen that different solutions – from a disaggregated approach to the construction of a multidimensional composite index – can be adopted to deal with measurement problems related to state capacity. In particular, the composite index I have created in this study can be used by researchers working on the topic to test their theories and to analyse statistically state capacity (and its two core components) across countries and over time. Nevertheless, the findings of this dissertation are not relevant only for academics, but also for policymakers and in particular for the international development community.

First of all, as already mentioned, achieving capable state institutions is a top priority of the world's major international institutions and one of the objectives of the United Nation's 2030 Agenda through its Sustainable Development Goals. Therefore, acquiring knowledge on the measurement of state capacity is of interest also for policymakers, because they use measures of state capacity in making decisions. The results of Chapter II suggest that some of the most established measures should be preferred to others, depending on the "content universe" one aspires to analyse. Moreover, Chapter II provides indications on which of these measures should be used and for what purpose.

Another topical issue for the international development community and development policies in general is state-building in the Global South. Since we know that capable states are crucial for socio-economic development, there have been many efforts to create well-functioning state institutions in countries where such institutions were inexistent or weak. Perhaps also due to the inconclusiveness of studies on the state-

democracy nexus, there have been different understandings about how to create these well-functioning state institutions in practice and how to sustain them over time. In general, the most widely adopted current practice seems to be that state capacity is supported by state capacity. It appears that the once vigorous agenda of exporting democracy around the world has lost its moment. In fact, the idea that too much democracy before a building a strong state can be even harmful has taken ground in developing contexts around the world.

Nevertheless, the findings in Chapter III suggest that it could be a good idea to clean the dust from democracy-building efforts. We have found convincing evidence that civic participation plays a key role in building impartial bureaucracies around the world. Given that a strong civic society is one of the core characteristics of well-functioning democracies and that impartiality in the exercise of public power is one of the core characteristics of well-functioning states, we have found at least one mechanism through which democracy can help state-building efforts in developing countries. Hence, to achieve the goal of building capable states, policymakers working in international development should turn their attention to the empowerment of the civil society. According to my results, a good way to tackle dysfunctional state apparatuses is to strengthen civil society, and in particular, to encourage civic participation into public affairs. Development efforts should, thus, focus on providing the means to increase the participation in the public sphere in underdeveloped countries and through that develop a more impartial bureaucracy. Since we know that a well-functioning bureaucracy is probably the most important aspect of a modern capable state, such efforts would be likely to enhance state capacity in general as well.

To analyse and evaluate the level of state capacity in countries around the world or to assess the progress in reaching the targets of the SDG's, policymakers can be guided

by the tool of measurement developed in Chapter IV. Obviously, the composite index of state capacity (*CAP*), and its two sub-dimensions (*Power* and *Procedures*), are most appropriate for general-level analyses on the state. They cannot and should not be used for detailed accounts on very specific aspects of the state but are intended to provide a broad picture on the state of state capacity, its power, and its procedures. If users are interested in analysing and evaluating more specific more aspects of the state, they should turn to one of the sub-indicators employed to construct the broader measures. An illustrative example of such more detailed analysis can be found in Chapter III of the study at hand.

Some scholars have criticised the inability of composite indicators to provide detailed information on the state (e.g., Trapnell 2011; Andrews, Hay and Myers 2010: 397), but others have stressed their importance also for policymakers (e.g., Gisselquist 2014). At the end of the day, using broad indices seems to be inevitable and essential also for policymakers, and even for a more general audience. As well put by Gisselquist (2014),

> *"Experts use indexes to facilitate cross-national comparisons and quantitative analysis, to explore trends over time, and to identify relationships for further study using other sources of information. For experts focused on particular countries, indexes complement – not substitute for – other types of assessments. They can further be useful tools for engagement with nonexpert audiences who lack the time or interest to engage more broadly."*

(Gisselquist 2014: 515)

The novel composite index, with its two sub-components, is well suited precisely for the above described usage of the tools. Finally, compared to many other frequently used composite measures of state capacity, the users of my indices can find helpful its theoretical roots, its analytical utility, and its relatively consistent time dimensions, which allows users of the indices presented in this study to make interesting comparisons within a country over time and across countries synchronically.

Potential Research Avenues for Future Studies

By now we have seen that the findings of this dissertation are manifold. Each of the chapters has played a part in generating knowledge on state capacity, and in particular its measurement. Furthermore, since the state-democracy nexus was used as a "frame" for the rest of the analysis throughout my study, I have addressed also important questions regarding the relationship between state capacity and democracy. Besides the already discussed direct conclusions and the most evident policy implications, the findings of my dissertation open up many interesting avenues for future research in various social science fields and sub-field. Many of the findings provide fruitful insight towards issues that have not been analysed more thoroughly in the study at hand for the simple reason that they are beyond the intended scope of my dissertation.

The potential fallacy of classic convergence validity

A classic way to show the convergence between two measures has been correlation. Yet, as the study at hand has robustly demonstrated, even highly correlated measures can in fact rate countries in surprisingly different ways and can have a significant impact on statistical inferences. Convergence does not imply equivalence nor interchangeability.

My results, thus, question the credibility of the most traditional tool of convergent validation, namely correlation analysis, as the one and only method to assess convergence. While bivariate correlations can effectively show the general relationship between two measures, researchers should avoid assuming that if two similar indicators are highly correlated picking one or another does not matter. The fallacy of this argument has been shown persuasively by the results in Chapter II. The findings of my study, instead, support the argument raised by Seawright and Collier (2014), according to whom if scholars aim for causal inference, they might be better off by testing the interchangeability of two similar indicators by regressing them as dependent variables on an identical set of predictors. To put it simply, researchers who are interested in testing the robustness of their results to several measures of the same concept, should not simply conclude that the one-fits-for-all approach works well when measures are correlated, but should instead, demonstrate that the measures lead to similar inferences.

Furthermore, the findings in Chapter II suggest that to grasp more detailed differences between measures, researchers should analyse and compare different cases through individual country scores. Since we have seen that convergent measures can rate countries with a substantial degree of disagreement, we are left wondering whether two measures can really be considered as equivalent, based solely on convergence validation. This is also why in validating the newly developed index of state capacity I have focused more on case-based validation approaches than mere correlations. If correlations and actual hypothesis testing through interchangeability provide the general level picture on the validity of a particular measure, by looking at individual country we can understand more nuanced differences among the measures.

Further studies evaluating best practices for different research situations are required to analyse more thoroughly the limitations of different tools of measurement validation.

The importance of the neglected practice of replication

In natural science replication of previous research is common. In the social sciences, by contrast, the vast majority of quantitative studies never gets replicated. In fact, by looking at the number of replication studies published in respectable social science journals, it seems that the social science community does not think much of replications. Why? According to Schmidt (2009: 95) the most important reason is precisely "that they are not acknowledged". Anyhow, the low esteem of replications does not mean that they are useless. As well put by King (1995: 451), "good science requires that we be able to reproduce existing numerical results, and that other scholars be able to show how substantive findings change as we apply the same methods in new contexts". The findings of the study at hand suggest that neglecting the replication of existing research undermines the creation of knowledge.

In this dissertation I have replicated various previous studies on the state-democracy nexus. These replications have shown that building up on existing research is one solution to push forward the research agenda on the topic. In Chapter II we found that some of the most influential empirical findings on the curvilinear state-democracy relationship are not robust to many common measures of state capacity. In Chapter III we found that some of the earlier conclusions on the topic are not valid if analysed only through specific aspects of state capacity and democracy. In Chapter IV we found that the novel measures of state capacity provided similar, although more nuanced results than in the original (replicated) study on the state-democracy relationship in relation to infant mortality.

These findings provide evidence that the replication of previous studies should not be neglected in the social sciences. On the contrary, besides testing the robustness of previous results, they can be used as a starting point for building new original knowledge.

Future studies on the state-democracy nexus are invited to replicate the statistical parts of the study at hand. Only in such way, it is possible to verify the generalisability of the results outside the setting presented here. The development of a common framework for the replication of previous social science studies might be a fruitful avenue for future research in social science methodology. Last, to stress the importance of replications in science, it might be useful to recall Popper's evergreen assertion on the issue:

"We do not take even our own observations quite seriously, or accept them as scientific observations, until we have repeated and tested them. Only by such repetitions can we convince ourselves that we are not dealing with a mere isolated 'coincidence,' but with events which, on account of their regularity and reproducibility, are in principle inter-subjectively testable."

(Popper 1959/2005: 23)

The problem of systematic rating discrepancy

We found in Chapter II that some of the most established measures of state capacity disagree on what is the true level of state capacity many countries. Nevertheless, as to this "rating discrepancy", measures seem to agree at least on one pattern: the disagreement in country scores is high at intermediate levels of state capacity but low at the two extremes of the scale. The study at hand is not by any means the first one to analyse systematic disagreement in social science measures of a particular concept. Goertz (2020) uses the term "gray zone" to indicate the conceptual area between two ideal types, which often coincides with measurement disagreement. In the case of perception-based measures of state capacity, it is understandable that disagreement across measures is strongest at intermediate levels. Since these cases are not high capacity states nor highly

dysfunctional states, they are simply harder to code. Anyhow, more detailed studies on the causes of this systematic disagreement are needed. Such studies are likely to provide valuable contributions to the state capacity literature.

Chapter II analysed thoroughly only measures that included subjective data. Future research might find interesting to analyse whether systematic disagreement exists also in different measures based on objective data. While it may seem surprising, studies have shown that substantial disagreement among measures of the same concept exists also in objective "facts-based" data. For instance, Jerven (2013) compares GDP/capita estimates from different databases and finds that the estimates disagree considerably among one another for some countries and that such disagreement is particularly large in Africa. Could there be an analogous pattern of systematic discrepancy among objective measures of state capacity as well?

On the state-democracy nexus

The relationship between state capacity and democracy has been a recurring theme in this dissertation. Using the state-democracy nexus as a frame for the other parts of my study implied the assumption that state capacity and democracy are not the same. Otherwise we would not have been able to gain any knowledge on the relationship between the two. Overall, my findings provide interesting insights on the state-democracy nexus and point out possibly rewarding routes for further research on the "sequencing" literature.

Many scholars have accepted the theory according to which the relationship between democracy and state capacity depends on the level of democracy. Some of the most influential studies on the topic, replicated in Chapter II, have argued that at low levels of democracy the relationship between democracy and state capacity is negative, whereas at high levels of democracy the relationship between democracy and state capacity is

positive. My findings have shown that this general-level argument is not valid across many commonly used measures of state capacity. Future research should continue analysing the relationship between state capacity and democracy, without assuming that it is necessarily conditional on the level of democracy.

In fact, the results in Chapter III indicate persuasively that the relationship between two of the core aspects of state capacity and democracy – respectively impartial public officials and civic participation – is positive, linear, and does not depend on the level of democracy (nor on the level of economic development). To be more specific, the findings in Chapter III show that civic participation increases the impartial behaviour of public officials. While some scholars (Andersen and Doucette 2020) have advocated and engaged in analysing the relationship between the component parts of the state-democracy nexus, literature on the disaggregated relationship between state capacity and democracy is at very initial stages. The formulation of theories and the testing of these theories regarding how different components of state capacity and democracy are related among one another is likely to be a top priority on the "sequencing" agenda in the years to come. Likewise, the causal direction of these different potential accounts needs to be determined, although mutual and synchronic causation should not be excluded.

The brief attempt to replicate Hanson's (2015) study on how the interaction between state capacity and democracy is related to development outcomes, backs up previous studies that have corroborated a substituting rather than a mutually reinforcing effect of the two. As before, a disaggregation of the core aspects of state capacity and democracy, both in terms of measures and in terms of theory, is likely to provide fruitful avenues for future research. In the study at hand, with the newly developed measures, I have conducted a general-level disaggregation of state capacity into its two core components – power and procedural constraints – and used them to further explore the above

interactive relationship. My findings, which suggest that the "interactive" effect depends on the chosen component of state capacity, confirm the advantages of a disaggregated approach. These differences between how the two components interact with democracy should be further analysed as well. Indeed, in general, my findings show that we are still in the early days of understanding comprehensively the relationship between state capacity and democracy.

On the conceptual dimensions of state capacity

While we have seen that nowadays most scholars agree on the background concept of state capacity, when it comes to the dimensions of the concept, expert opinions get less congruent among one another. For now, the dimensions of state capacity have been mainly classified through functional perspectives on the issue. Differences among scholarly opinions may not be large but are important when we turn into the measurement of state capacity, because most efforts to capture state capacity in its entirety have relied in one way or another on these functional dimensions.

Recently, however, researchers have started advocating the conception of state capacity not so much by focusing on the core functions of the state, but instead by focusing on "the ability to implement any political decision" (D'Arcy and Nistotskaya 2021: 4), or in other words "to the state's ability to 'get things done'" (Lindvall and Teorell 2016: 5). Within this "alternative" conceptual framework researchers have identified especially two dimensions of a broader concept of state capacity: power and procedural constraints (Fukuyama 2013; Bardhan 2016; Lindvall and Teorell 2016; Johnson and Koyama 2017; D'Arcy and Nistotskaya 2021). While scholars disagree on whether state capacity should refer to both dimensions or only the former one, my findings have suggested that statistically power and procedural constraints can be

considered to be the most important dimension of a more general concept of state capacity.

My results indicate that future theory-building on the topic could find it useful to put aside for a moment the functional aspects of the state and focus more decisively on the two dimensions recently conceived by some scholars and statistically confirmed by the study at hand. As a consequence, creating measures of state capacity that capture these two dimensions could prove more useful to testing theories on the topic. The measure(s) developed Chapter IV should be considered as an initial step in such measurement-creation efforts.

Appendix A

Table A1. Correlations between FSI and other measures of state capacity over time

Year	QOG	HSI	SFI	WGI	CPI	VDEM
2005	0.88 (124)	0.84 (145)	0.87 (145)	0.92 (145)	0.88 (138)	0.82 (145)
2006	0.87 (137)	0.85 (160)	0.85 (163)	0.89 (176)	0.89 (158)	0.81 (169)
2007	0.86 (137)	0.85 (160)	0.87 (163)	0.89 (176)	0.90 (170)	0.81 (169)
2008	0.86 (137)	0.84 (160)	0.87 (163)	0.89 (176)	0.90 (169)	0.81 (169)
2009	0.86 (137)	0.85 (160)	0.88 (163)	0.90 (176)	0.89 (169)	0.81 (169)
2010	0.87 (137)		0.89 (163)	0.90 (176)	0.88 (168)	0.81 (169)
2011	0.87 (137)		0.89 (163)	0.90 (176)	0.87 (171)	0.81 (170)
2012	0.88 (137)		0.89 (164)	0.91 (177)	0.87 (168)	0.81 (170)
2013	0.87 (137)		0.90 (164)	0.91 (177)	0.87 (169)	0.79 (170)
2014	0.87 (137)		0.90 (164)	0.91 (177)	0.88 (167)	0.80 (170)
2015	0.88 (137)		0.90 (164)	0.92 (177)	0.89 (164)	0.80 (170)
2016	0.89 (137)		0.89 (164)	0.92 (177)	0.90 (169)	0.79 (170)
2017	0.89 (137)		0.89 (164)	0.91 (177)	0.89 (172)	0.79 (170)

Note: Pearson's correlation coefficients; number of observations in parentheses; all coefficients significant at the $p < 0.001$ level.

Table A2. Correlations between QOG and other measures of state capacity over time

Year	FSI	HSI	SFI	WGI	CPI	VDEM
1995		0.84 (122)	0.83 (124)		0.82 (40)	0.75 (126)
1996		0.85 (122)	0.82 (124)	0.88 (128)	0.90 (53)	0.77 (126)
1997		0.86 (122)	0.84 (124)		0.93 (51)	0.80 (126)
1998		0.87 (124)	0.84 (126)	0.92 (130)	0.91 (81)	0.81 (128)
1999		0.86 (133)	0.81 (135)		0.91 (93)	0.80 (137)
2000		0.84 (133)	0.82 (135)	0.92 (139)	0.91 (87)	0.79 (137)
2001		0.84 (133)	0.80 (135)		0.92 (88)	0.79 (137)
2002		0.86 (133)	0.78 (135)	0.92 (139)	0.92 (98)	0.79 (137)
2003		0.84 (133)	0.77 (135)	0.92 (139)	0.91 (123)	0.80 (137)
2004		0.85 (133)	0.79 (135)	0.93 (139)	0.91 (129)	0.80 (137)
2005	0.88 (124)	0.87 (133)	0.79 (135)	0.93 (139)	0.91 (133)	0.81 (137)
2006	0.87 (137)	0.88 (133)	0.77 (135)	0.93 (139)	0.91 (133)	0.80 (137)
2007	0.86 (137)	0.88 (133)	0.76 (135)	0.93 (139)	0.91 (136)	0.79 (137)
2008	0.86 (137)	0.85 (133)	0.74 (135)	0.93 (139)	0.91 (136)	0.79 (137)
2009	0.86 (137)	0.84 (133)	0.75 (135)	0.93 (139)	0.91 (137)	0.79 (137)
2010	0.87 (137)		0.75 (135)	0.94 (139)	0.91 (136)	0.80 (137)
2011	0.87 (137)		0.75 (135)	0.94 (139)	0.92 (139)	0.79 (137)
2012	0.88 (137)		0.75 (135)	0.94 (139)	0.92 (139)	0.80 (137)
2013	0.87 (137)		0.74 (135)	0.93 (139)	0.92 (139)	0.78 (137)
2014	0.87 (137)		0.73 (135)	0.93 (139)	0.92 (138)	0.78 (137)
2015	0.88 (137)		0.75 (135)	0.93 (139)	0.93 (137)	0.78 (137)
2016	0.89 (137)		0.75 (135)	0.93 (139)	0.93 (139)	0.78 (137)
2017	0.89 (137)		0.75 (135)	0.93 (139)	0.93 (139)	0.78 (137)

Note: Pearson's correlation coefficients; number of observations in parentheses; all coefficients significant at the $p < 0.001$ level.

Table A3. Correlations between HSI and other measures of state capacity over time

Year	FSI	QOG	SFI	WGI	CPI	VDEM
1995		0.84 (122)	0.83 (160)		0.79 (40)	0.71 (160)
1996		0.85 (122)	0.82 (160)	0.84 (159)	0.86 (53)	0.71 (160)
1997		0.86 (122)	0.82 (160)		0.89 (50)	0.72 (160)
1998		0.87 (124)	0.83 (160)	0.87 (160)	0.87 (82)	0.72 (160)
1999		0.86 (133)	0.83 (160)		0.88 (96)	0.73 (160)
2000		0.84 (133)	0.84 (160)	0.85 (160)	0.83 (87)	0.71 (160)
2001		0.84 (133)	0.84 (160)		0.89 (88)	0.71 (160)
2002		0.86 (133)	0.83 (161)	0.87 (161)	0.88 (99)	0.68 (161)
2003		0.84 (133)	0.83 (161)	0.88 (161)	0.87 (128)	0.67 (161)
2004		0.85 (133)	0.83 (161)	0.89 (161)	0.85 (137)	0.69 (161)
2005	0.84 (145)	0.87 (133)	0.83 (161)	0.89 (161)	0.83 (150)	0.73 (161)
2006	0.85 (160)	0.88 (133)	0.83 (162)	0.90 (162)	0.85 (152)	0.72 (162)
2007	0.85 (160)	0.88 (133)	0.83 (162)	0.90 (162)	0.83 (160)	0.72 (162)
2008	0.84 (160)	0.85 (133)	0.83 (162)	0.90 (162)	0.84 (160)	0.70 (162)
2009	0.85 (160)	0.84 (133)	0.83 (162)	0.90 (162)	0.83 (160)	0.71 (162)

Note: Pearson's correlation coefficients; number of observations in parentheses; all coefficients significant at the $p < 0.001$ level.

Table A4. Correlations between SFI and other measures of state capacity over time

Year	FSI	QOG	HSI	WGI	CPI	VDEM
1995		0.83 (124)	0.83 (160)		0.73 (40)	0.76 (163)
1996		0.82 (124)	0.82 (160)	0.82 (161)	0.80 (53)	0.76 (163)
1997		0.84 (124)	0.82 (160)		0.77 (51)	0.77 (163)
1998		0.84 (126)	0.83 (160)	0.82 (163)	0.73 (83)	0.77 (163)
1999		0.81 (135)	0.83 (160)		0.73 (97)	0.76 (163)
2000		0.82 (135)	0.84 (160)	0.82 (163)	0.74 (88)	0.75 (163)
2001		0.80 (135)	0.84 (160)		0.75 (89)	0.73 (163)
2002		0.78 (135)	0.83 (161)	0.82 (164)	0.74 (100)	0.72 (164)
2003		0.77 (135)	0.83 (161)	0.82 (164)	0.75 (129)	0.71 (164)
2004		0.79 (135)	0.83 (161)	0.83 (164)	0.75 (139)	0.71 (164)
2005	0.87 (145)	0.79 (135)	0.83 (161)	0.83 (164)	0.73 (152)	0.71 (164)
2006	0.85 (163)	0.77 (135)	0.83 (162)	0.82 (165)	0.73 (154)	0.70 (165)
2007	0.87 (163)	0.76 (135)	0.83 (162)	0.80 (165)	0.76 (163)	0.70 (165)
2008	0.87 (163)	0.74 (135)	0.83 (162)	0.80 (166)	0.76 (163)	0.69 (166)
2009	0.88 (163)	0.75 (135)	0.83 (162)	0.81 (166)	0.76 (163)	0.70 (166)
2010	0.89 (163)	0.75 (135)		0.80 (166)	0.73 (163)	0.68 (166)
2011	0.89 (163)	0.75 (135)		0.81 (166)	0.72 (165)	0.69 (166)
2012	0.89 (164)	0.75 (135)		0.81 (167)	0.73 (164)	0.68 (167)
2013	0.90 (164)	0.74 (135)		0.81 (167)	0.74 (165)	0.69 (167)
2014	0.90 (164)	0.73 (135)		0.83 (167)	0.74 (164)	0.68 (167)
2015	0.90 (164)	0.75 (135)		0.83 (167)	0.75 (163)	0.69 (167)
2016	0.89 (164)	0.75 (135)		0.83 (167)	0.75 (164)	0.70 (167)
2017	0.89 (164)	0.75 (135)		0.81 (167)	0.75 (166)	0.69 (167)

Note: Pearson's correlation coefficients; number of observations in parentheses; all coefficients significant at the $p < 0.001$ level.

Table A5. Correlations between WGI and other measures of state capacity over time

Year	FSI	QOG	HSI	SFI	CPI	VDEM
1996		0.88 (128)	0.84 (159)	0.82 (161)	0.94 (53)	0.83 (168)
1998		0.92 (130)	0.87 (160)	0.82 (163)	0.93 (84)	0.83 (171)
2000		0.92 (139)	0.85 (160)	0.82 (163)	0.94 (89)	0.85 (171)
2002		0.92 (139)	0.87 (161)	0.82 (164)	0.92 (101)	0.83 (172)
2003		0.92 (139)	0.88 (161)	0.82 (164)	0.93 (131)	0.83 (172)
2004		0.93 (139)	0.89 (161)	0.83 (164)	0.94 (144)	0.84 (172)
2005	0.92 (145)	0.93 (139)	0.89 (161)	0.83 (164)	0.93 (157)	0.84 (173)
2006	0.89 (176)	0.93 (139)	0.90 (162)	0.82 (165)	0.93 (161)	0.83 (174)
2007	0.89 (176)	0.93 (139)	0.90 (162)	0.80 (165)	0.93 (178)	0.83 (174)
2008	0.89 (176)	0.93 (139)	0.90 (162)	0.80 (166)	0.94 (177)	0.83 (174)
2009	0.90 (176)	0.93 (139)	0.90 (162)	0.81 (166)	0.94 (177)	0.84 (174)
2010	0.90 (176)	0.94 (139)		0.80 (166)	0.93 (175)	0.84 (174)
2011	0.90 (176)	0.94 (139)		0.81 (166)	0.93 (180)	0.83 (175)
2012	0.91 (177)	0.94 (139)		0.81 (167)	0.94 (174)	0.83 (175)
2013	0.91 (177)	0.93 (139)		0.81 (167)	0.94 (175)	0.82 (175)
2014	0.91 (177)	0.93 (139)		0.83 (167)	0.92 (173)	0.80 (175)
2015	0.92 (177)	0.93 (139)		0.83 (167)	0.93 (167)	0.80 (175)
2016	0.92 (177)	0.93 (139)		0.83 (167)	0.93 (175)	0.80 (175)
2017	0.91 (177)	0.93 (139)		0.81 (167)	0.92 (179)	0.79 (175)

Note: Pearson's correlation coefficients; number of observations in parentheses; all coefficients significant at the $p < 0.001$ level.

Table A6. Correlations between CPI and other measures of state capacity over time

Year	FSI	QOG	HSI	SFI	WGI	VDEM
1995		0.82 (40)	0.79 (40)	0.73 (40)		0.81 (40)
1996		0.90 (53)	0.86 (53)	0.80 (53)	0.94 (53)	0.87 (53)
1997		0.93 (51)	0.89 (50)	0.77 (51)		0.85 (51)
1998		0.91 (81)	0.87 (82)	0.73 (83)	0.93 (84)	0.81 (84)
1999		0.91 (93)	0.88 (96)	0.73 (97)		0.82 (98)
2000		0.91 (87)	0.83 (87)	0.74 (88)	0.94 (89)	0.82 (89)
2001		0.92 (88)	0.89 (88)	0.75 (89)		0.82 (90)
2002		0.92 (98)	0.88 (99)	0.74 (100)	0.92 (101)	0.82 (101)
2003		0.91 (123)	0.87 (128)	0.75 (129)	0.93 (131)	0.81 (130)
2004		0.91 (129)	0.85 (137)	0.75 (139)	0.94 (144)	0.81 (143)
2005	0.88 (138)	0.91 (133)	0.83 (150)	0.73 (152)	0.93 (157)	0.82 (156)
2006	0.89 (158)	0.91 (133)	0.85 (152)	0.73 (154)	0.93 (161)	0.82 (158)
2007	0.90 (170)	0.91 (136)	0.83 (160)	0.76 (163)	0.93 (178)	0.84 (170)
2008	0.90 (169)	0.91 (136)	0.84 (160)	0.76 (163)	0.94 (177)	0.84 (170)
2009	0.89 (169)	0.91 (137)	0.83 (160)	0.76 (163)	0.94 (177)	0.84 (170)
2010	0.88 (168)	0.91 (136)		0.73 (163)	0.93 (175)	0.84 (170)
2011	0.87 (171)	0.92 (139)		0.72 (165)	0.93 (180)	0.83 (172)
2012	0.87 (168)	0.92 (139)		0.73 (164)	0.94 (174)	0.86 (169)
2013	0.87 (169)	0.92 (139)		0.74 (165)	0.94 (175)	0.84 (170)
2014	0.88 (167)	0.92 (138)		0.74 (164)	0.92 (173)	0.85 (169)
2015	0.89 (164)	0.93 (137)		0.75 (163)	0.93 (167)	0.84 (167)
2016	0.90 (169)	0.93 (139)		0.75 (164)	0.93 (175)	0.86 (169)
2017	0.89 (172)	0.93 (139)		0.75 (166)	0.92 (179)	0.86 (173)

Note: Pearson's correlation coefficients; number of observations in parentheses; all coefficients significant at the $p < 0.001$ level.

Table A7. Correlations between VDEM and other measures of state capacity over time

Year	FSI	QOG	HSI	SFI	WGI	CPI
1995		0.75 (126)	0.71 (160)	0.76 (163)		0.81 (40)
1996		0.77 (126)	0.71 (160)	0.76 (163)	0.83 (168)	0.87 (53)
1997		0.80 (126)	0.72 (160)	0.77 (163)		0.85 (51)
1998		0.81 (128)	0.72 (160)	0.77 (163)	0.83 (171)	0.81 (84)
1999		0.80 (137)	0.73 (160)	0.76 (163)		0.82 (98)
2000		0.79 (137)	0.71 (160)	0.75 (163)	0.85 (171)	0.82 (89)
2001		0.79 (137)	0.71 (160)	0.73 (163)		0.82 (90)
2002		0.79 (137)	0.68 (161)	0.72 (164)	0.83 (172)	0.82 (101)
2003		0.80 (137)	0.67 (161)	0.71 (164)	0.83 (172)	0.81 (130)
2004		0.80 (137)	0.69 (161)	0.71 (164)	0.84 (172)	0.81 (143)
2005	0.82 (145)	0.81 (137)	0.73 (161)	0.71 (164)	0.84 (173)	0.82 (156)
2006	0.81 (169)	0.80 (137)	0.72 (162)	0.70 (165)	0.83 (174)	0.82 (158)
2007	0.81 (169)	0.79 (137)	0.72 (162)	0.70 (165)	0.83 (174)	0.84 (170)
2008	0.81 (169)	0.79 (137)	0.70 (162)	0.69 (166)	0.83 (174)	0.84 (170)
2009	0.81 (169)	0.79 (137)	0.71 (162)	0.70 (166)	0.84 (174)	0.84 (170)
2010	0.81 (169)	0.80 (137)		0.68 (166)	0.84 (174)	0.84 (170)
2011	0.81 (170)	0.79 (137)		0.69 (166)	0.83 (175)	0.83 (172)
2012	0.81 (170)	0.80 (137)		0.68 (167)	0.83 (175)	0.86 (169)
2013	0.79 (170)	0.78 (137)		0.69 (167)	0.82 (175)	0.84 (170)
2014	0.80 (170)	0.78 (137)		0.68 (167)	0.80 (175)	0.85 (169)
2015	0.80 (170)	0.78 (137)		0.69 (167)	0.80 (175)	0.84 (167)
2016	0.79 (170)	0.78 (137)		0.70 (167)	0.80 (175)	0.86 (169)
2017	0.79 (170)	0.78 (137)		0.69 (167)	0.79 (175)	0.86 (173)

Note: Pearson's correlation coefficients; number of observations in parentheses; all coefficients significant at the $p < 0.001$ level.

Table A8. Principal component analysis of measures of state capacity (1996-2009)

Component	Eigenvalue	% of explained variance	Cumulative % of explained variance
1	5.168	86.13	86.13
2	0.314	5.23	91.36
3	0.266	4.43	95.79
4	0.111	1.84	97.63
5	0.092	1.54	99.17
6	0.050	0.83	100.00

Note: FSI not included.

Table A9. Principal component analysis of measures of state capacity (2005-2017)

Component	Eigenvalue	% of explained variance	Cumulative % of explained variance
1	5.239	87.32	87.32
2	0.328	5.47	92.79
3	0.236	3.94	96.73
4	0.086	1.43	98.16
5	0.064	1.07	99.23
6	0.046	0.77	100.00

Note: HSI not included.

Table A10. Summary statistics of variables used to replicate Bäck and Hadenius (2008)

Variable		Mean	SD	Min	Max	Observations
BHI	Overall	0.556	0.241	0.000	1.000	N = 2006
	Between		0.220	0.113	0.997	n = 132
	Within		0.095	0.199	0.894	T-bar = 15.197
QOG	Overall	0.558	0.241	0.014	1.000	N = 2006
	Between		0.218	0.093	0.998	n = 132
	Within		0.091	0.232	0.886	T-bar = 15.450
HSI	Overall	0.564	0.165	0.161	1.000	N = 2317
	Between		0.157	0.241	0.889	n = 150
	Within		0.040	0.327	0.721	T-bar = 15.447
VDEM	Overall	0.536	0.198	0.087	1.000	N = 2443
	Between		0.190	0.144	1.000	n = 158
	Within		0.054	0.180	0.767	T-bar = 15.462
Democracy	Overall	5.810	3.333	0.000	10.000	N = 2472
	Between		3.108	0.162	10.000	n = 160
	Within		1.310	0.273	10.305	T-bar = 15.450
Democracy^2	Overall	44.858	37.064	0.000	100.000	N = 2472
	Between		34.955	0.063	100.000	n = 160
	Within		12.870	-10.032	91.837	T-bar = 15.450
Ln(GDP/capita)	Overall	8.130	1.543	4.880	11.485	N = 2472
	Between		1.539	5.256	11.179	n = 160
	Within		0.174	7.127	10.135	T-bar = 15.450
Trade	Overall	74.256	49.195	0.021	531.737	N = 2468
	Between		43.993	0.714	331.846	n = 160
	Within		17.977	-64.238	314.437	T-bar = 15.425
British colony	Overall	0.266	0.442	0.000	1.000	N = 2472
	Between		0.445	0.000	1.000	n = 160
	Within		0.000	0.266	0.266	T-bar = 15.450

Table A11. Summary statistics of variables used to replicate Carbone and Memoli (2015)

Variable		Mean	SD	Min	Max	Observations
CMI	Overall	0.528	0.266	0.010	1.000	N = 344
	Between		0.261	0.010	1.000	n = 122
	Within		0.059	0.319	0.824	T-bar = 2.820
FSI	Overall	0.378	0.182	0.007	0.837	N = 344
	Between		0.182	0.032	0.827	n = 121
	Within		0.019	0.328	0.447	T-bar = 2.843
QOG	Overall	0.406	0.140	0.030	0.848	N = 284
	Between		0.139	0.040	0.848	n = 99
	Within		0.016	0.346	0.472	T-bar = 2.869
HSI	Overall	0.540	0.142	0.179	1.000	N = 344
	Between		0.143	0.218	0.960	n = 121
	Within		0.019	0.468	0.604	T-bar = 2.843
SFI	Overall	0.607	0.239	0.040	1.000	N = 345
	Between		0.237	0.053	1.000	n = 122
	Within		0.029	0.487	0.687	T-bar = 2.828
WGI	Overall	0.424	0.151	0.138	1.000	N = 345
	Between		0.149	0.151	0.971	n = 122
	Within		0.017	0.349	0.477	T-bar = 2.828
CPI	Overall	0.266	0.155	0.047	0.977	N = 343
	Between		0.154	0.058	0.965	n = 122
	Within		0.028	0.169	0.367	T-bar = 2.811
VDEM	Overall	0.372	0.159	0.000	0.769	N = 345
	Between		0.157	0.026	0.751	n = 122
	Within		0.024	0.204	0.471	T-bar = 2.828
Democracy	Overall	-0.086	0.935	-1.976	1.097	N = 345
	Between		0.942	-1.976	1.097	n = 122
	Within		0.123	-1.016	0.466	T-bar = 2.828
Democracy^2	Overall	0.878	0.822	0.003	3.903	N = 345
	Between		0.840	0.004	3.903	n = 122
	Within		0.136	0.069	1.823	T-bar = 2.828
Duration of democracy	Overall	8.542	14.142	0.000	91.000	N = 345
	Between		13.537	0.000	89.000	n = 122
	Within		3.171	-30.791	29.209	T-bar = 2.828
Duration of democracy^2	Overall	272.386	902.675	0.000	8281	N = 345
	Between		865.963	0.000	7923.667	n = 122
	Within		165.408	-2048.948	1551.052	T-bar = 2.828
Democracy × duration of democracy	Overall	6.866	13.141	0.000	99.819	N = 345
	Between		12.795	0.000	97.625	n = 122
	Within		1.538	-0.349	19.602	T-bar = 2.828
Ethnic fractionalization	Overall	0.488	0.244	0.000	0.930	N = 345
	Between		0.248	0.000	0.930	n = 122
	Within		0.000	0.488	0.488	T-bar = 2.828
Log(GDP/capita)	Overall	7.753	1.317	4.812	10.975	N = 345
	Between		1.331	5.020	10.975	n = 122
	Within		0.181	7.225	8.088	T-bar = 2.828
Log(Area)	Overall	12.340	1.741	6.537	16.611	N = 345
	Between		1.771	6.546	16.611	n = 122
	Within		0.001	12.326	12.353	T-bar = 2.828

Table A12. Summary statistics of variables used to replicate Memoli and Grassi (2016)

Variable		Mean	SD	Min	Max	Observations	
GMI	Overall	0.324	0.194	0.000	1.000	N =	197
	Between		0.188	0.105	0.900	n =	18
	Within		0.063	0.139	0.504	T-bar = 10.944	
QOG	Overall	0.449	0.194	0.000	1.000	N =	197
	Between		0.177	0.138	0.949	n =	18
	Within		0.089	0.240	0.723	T-bar = 10.944	
SFI	Overall	0.565	0.278	0.000	1.000	N =	197
	Between		0.274	0.164	1.000	n =	18
	Within		0.079	0.250	0.845	T-bar = 10.944	
WGI	Overall	0.383	0.231	0.000	1.000	N =	161
	Between		0.231	0.065	0.962	n =	18
	Within		0.047	0.272	0.524	T-bar = 8.944	
CPI	Overall	0.334	0.226	0.000	1.000	N =	186
	Between		0.221	0.083	0.951	n =	18
	Within		0.057	0.082	0.506	T-bar = 10.333	
VDEM	Overall	0.374	0.247	0.000	1.000	N =	197
	Between		0.247	0.115	0.949	n =	18
	Within		0.052	0.058	0.496	T-bar = 10.944	
Lagged stateness	Overall	0.027	0.252	-0.948	0.836	N =	197
	Between		0.115	-0.147	0.275	n =	18
	Within		0.225	-0.808	0.769	T-bar = 10.944	
Democracy	Overall	6.147	1.052	-2.000	7.000	N =	197
	Between		0.810	4.091	7.000	n =	18
	Within		0.692	0.056	8.147	T-bar = 10.944	
Democracy^2	Overall	38.888	10.058	1.000	49.000	N =	197
	Between		8.875	20.636	49.000	n =	18
	Within		5.103	-1.012	62.161	T-bar = 10.944	
Executive partisan balance	Overall	0.167	0.329	0.000	1.000	N =	197
	Between		0.320	0.000	1.000	n =	18
	Within		0.112	-0.474	0.537	T-bar = 10.944	
Level of economic development (1994)	Overall	7.870	0.629	6.724	8.920	N =	197
	Between		0.644	6.724	8.920	n =	18
	Within		0.000	7.870	7.870	T-bar = 10.944	
Land size (km^2)	Overall	12.608	1.768	9.290	15.951	N =	197
	Between		1.813	9.290	15.951	n =	18
	Within		0.000	12.608	12.608	T-bar = 10.944	
Oil rents	Overall	0.112	0.316	0.000	1.000	N =	197
	Between		0.323	0.000	1.000	n =	18
	Within		0.000	0.112	0.112	T-bar = 10.944	
Log(Ethnic fractionalization)	Overall	-0.923	0.472	-1.778	-0.302	N =	197
	Between		0.485	-1.778	-0.302	n =	18
	Within		0.000	-0.923	-0.923	T-bar = 10.944	

Table A13. Summary statistics of variables used to replicate Charron and Lapuente (2010)

Variable	Observations	Mean	SD	Min	Max
QOG	108	0.481	0.234	0.094	1.000
FSI	113	0.477	0.259	0.032	1.000
HSI	120	0.499	0.198	0.097	0.951
SFI	122	0.626	0.277	0.083	1.000
WGI	133	0.508	0.227	0.153	1.000
CPI	93	0.391	0.276	0.000	1.000
VDEM	126	0.511	0.217	0.120	1.000
CMI	89	0.531	0.261	0.051	1.000
Democracy	133	7.037	2.815	0.750	10.000
Democracy^2	133	57.380	34.766	0.563	100.000
Ln(GDP/capita)	133	8.606	1.126	6.306	10.791
Democracy × Ln(GDP/capita)	133	62.314	29.603	5.708	107.912
Trade openness	133	85.441	45.542	18.218	279.558
British colony	133	0.241	0.429	0.000	1.000

Table A14. Replication of Bäck and Hadenius (2008) and re-analysis with alternative measures of state capacity: robustness tests with common sample of country-years across models

	Replication (1)	QOG (2)	HSI (3)	VDEM (4)
Democracy	-0.004	-0.004	0.0000	0.003*
	(0.002)	(0.003)	(0.001)	(0.001)
Democracy^2	0.001*	0.001*	0.0001	-0.0002
	(0.0002)	(0.0003)	(0.0001)	(0.0001)
Ln(GDP/capita)	0.003	0.004	0.008***	0.004***
	(0.002)	(0.002)	(0.002)	(0.001)
Trade	0.000005	0.00002	0.00004*	-0.00001
	(0.00003)	(0.00003)	(0.00002)	(0.00001)
British colony	0.005	0.005	0.001	-0.003
	(0.003)	(0.003)	(0.001)	(0.002)
Lagged dependent variable	0.938***	0.937***	0.913***	0.958***
	(0.019)	(0.018)	(0.024)	(0.012)
Constant	0.004	0.003	-0.018	-0.014*
	(0.012)	(0.013)	(0.009)	(0.007)
R^2	0.96	0.97	0.97	0.98
Observations	1895	1895	1895	1895
Countries	126	126	126	126

Note: Pooled OLS models with panel-corrected standard errors in parentheses; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Model 1 is a replication of Model 3 in Table 1 in the original study; dependent variable in Model 1 is an additive index of two ICRG indicators (PRS Group): *Corruption* and *Bureaucracy Quality*. Models reproduce the specification and estimation methods of the original study. Independent variables are taken from the QoG Standard Dataset (Teorell et al. 2019). Democracy = fh_ipolity2. GDP/capita = wdi_gdpcapcon2010. Trade = wdi_trade. British colony = ht_colonial.

Table A15. Replication of Carbone and Memoli (2015) and re-analysis with alternative measures of state capacity: robustness tests with common sample of country-years across models

	Replication (1)	FSI (2)	QOG (3)	HSI (4)	SFI (5)	WGI (6)	CPI (7)	VDEM (8)
Democracy	0.103***	0.085***	0.038*	0.037***	0.085***	0.032*	0.035	0.076***
	(0.024)	(0.014)	(0.016)	(0.011)	(0.015)	(0.016)	(0.018)	(0.022)
Democracy^2	0.087***	0.062***	0.020*	0.036***	0.054***	0.022	0.019	0.034*
	(0.021)	(0.015)	(0.010)	(0.011)	(0.014)	(0.014)	(0.015)	(0.016)
Duration of democracy	-0.003	-0.0004	-0.001	0.0002	0.001	-0.0003	-0.0005	0.0002
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)
Duration of democracy^2	-0.000003	-0.00001	0.00001	-0.000001	-0.00002	0.00001	0.000004	-0.000003
	(0.00004)	(0.00001)	(0.00002)	(0.00001)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
Democracy × Duration of democracy	0.003	0.001	-0.0003	-0.001	-0.001	-0.0004	0.001	0.001
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Ethnic fractionalization	-0.066	-0.115*	-0.061	-0.098*	-0.189***	-0.092*	0.002	0.063
	(0.080)	(0.052)	(0.049)	(0.041)	(0.049)	(0.045)	(0.042)	(0.042)
Log(GDP/capita)	0.092***	0.030***	0.031***	0.047***	0.084***	0.035***	0.043***	0.031***
	(0.013)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)	(0.008)	(0.007)
Log(Area)	-0.025**	-0.018*	-0.007	-0.007	-0.023***	-0.017*	-0.027**	-0.011
	(0.009)	(0.008)	(0.009)	(0.007)	(0.006)	(0.009)	(0.010)	(0.007)
Constant	0.106	0.381**	0.271*	0.304**	0.289**	0.400***	0.254*	0.197*
	(0.190)	(0.125)	(0.120)	(0.103)	(0.107)	(0.119)	(0.128)	(0.094)
Sigma_u	0.16	0.09	0.11	0.07	0.11	0.10	0.11	0.10
Sigma_e	0.06	0.02	0.02	0.02	0.03	0.02	0.03	0.03
Rho	0.86	0.94	0.97	0.92	0.91	0.96	0.92	0.93
Between R^2	0.63	0.68	0.38	0.68	0.79	0.54	0.53	0.55
Wald chi-square (8)	201.24	154.91	72.02	220.09	485.22	118.87	70.34	88.25
Prob > chi-square	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Countries	99	99	99	99	99	99	99	99
Observations	283	283	283	283	283	283	283	283

Note: Random effect models with robust standard errors in parentheses; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Model 1 is a replication of Model 4 in Table 1 in the original study; dependent variable in Model 1 is a multiplicative index of two Bertelsmann Stiftung indicators: *Monopoly on the Use of Force* and *Basic Administration*. Models reproduce the specification and estimation methods of the original study. Model 1 is run with original data provided by the authors; all independent variables are taken from this original dataset; see information about sources in Carbone and Memoli (2015). HSI provides data only until 2009. Thus, I have coded its scores in 2010 equal to its scores in 2009.

Table A16. Replication of Grassi and Memoli (2016) and re-analysis with alternative measures of state capacity: robustness tests with common sample of country-years across models

	Replication (1)	QOG (2)	SFI (3)	WGI (4)	CPI (5)	VDEM (6)
Lagged stateness	0.164***	-0.010	-0.007	0.012	0.034	0.021
	(0.018)	(0.036)	(0.041)	(0.016)	(0.026)	(0.015)
Democracy	-0.038***	0.039***	-0.007	0.001	0.016*	-0.017
	(0.003)	(0.006)	(0.008)	(0.008)	(0.006)	(0.011)
Democracy^2	0.005***	-0.004*	0.004	-0.0002	-0.001	0.007
	(0.001)	(0.002)	(0.003)	(0.003)	(0.002)	(0.004)
Executive partisan balance	0.116*	0.040	0.089**	-0.062	-0.093	0.065
	(0.056)	(0.057)	(0.031)	(0.081)	(0.057)	(0.076)
Level of economic development (1994)	0.099	0.061	0.282***	0.219***	0.147	0.187**
	(0.063)	(0.080)	(0.046)	(0.057)	(0.078)	(0.072)
Land size (km^2)	0.015	0.016	-0.033*	-0.006	-0.002	-0.016
	(0.021)	(0.025)	(0.015)	(0.022)	(0.026)	(0.022)
Oil rents	-0.123	-0.112	-0.177	-0.293**	-0.161*	-0.076
	(0.164)	(0.167)	(0.141)	(0.101)	(0.076)	(0.170)
Log(Ethnic fractionalization)	-0.055	-0.070	-0.138*	-0.026	-0.111	-0.075
	(0.142)	(0.172)	(0.069)	(0.139)	(0.157)	(0.086)
Constant	-0.664	-0.398	-1.458***	-1.234*	-0.905	-1.137*
	(0.485)	(0.461)	(0.405)	(0.494)	(0.655)	(0.575)
Sigma_u	0.17	0.21	0.14	0.17	0.19	0.18
Sigma_e	0.04	0.06	0.07	0.05	0.05	0.04
Rho	0.94	0.91	0.81	0.92	0.93	0.95
Between R^2	0.29	0.13	0.77	0.58	0.39	0.49
Wald chi-square (8)	85521.73	6336.73	131.39	54.36	595.29	18.35
Prob > chi-square	0.000	0.000	0.000	0.000	0.000	0.019
Countries	18	18	18	18	18	18
Observations	154	154	154	154	154	154

Note: Random effects models with robust standard error in parentheses; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Model 1 is a replication of Model 3 in Table 1 in the original study; dependent variable in Model 1 is HSI. Models reproduce the specification and estimation methods of the original study. Model 1 is run with original data provided by the authors; all independent variables are taken from this original dataset; see information about sources in Grassi and Memoli (2016). FSI is excluded because it provides data only from 2005 onwards.

Table A17. Replication of Charron and Lapuente (2010) and re-analysis with alternative measures of state capacity: robustness tests with common sample of country-years across models

	Replication (1)	FSI (2)	HSI (3)	SFI (4)	WGI (5)	CPI (6)	VDEM (7)	CMI (8)
Democracy	-0.125	-0.126	-0.062	-0.057	-0.150**	-0.131*	-0.212***	-0.246*
	(0.064)	(0.065)	(0.039)	(0.065)	(0.050)	(0.059)	(0.054)	(0.094)
Democracy^2	0.005	0.009*	0.004	0.008*	0.005	0.007*	0.006*	0.016**
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.005)
Ln(GDP/capita)	-0.014	0.022	0.110***	0.150**	-0.003	0.028	-0.121**	0.051
	(0.050)	(0.051)	(0.030)	(0.051)	(0.039)	(0.046)	(0.042)	(0.073)
Democracy × Ln(GDP/capita)	0.009	0.008	0.002	-0.0005	0.013*	0.007	0.022**	0.010
	(0.008)	(0.008)	(0.005)	(0.008)	(0.006)	(0.008)	(0.007)	(0.012)
Trade openness	0.001*	0.0003	0.0001	0.001*	0.001*	0.0004	0.0003	0.001*
	(0.0004)	(0.0004)	(0.0002)	(0.0004)	(0.0003)	(0.0004)	(0.0003)	(0.001)
British colony	0.017	0.005	0.030	-0.025	0.084**	0.057	0.004	0.066
	(0.037)	(0.037)	(0.022)	(0.037)	(0.029)	(0.034)	(0.031)	(0.054)
Constant	0.509	0.166	-0.384	-0.722	0.386	0.085	1.307***	0.228
	(0.411)	(0.420)	(0.248)	(0.420)	(0.324)	(0.381)	(0.347)	(0.605)
R^2	0.38	0.67	0.76	0.79	0.66	0.48	0.69	0.65
Countries/Observations	65	65	65	65	65	65	65	65

Note: Cross-sectional OLS models with standard errors in parentheses; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Model 1 is a replication of Model 2 in Table 2 in the original study; dependent variable in Model 1 is QOG. Models reproduce the specification and estimation methods of the original study. Independent variables are taken from the QoG Standard Dataset (Teorell, Holmberg, and Rothstein 2008). Democracy = fh_ipolity2. GDP/capita = gle_rgdp. Trade = pwt_openk. British colony = ht_colonial. CMI measured in 2006, because coded only from 2006 onwards. FSI measured in 2005, because coded only from 2005 onwards.

Figure A1. Scatter plots of state capacity with different measures in 2009
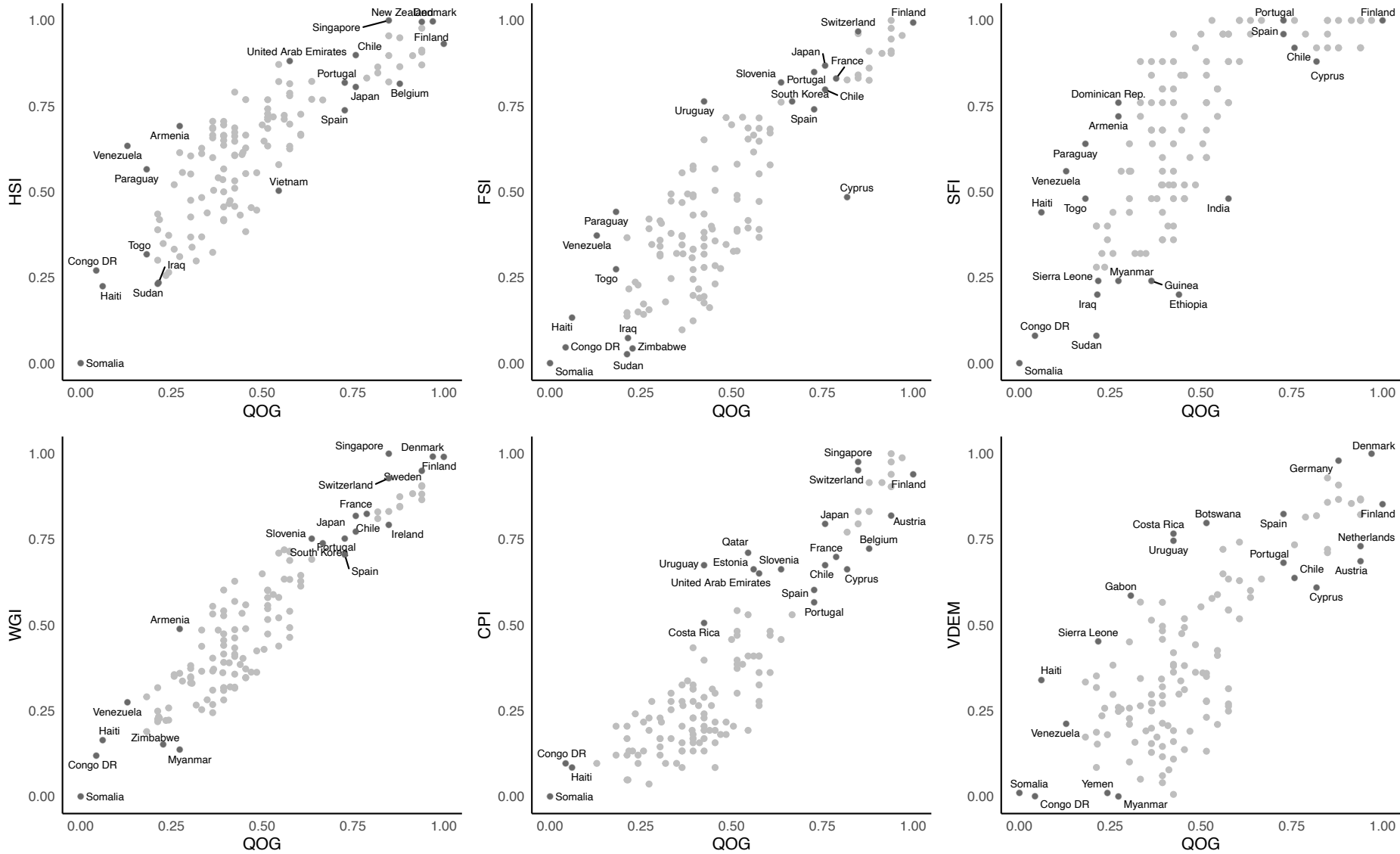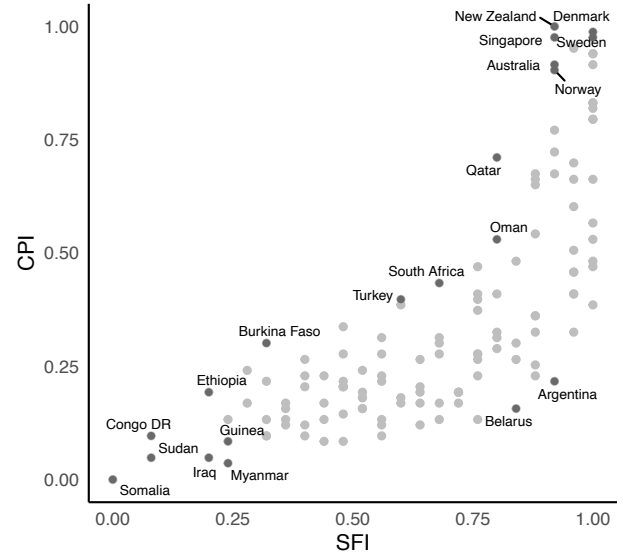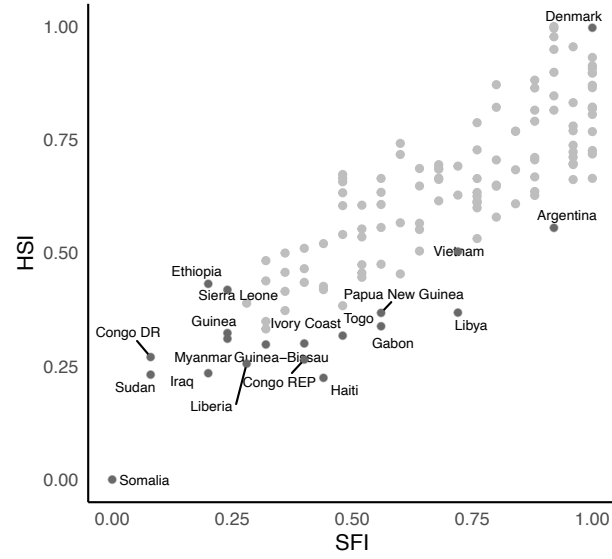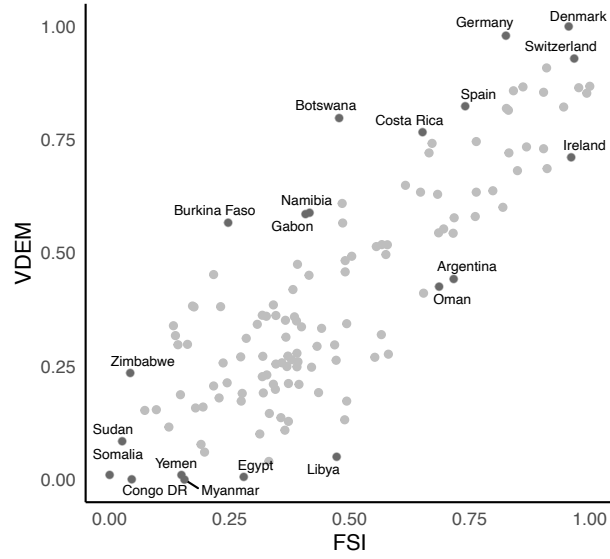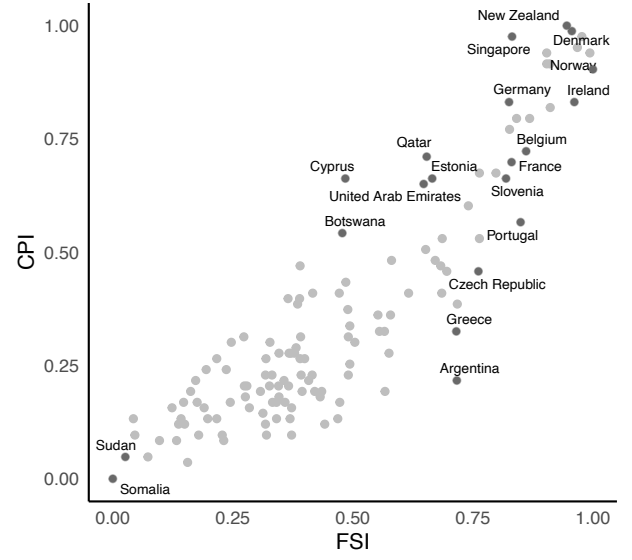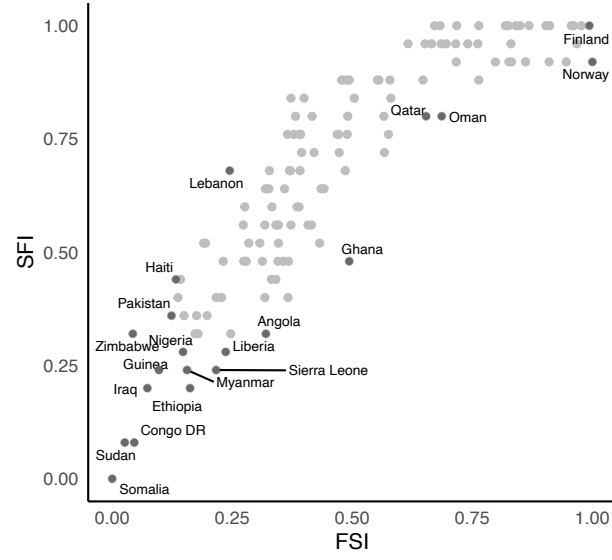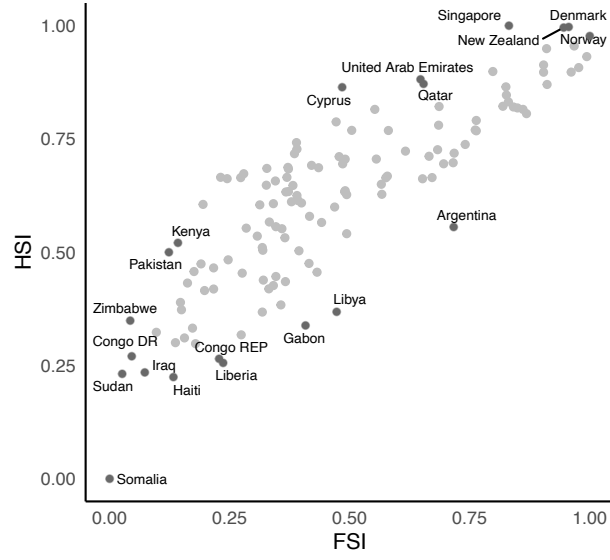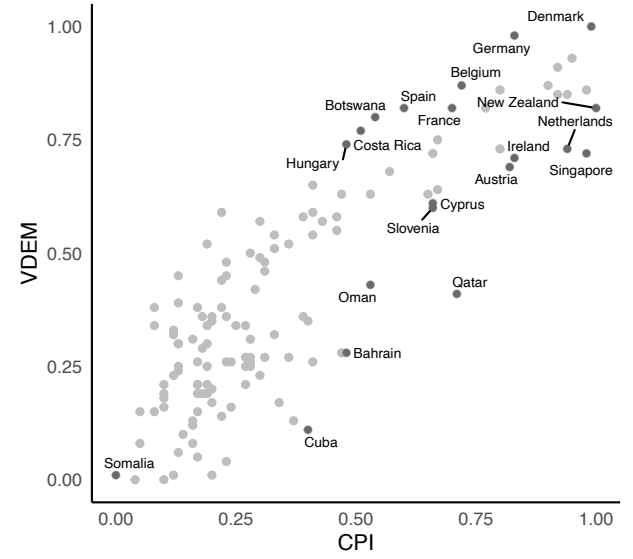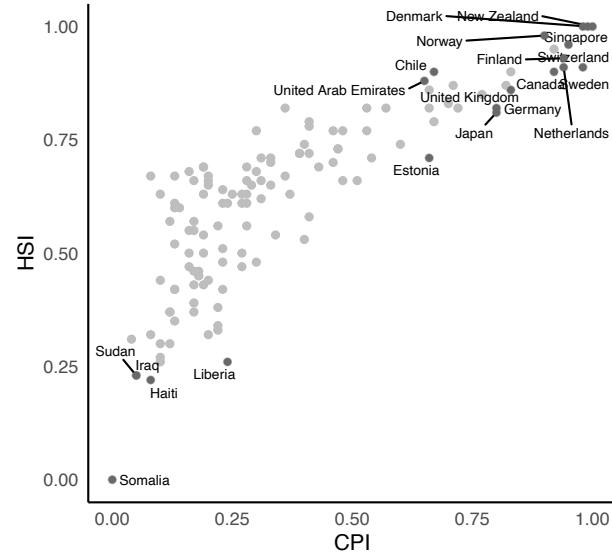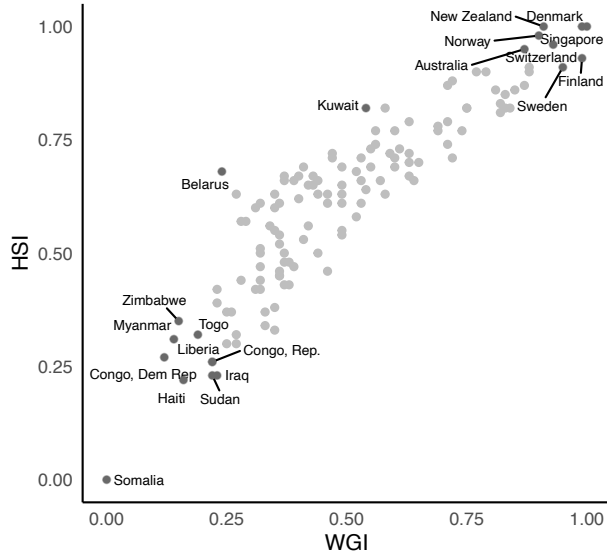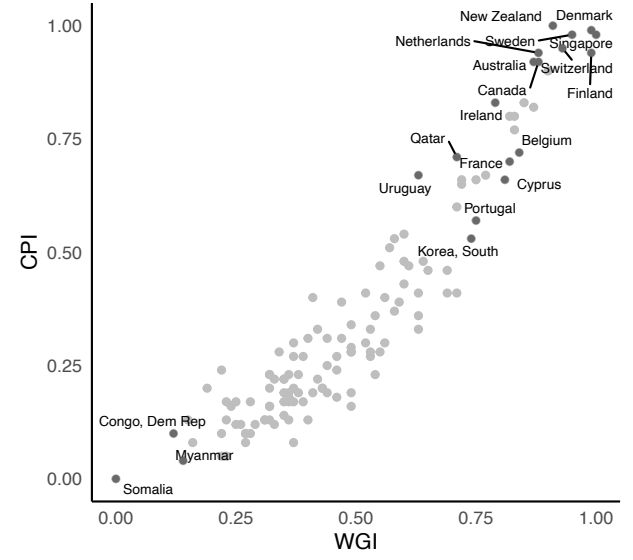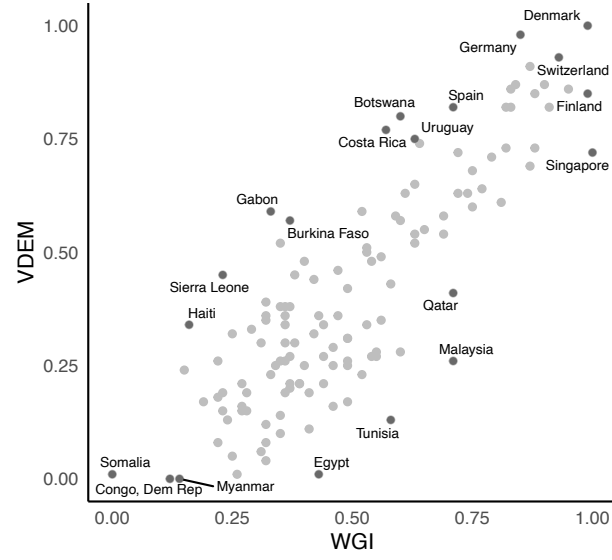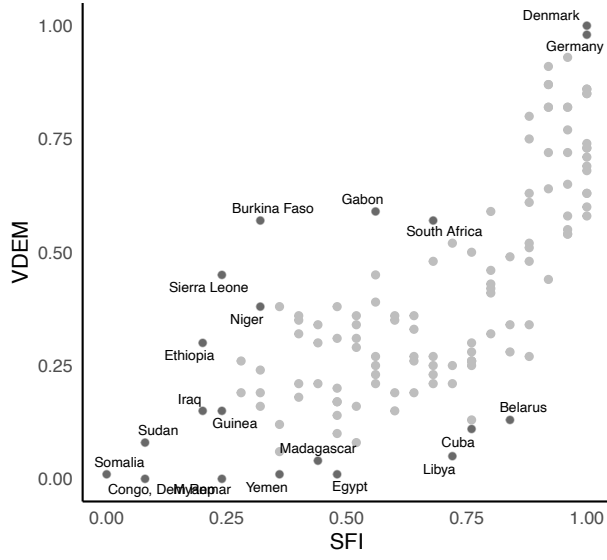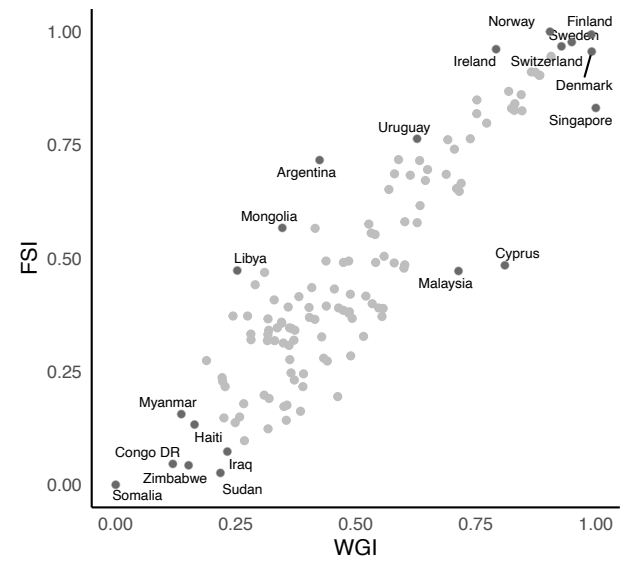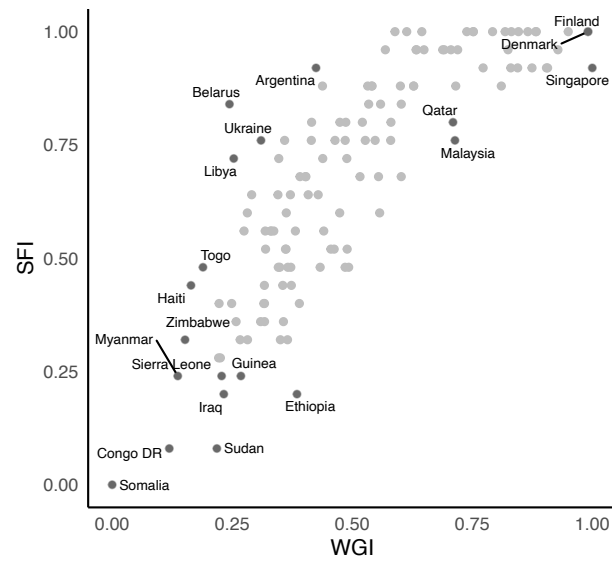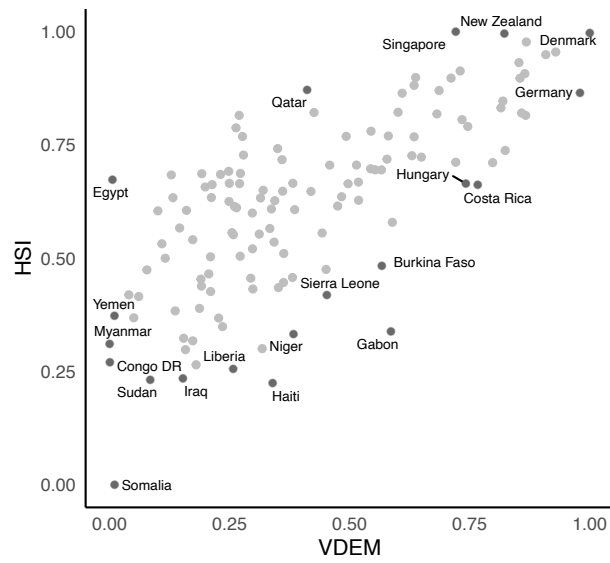
Note: Scores are normalised to range from 0 to 1.

Table A18. Observations with the highest pairwise rating discrepancy between measures of state capacity in 2009

Pair of measures	Observation (country)	Discrepancy (in units)	Higher score	Lower score
SFI and CPI	Argentina	0.70	0.92 with SFI	0.22 with CPI
	Belarus	0.68	0.84 with SFI	0.16 with CPI
	Jamaica	0.65	0.88 with SFI	0.23 with CPI
	Albania	0.63	0.88 with SFI	0.25 with CPI
	Greece	0.63	0.96 with SFI	0.33 with CPI
	Ukraine	0.63	0.76 with SFI	0.13 with CPI
SFI and VDEM	Belarus	0.71	0.84 with SFI	0.13 with VDEM
	Libya	0.67	0.72 with SFI	0.05 with VDEM
	Cuba	0.65	0.76 with SFI	0.11 with VDEM
	Tunisia	0.63	0.76 with SFI	0.13 with VDEM
	Kuwait	0.61	0.88 with SFI	0.27 with VDEM
SFI and WGI	Belarus	0.60	0.84 with SFI	0.24 with WGI
	Argentina	0.50	0.92 with SFI	0.42 with WGI
	Libya	0.47	0.72 with SFI	0.25 with WGI
	Ukraine	0.45	0.76 with SFI	0.31 with WGI
	Albania	0.44	0.88 with SFI	0.44 with WGI
SFI and QOG	Albania	0.55	0.88 with SFI	0.33 with QOG
	Costa Rica	0.54	0.96 with SFI	0.42 with QOG
	Bulgaria	0.52	0.88 with SFI	0.36 with QOG
	Dominican Rep.	0.49	0.76 with SFI	0.27 with QOG
	Jamaica	0.49	0.88 with SFI	0.39 with QOG
SFI and FSI	Belarus	0.47	0.84 with SFI	0.37 with FSI
	Lebanon	0.44	0.68 with SFI	0.24 with FSI
	Mexico	0.44	0.84 with SFI	0.40 with FSI
	Serbia	0.42	0.80 with SFI	0.38 with FSI
	Cyprus	0.40	0.88 with SFI	0.48 with FSI
	Botswana	0.40	0.88 with SFI	0.48 with FSI
SFI and HSI	Argentina	0.36	0.92 with SFI	0.56 with HSI
	Libya	0.35	0.72 with SFI	0.37 with HSI
	Hungary	0.34	1.00 with SFI	0.66 with HSI
	Costa Rica	0.30	0.96 with SFI	0.66 with HSI
	Italy	0.28	1.00 with SFI	0.72 with HSI
HSI and CPI	Iran	0.59	0.67 with HSI	0.08 with CPI
	Russia	0.54	0.67 with HSI	0.13 with CPI
	Venezuela	0.53	0.63 with HSI	0.10 with CPI
	Belarus	0.52	0.68 with HSI	0.16 with CPI
	Armenia	0.50	0.69 with HSI	0.19 with CPI
	Kazakhstan	0.50	0.69 with HSI	0.19 with CPI
HSI and VDEM	Egypt	0.66	0.67 with HSI	0.01 with VDEM
	Belarus	0.55	0.68 with HSI	0.13 with VDEM
	Kuwait	0.55	0.82 with HSI	0.27 with VDEM
	Malaysia	0.53	0.79 with HSI	0.26 with VDEM
	Tunisia	0.50	0.63 with HSI	0.13 with VDEM
	Azerbaijan	0.50	0.60 with HSI	0.10 with VDEM
	Kazakhstan	0.50	0.69 with HSI	0.19 with VDEM
HSI and FSI	Iran	0.44	0.67 with HSI	0.23 with FSI
	Sri Lanka	0.42	0.61 with HSI	0.19 with FSI
	Lebanon	0.42	0.66 with HSI	0.24 with FSI
	Colombia	0.39	0.66 with HSI	0.27 with FSI
	Egypt	0.39	0.67 with HSI	0.28 with FSI

HSI and QOG	Venezuela	0.50	0.63 with HSI	0.13 with QOG
	Armenia	0.42	0.69 with HSI	0.27 with QOG
	Paraguay	0.39	0.57 with HSI	0.18 with QOG
	Uruguay	0.37	0.79 with HSI	0.42 with QOG
	Bulgaria	0.35	0.71 with HSI	0.36 with QOG
HSI and WGI	Belarus	0.44	0.68 with HSI	0.24 with WGI
	Venezuela	0.36	0.63 with HSI	0.27 with WGI
	Iran	0.30	0.67 with HSI	0.37 with WGI
	Algeria	0.29	0.66 with HSI	0.37 with WGI
	Ecuador	0.29	0.61 with HSI	0.32 with WGI
	Ukraine	0.29	0.60 with HSI	0.31 with WGI
	Nicaragua	0.29	0.57 with HSI	0.28 with WGI
CPI and FSI	Argentina	0.50	0.72 with FSI	0.22 with CPI
	Greece	0.39	0.72 with FSI	0.33 with CPI
	Mongolia	0.38	0.57 with FSI	0.19 with CPI
	Ukraine	0.34	0.47 with FSI	0.13 with CPI
	Italy	0.33	0.72 with FSI	0.39 with CPI
CPI and VDEM	Gabon	0.37	0.59 with VDEM	0.22 with CPI
	Mongolia	0.33	0.52 with VDEM	0.19 with CPI
	Sierra Leone	0.32	0.45 with VDEM	0.13 with CPI
	Iran	0.30	0.38 with VDEM	0.08 with CPI
	Qatar	0.30	0.71 with CPI	0.41 with VDEM
CPI and QOG	Iran	0.37	0.45 with QOG	0.08 with CPI
	Vietnam	0.36	0.55 with QOG	0.19 with CPI
	Indonesia	0.31	0.51 with QOG	0.20 with CPI
	Morocco	0.31	0.58 with QOG	0.27 with CPI
	Tanzania	0.30	0.48 with QOG	0.18 with CPI
	India	0.30	0.58 with QOG	0.28 with CPI
WGI and CPI	Philippines	0.33	0.49 with WGI	0.16 with CPI
	Jamaica	0.31	0.54 with WGI	0.23 with CPI
	Greece	0.30	0.63 with WGI	0.33 with CPI
	Armenia	0.30	0.49 with WGI	0.19 with CPI
	Malaysia	0.30	0.71 with WGI	0.41 with CPI
WGI and FSI	Cyprus	0.33	0.81 with WGI	0.48 with FSI
	Argentina	0.30	0.72 with FSI	0.42 with WGI
	Sri Lanka	0.27	0.46 with WGI	0.19 with FSI
	Malaysia	0.24	0.71 with WGI	0.47 with FSI
	Ethiopia	0.22	0.38 with WGI	0.16 with FSI
	Kenya	0.22	0.36 with WGI	0.14 with FSI
	Libya	0.22	0.47 with FSI	0.25 with WGI
	Mongolia	0.22	0.57 with FSI	0.35 with WGI
WGI and QOG	Armenia	0.22	0.49 with WGI	0.27 with QOG
	Uruguay	0.21	0.63 with WGI	0.42 with QOG
	South Africa	0.21	0.60 with WGI	0.39 with QOG
	Thailand	0.19	0.55 with WGI	0.36 with QOG
	Bulgaria	0.17	0.53 with WGI	0.36 with QOG
WGI and VDEM	Tunisia	0.45	0.58 with WGI	0.13 with VDEM
	Malaysia	0.45	0.71 with WGI	0.26 with VDEM
	Egypt	0.42	0.43 with WGI	0.01 with VDEM
	Ghana	0.32	0.49 with WGI	0.17 with VDEM
	Bahrain	0.32	0.60 with WGI	0.28 with VDEM
QOG and VDEM	Egypt	0.41	0.42 with QOG	0.01 with VDEM
	Tunisia	0.39	0.52 with QOG	0.13 with VDEM
	Costa Rica	0.35	0.77 with VDEM	0.42 with QOG
	Madagascar	0.35	0.39 with QOG	0.04 with VDEM
	Vietnam	0.34	0.55 with QOG	0.21 with VDEM

QOG and FSI	Uruguay	0.34	0.76 with FSI	0.42 with QOG
	Cyprus	0.34	0.82 with QOG	0.48 with FSI
	Ethiopia	0.28	0.44 with QOG	0.16 with FSI
	Pakistan	0.27	0.39 with QOG	0.12 with FSI
	Paraguay	0.26	0.44 with FSI	0.18 with QOG
	Guinea	0.26	0.36 with QOG	0.10 with FSI
FSI and VDEM	Libya	0.42	0.47 with FSI	0.05 with VDEM
	Tunisia	0.36	0.49 with FSI	0.13 with VDEM
	Botswana	0.32	0.80 with VDEM	0.48 with FSI
	Ghana	0.32	0.49 with FSI	0.17 with VDEM
	Burkina Faso	0.32	0.57 with VDEM	0.25 with FSI

Note: Scores are normalised to range from 0 to 1. For each pair of measures, the five observations with the highest rating discrepancy are reported. If the fifth "position" is shared by multiple observations, each of them is reported.

Appendix B

Table B1. Data sources

Variable	Source
Civic participation (CCSI)	Varieties of Democracy (Coppedge et al. 2020a)
Civic participation (CSPART)	Varieties of Democracy (Coppedge et al. 2020a)
Impartial public officials (V-Dem)	Varieties of Democracy (Coppedge et al. 2020a)
Impartial public officials (QoG)	QoG Expert Survey II (Dahlström et al. 2015)
GDP/capita	The Maddison Project Database (Bolt et al. 2018)
Natural resources rents	World Development Indicators (World Bank 2019)
Total population	World Development Indicators (World Bank 2019)
Years of education	Varieties of Democracy (Coppedge et al. 2020a)
Ethnic fractionalisation	Fractionalization dataset (Alesina et al. 2003)
Regime type dummies	Varieties of Democracy (Coppedge et al. 2020a)
Region dummies	Varieties of Democracy (Coppedge et al. 2020a)

Note: The set of data covers all available countries in the world from 1945 to 2017.

Table B2. Summary statistics of variables used in panel regressions (Tables 3.1 and 3.2)

Variable		Mean	SD	Min	Max	Observations
Impartial pub. officials	Overall	0.186	1.505	-3.685	4.455	N = 9935
(V-Dem)	Between		1.303	-2.033	3.839	n = 162
	Within		0.725	-3.113	2.592	T-bar = 61.327
Civic participation	Overall	0.539	0.317	0.008	0.977	N = 9935
(CCSI)	Between		0.244	0.030	0.967	n = 162
	Within		0.207	-0.089	1.285	T-bar = 61.327
Civic participation	Overall	0.533	0.296	0.021	0.989	N = 9935
(CSPART)	Between		0.223	0.066	0.975	n = 162
	Within		0.193	-0.051	1.140	T-bar = 61.327
Civic participation^2	Overall	0.392	0.335	0.0001	0.955	N = 9935
(CCSI)	Between		0.264	0.002	0.936	n = 162
	Within		0.211	-0.276	1.145	T-bar = 61.327
Civic participation^2	Overall	0.371	0.314	0.0001	0.978	N = 9935
(CSPART)	Between		0.246	0.006	0.950	n = 162
	Within		0.191	-0.151	1.123	T-bar = 61.327
Ln(GDP/capita)	Overall	8.482	1.168	4.898	12.249	N = 9770
	Between		1.048	6.424	11.314	n = 162
	Within		0.538	6.061	10.820	T-bar = 60.309
Civic participation	Overall	4.720	3.095	0.059	11.031	N = 9770
(CCSI) × GDP/capita	Between		2.501	0.209	9.861	n = 162
	Within		1.848	-1.041	12.024	T-bar = 60.309
Civic participation	Overall	4.654	2.925	0.135	11.074	N = 9770
(CSPART) × GDP/capita	Between		2.335	0.473	9.828	n = 162
	Within		1.727	-0.627	10.621	T-bar = 60.309
Ln(Natural resources	Overall	0.772	2.101	-8.337	4.489	N = 6396
rents)	Between		2.081	-8.063	3.748	n = 158
	Within		0.660	-3.371	4.620	T-bar = 40.481
Ln(Total population)	Overall	15.922	1.643	11.010	21.050	N = 7974
	Between		1.608	11.238	20.773	n = 159
	Within		0.337	14.598	17.374	T-bar = 50.151
Years of education	Overall	5.869	3.463	0.043	13.610	N = 8453
	Between		3.191	0.594	11.819	n = 135
	Within		1.641	0.188	10.441	T-bar = 62.615

		Overall		Between		Within
		Freq.	%	Freq.	%	%
Regime type dummies		9915	100.00	391	241.36	41.43
Closed autocracy		3439	34.68	123	75.93	45.48
Electoral autocracy		2739	27.62	125	77.16	36.86
Electoral democracy		1763	17.78	93	57.41	31.39
Liberal democracy		1974	19.91	50	30.86	61.58

Table B3. Summary statistics for variables used in cross-section regressions (Table 3.6)

Variable	Observations	Mean	SD	Min	Max
Impartial pub. officials (QoG)	110	3.971	1.198	1.573	6.292
Civic participation (CCSI)	110	0.755	0.217	0.126	0.975
Civic participation (CSPART)	110	0.759	0.1797967	0.096	0.985
Ln(GDP/capita)	110	9.408	1.063855	6.867	11.242
Ln(Natural resources rents)	106	0.019	2.262254	-8.134	3.559
Ln(Total population)	109	16.593	1.562412	12.557	21.039
Years of education	99	8.679	2.96151	2.680	13.610
Ethnic fractionalisation	109	0.414	0.2474137	0.002	0.930

	Frequency	%	
Regime type dummies	110	100.00	
Closed autocracy	5	4.55	
Electoral autocracy	30	27.27	
Electoral democracy	36	32.73	
Liberal democracy	39	35.45	
Region dummies	110	100.00	
Former Soviet Union	24	21.82	
Latin America	17	15.45	
Middle East & N. Africa	8	7.27	
Sub-Saharan Africa	22	20.00	
Western countries	23	20.91	
Asia	16	14.55	

Table B4. Unit-root tests of key variables

Variable	Phillips-Perron test
Impartial public officials (V-Dem)	-3.823**
Civic participation (CCSI)	-7.197**
Civic participation (CSPART)	-1.990*

Note: Inverse normal Z statistics; *$p < 0.05$, **$p < 0.01$.

Table B5. Civic participation and impartial public officials: robustness tests with common sample of country-years across models

	Dependent variable: Impartial public officials $_t$			
	Baseline model	Additional controls	Regime types	Full model
	(1)	(2)	(3)	(4)
Civic participation $_{t-1}$	0.126***	0.133***	0.148***	0.154***
	(0.034)	(0.037)	(0.043)	(0.044)
Ln(GDP/capita) $_{t-1}$	0.017	0.024	0.016	0.024
	(0.016)	(0.017)	(0.016)	(0.017)
Civic participation^2 $_{t-1}$				
Civic participation × GDP/capita $_{t-1}$				
Ln(Natural resources) $_{t-1}$		0.004		0.004
		(0.005)		(0.006)
Ln(Population) $_{t-1}$		-0.010		-0.004
		(0.034)		(0.035)
Years of education $_{t-1}$		-0.016		-0.016
		(0.010)		(0.010)
Impartial pub. officials $_{t-1}$	0.899***	0. 898***	0.900***	0.899***
	(0.008)	(0.008)	(0.008)	(0.008)
Long-run multiplier				
Civic participation $_{t-1}$	1.253***	1.302***	1.490***	1.532***
	(0.285)	(0.305)	(0.395)	(0.395)
Within R^2	0.86	0.86	0.86	0.86
N	5433	5433	5433	5433
Sample	1971-2017	1971-2017	1971-2017	1971-2017
Countries	133	133	133	133
Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Regime type dummies $_{t-1}$	No	No	Yes	Yes

Note: Robust standard errors in parentheses; *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$. Constant coefficient measured but not reported. Sample of years refers to dependent variable. Civic participation is measured with CCSI.

Appendix C

Figure C1. Missingness pattern of the imputation dataset
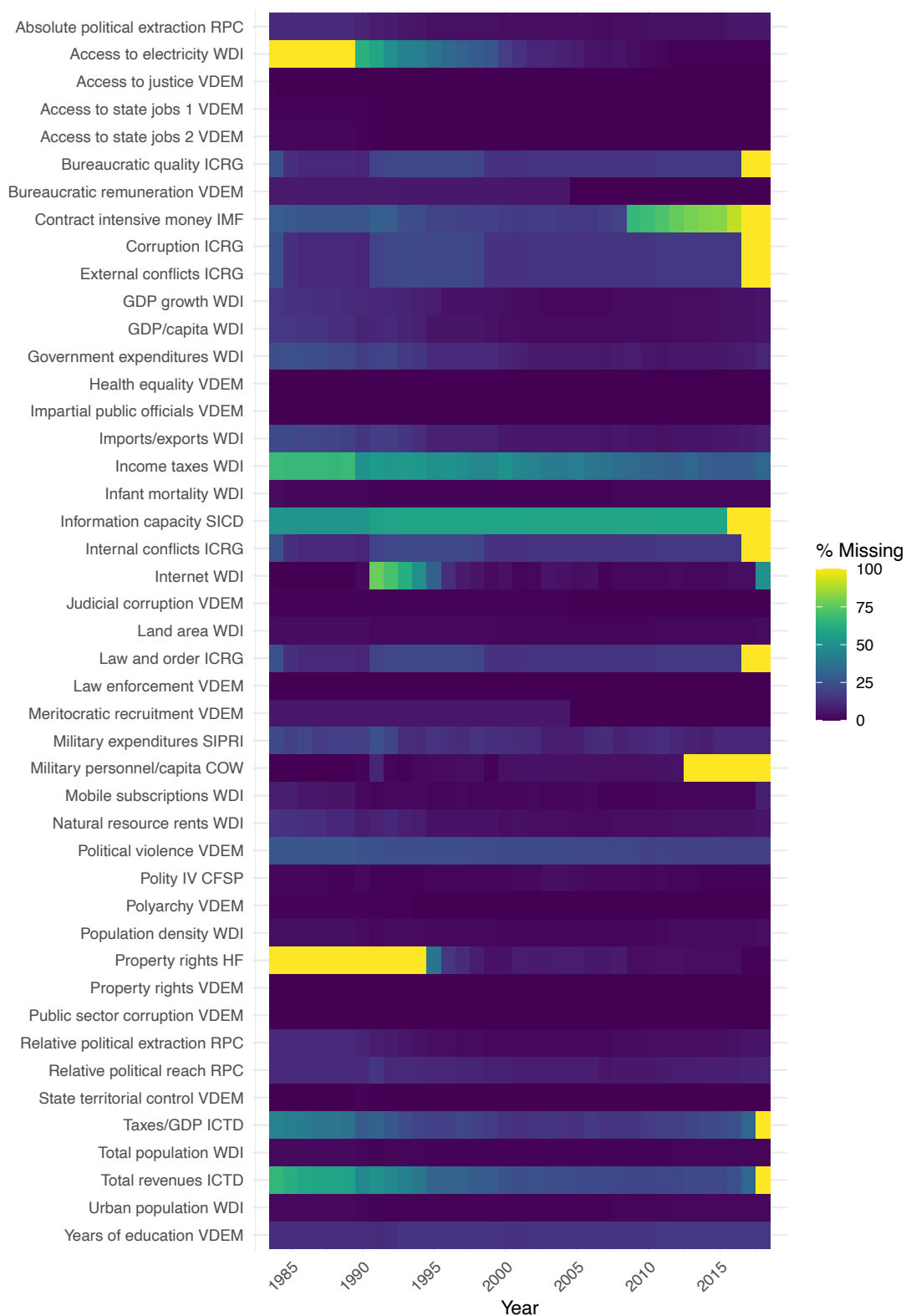
Note: Information about data sources in Table C7 (Appendix C). Internet (WDI) coded as 0 before year 1990.

Table C1. Correlations between observed measures of state capacity before (lower-left quadrant) and after (upper-right quadrant) imputation

Ind.	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20
V1	1.00	0.89	0.59	0.37	0.52	0.41	0.34	0.37	0.41	0.45	0.53	0.49	0.56	0.29	0.40	0.50	0.55	0.41	0.58	0.46
V2	0.94	1.00	0.66	0.38	0.53	0.49	0.40	0.39	0.38	0.45	0.55	0.49	0.60	0.30	0.43	0.51	0.58	0.44	0.58	0.48
	(3133)																			
V3	0.65	0.70	1.00	0.31	0.48	0.69	0.57	0.36	0.37	0.40	0.47	0.32	0.50	0.39	0.46	0.48	0.49	0.37	0.50	0.39
	(3362)	(3888)																		
V4	0.39	0.43	0.31	1.00	0.34	0.36	0.33	0.39	0.24	0.27	0.39	0.27	0.45	0.34	0.29	0.29	0.29	0.39	0.40	0.34
	(3289)	(3788)	(4800)																	
V5	0.59	0.62	0.53	0.38	1.00	0.49	0.45	0.42	0.36	0.49	0.68	0.66	0.70	0.42	0.46	0.64	0.59	0.62	0.68	0.75
	(2806)	(3280)	(4001)	(3998)																
V6	0.41	0.55	0.66	0.38	0.56	1.00	0.75	0.36	0.37	0.38	0.44	0.27	0.46	0.39	0.47	0.42	0.40	0.34	0.43	0.43
	(2687)	(2950)	(3616)	(3608)	(2926)															
V7	0.39	0.47	0.59	0.34	0.49	0.75	1.00	0.30	0.31	0.36	0.50	0.33	0.52	0.42	0.46	0.41	0.40	0.36	0.50	0.51
	(3422)	(3950)	(4988)	(5022)	(4112)	(3689)														
V8	0.36	0.43	0.37	0.39	0.45	0.33	0.30	1.00	0.48	0.51	0.39	0.33	0.42	0.40	0.31	0.44	0.44	0.32	0.43	0.32
	(3422)	(3950)	(4988)	(5081)	(4157)	(3705)	(5214)													
V9	0.50	0.44	0.45	0.26	0.41	0.42	0.35	0.54	1.00	0.58	0.43	0.33	0.46	0.19	0.31	0.48	0.47	0.31	0.48	0.37
	(2663)	(3088)	(3865)	(3988)	(3291)	(2973)	(4065)	(4107)												
V10	0.51	0.49	0.42	0.30	0.50	0.36	0.38	0.54	0.63	1.00	0.39	0.40	0.42	0.29	0.38	0.68	0.48	0.34	0.50	0.45
	(2806)	(3280)	(4001)	(3998)	(4158)	(2926)	(4112)	(4157)	(3291)											
V11	0.55	0.60	0.49	0.40	0.73	0.47	0.50	0.39	0.47	0.44	1.00	0.64	0.84	0.43	0.51	0.55	0.71	0.74	0.85	0.76
	(3422)	(3950)	(4988)	(5081)	(4158)	(3705)	(5214)	(5279)	(4107)	(4158)										
V12	0.53	0.55	0.38	0.32	0.68	0.37	0.39	0.37	0.39	0.43	0.68	1.00	0.63	0.22	0.32	0.58	0.46	0.53	0.63	0.65
	(2806)	(3280)	(4001)	(3998)	(4158)	(2926)	(4112)	(4157)	(3291)	(4158)	(4158)									
V13	0.58	0.63	0.51	0.45	0.74	0.48	0.53	0.42	0.50	0.46	0.84	0.67	1.00	0.49	0.43	0.58	0.64	0.70	0.78	0.74
	(3401)	(3944)	(4967)	(5060)	(4137)	(3690)	(5193)	(5258)	(4086)	(4137)	(5259)	(4137)								
V14	0.33	0.35	0.45	0.37	0.56	0.44	0.45	0.41	0.25	0.33	0.50	0.35	0.54	1.00	0.34	0.32	0.41	0.41	0.42	0.43
	(2196)	(2548)	(3134)	(3022)	(2704)	(2028)	(3195)	(3195)	(2465)	(2704)	(3195)	(2704)	(3174)							
V15	0.47	0.51	0.50	0.29	0.50	0.44	0.46	0.31	0.34	0.39	0.51	0.37	0.43	0.34	1.00	0.36	0.73	0.55	0.63	0.51
	(3422)	(3950)	(4988)	(5081)	(4158)	(3705)	(5214)	(5279)	(4107)	(4158)	(5280)	(4158)	(5259)	(3195)						
V16	0.57	0.58	0.51	0.31	0.65	0.44	0.45	0.46	0.54	0.67	0.60	0.61	0.63	0.44	0.38	1.00	0.53	0.45	0.58	0.63
	(2806)	(3280)	(4001)	(3998)	(4158)	(2926)	(4112)	(4157)	(3291)	(4158)	(4158)	(4158)	(4137)	(2704)	(4158)					
V17	0.57	0.62	0.51	0.28	0.63	0.35	0.40	0.44	0.51	0.50	0.71	0.51	0.63	0.44	0.73	0.56	1.00	0.65	0.81	0.67
	(3422)	(3950)	(4988)	(5081)	(4158)	(3705)	(5214)	(5279)	(4107)	(4158)	(5280)	(4158)	(5259)	(3195)	(5280)	(4158)				
V18	0.45	0.50	0.39	0.39	0.68	0.37	0.37	0.31	0.32	0.36	0.74	0.58	0.69	0.43	0.55	0.50	0.66	1.00	0.73	0.67
	(3289)	(3794)	(4817)	(5081)	(3998)	(3608)	(5039)	(5098)	(3988)	(3998)	(5098)	(3998)	(5077)	(3029)	(5098)	(3998)	(5098)			
V19	0.61	0.63	0.52	0.40	0.72	0.40	0.50	0.43	0.51	0.53	0.85	0.67	0.78	0.47	0.63	0.63	0.81	0.74	1.00	0.76
	(3422)	(3950)	(4988)	(5081)	(4158)	(3705)	(5214)	(5279)	(4107)	(4158)	(5280)	(4158)	(5259)	(3195)	(5280)	(4158)	(5280)	(5098)		
V20	0.49	0.49	0.39	0.36	0.78	0.41	0.50	0.32	0.41	0.45	0.77	0.73	0.77	0.46	0.53	0.62	0.68	0.71	0.76	1.00
	(2552)	(2774)	(3345)	(3392)	(2723)	(3181)	(3435)	(3472)	(2795)	(2723)	(3472)	(2723)	(3462)	(1900)	(3472)	(2723)	(3472)	(3392)	(3472)	

Note: Entries are Pearson's correlation coefficients. Number of observations (N) in parentheses; N = 5280 throughout the upper-right quadrant. Abbreviations of indicators explained in Table C7 (Appendix C). Ind. = Indicator.

Table C2. Factor analysis output with imputation dataset 1

Observed indicator	Factor 1	Factor 2	Communality	Uniqueness
Total revenues	0.17	0.61	0.54	0.46
Taxes/GDP	0.13	0.69	0.63	0.37
Absolute political extraction	-0.06	0.86	0.66	0.34
Bureaucratic remuneration	0.20	0.33	0.24	0.76
Bureaucratic quality	0.68	0.15	0.63	0.37
Access to electricity	-0.06	0.78	0.54	0.46
Urban population	0.13	0.55	0.43	0.57
State territorial control	0.20	0.41	0.32	0.68
Political violence	0.25	0.35	0.31	0.69
Internal conflicts	0.30	0.37	0.38	0.62
Public sector corruption	0.90	-0.01	0.81	0.19
Corruption	0.66	0.10	0.54	0.46
Judicial corruption	0.76	0.15	0.76	0.24
Contract intensive money	0.34	0.19	0.24	0.76
Property rights (V-Dem)	0.53	0.15	0.41	0.59
Law and order	0.46	0.30	0.49	0.51
Access to justice	0.77	0.06	0.67	0.33
Meritocratic recruitment	0.91	-0.16	0.65	0.35
Impartial public officials	0.94	-0.02	0.85	0.15
Property rights (HF)	0.88	-0.02	0.75	0.25
Eigen value	9.84	1.00		
Sum of squared loadings	6.92	3.92		
Proportion of explained variance	0.35	0.20		
Relative explained variance	0.64	0.36		
Overall KMO			0.92	
Correlation between factors			0.71	

Table C3. Factor analysis output with imputation dataset 2

Observed indicator	Factor 1	Factor 2	Communality	Uniqueness
Total revenues	0.18	0.59	0.53	0.47
Taxes/GDP	0.14	0.68	0.62	0.38
Absolute political extraction	-0.06	0.86	0.66	0.34
Bureaucratic remuneration	0.20	0.33	0.24	0.76
Bureaucratic quality	0.67	0.16	0.63	0.37
Access to electricity	-0.07	0.79	0.55	0.45
Urban population	0.14	0.55	0.43	0.57
State territorial control	0.18	0.42	0.32	0.68
Political violence	0.23	0.35	0.30	0.70
Internal conflicts	0.28	0.40	0.39	0.61
Public sector corruption	0.90	0.00	0.81	0.19
Corruption	0.64	0.12	0.53	0.47
Judicial corruption	0.76	0.15	0.76	0.24
Contract intensive money	0.34	0.20	0.25	0.75
Property rights (V-Dem)	0.53	0.14	0.41	0.59
Law and order	0.45	0.32	0.50	0.50
Access to justice	0.78	0.06	0.67	0.33
Meritocratic recruitment	0.91	-0.16	0.65	0.35
Impartial public officials	0.93	-0.01	0.85	0.15
Property rights (HF)	0.88	-0.02	0.74	0.26
Eigen value	9.87	0.98		
Sum of squared loadings	6.89	3.97		
Proportion of explained variance	0.34	0.20		
Relative explained variance	0.63	0.37		
Overall KMO			0.92	
Correlation between factors			0.71	

Table C4. Factor analysis output with imputation dataset 3

Observed indicator	Factor 1	Factor 2	Communality	Uniqueness
Total revenues	0.18	0.59	0.53	0.47
Taxes/GDP	0.14	0.68	0.62	0.38
Absolute political extraction	-0.06	0.86	0.66	0.34
Bureaucratic remuneration	0.20	0.33	0.24	0.76
Bureaucratic quality	0.67	0.16	0.63	0.37
Access to electricity	-0.07	0.79	0.55	0.45
Urban population	0.14	0.55	0.43	0.57
State territorial control	0.18	0.42	0.32	0.68
Political violence	0.23	0.35	0.30	0.7
Internal conflicts	0.28	0.40	0.39	0.61
Public sector corruption	0.90	0.00	0.81	0.19
Corruption	0.64	0.12	0.53	0.47
Judicial corruption	0.76	0.15	0.76	0.24
Contract intensive money	0.34	0.20	0.25	0.75
Property rights (V-Dem)	0.53	0.14	0.41	0.59
Law and order	0.45	0.32	0.50	0.50
Access to justice	0.78	0.06	0.67	0.33
Meritocratic recruitment	0.91	-0.16	0.65	0.35
Impartial public officials	0.93	-0.01	0.85	0.15
Property rights (HF)	0.88	-0.02	0.74	0.26
Eigen value	9.87	0.99		
Sum of squared loadings	6.89	3.97		
Proportion of explained variance	0.34	0.20		
Relative explained variance	0.63	0.37		
Overall KMO			0.92	
Correlation between factors			0.71	

Table C5. Factor analysis output with imputation dataset 4

Observed indicator	Factor 1	Factor 2	Communality	Uniqueness
Total revenues	0.17	0.59	0.53	0.47
Taxes/GDP	0.12	0.69	0.61	0.39
Absolute political extraction	-0.06	0.85	0.65	0.35
Bureaucratic remuneration	0.20	0.33	0.24	0.76
Bureaucratic quality	0.68	0.15	0.63	0.37
Access to electricity	-0.06	0.79	0.55	0.45
Urban population	0.13	0.55	0.43	0.57
State territorial control	0.17	0.43	0.32	0.68
Political violence	0.21	0.39	0.31	0.69
Internal conflicts	0.29	0.39	0.40	0.60
Public sector corruption	0.90	-0.01	0.81	0.19
Corruption	0.64	0.11	0.53	0.47
Judicial corruption	0.76	0.15	0.76	0.24
Contract intensive money	0.35	0.20	0.26	0.74
Property rights (V-Dem)	0.53	0.15	0.41	0.59
Law and order	0.44	0.33	0.51	0.49
Access to justice	0.77	0.07	0.67	0.33
Meritocratic recruitment	0.92	-0.16	0.66	0.34
Impartial public officials	0.93	-0.01	0.85	0.15
Property rights (HF)	0.88	-0.02	0.75	0.25
Eigen value	9.90	0.98		
Sum of squared loadings	6.87	4.01		
Proportion of explained variance	0.34	0.20		
Relative explained variance	0.63	0.37		
Overall KMO			0.93	
Correlation between factors			0.72	

Table C6. Factor analysis output with imputation dataset 5

Observed indicator	Factor 1	Factor 2	Communality	Uniqueness
Total revenues	0.15	0.62	0.54	0.46
Taxes/GDP	0.11	0.71	0.62	0.38
Absolute political extraction	-0.06	0.84	0.64	0.36
Bureaucratic remuneration	0.20	0.33	0.24	0.76
Bureaucratic quality	0.67	0.16	0.63	0.37
Access to electricity	-0.08	0.79	0.54	0.46
Urban population	0.13	0.55	0.43	0.57
State territorial control	0.16	0.44	0.32	0.68
Political violence	0.21	0.39	0.31	0.69
Internal conflicts	0.25	0.44	0.41	0.59
Public sector corruption	0.90	0.00	0.81	0.19
Corruption	0.63	0.12	0.53	0.47
Judicial corruption	0.75	0.16	0.76	0.24
Contract intensive money	0.36	0.19	0.26	0.74
Property rights (V-Dem)	0.53	0.14	0.41	0.59
Law and order	0.43	0.34	0.51	0.49
Access to justice	0.77	0.07	0.67	0.33
Meritocratic recruitment	0.92	-0.16	0.66	0.34
Impartial public officials	0.93	-0.01	0.85	0.15
Property rights (HF)	0.88	-0.02	0.75	0.25
Eigen value	9.90	0.99		
Sum of squared loadings	6.79	4.10		
Proportion of explained variance	0.34	0.21		
Relative explained variance	0.62	0.38		
Overall KMO			0.92	
Correlation between factors			0.72	

Table C7. Data sources of multiple imputation dataset

Abbr.	Variable	Source
V1	Total revenues	ICTD/UNU-WIDER GRD (Prichard 2016)
V2	Taxes/GDP	ICTD/UNU-WIDER GRD (Prichard 2016)
V3	Absolute political extraction	Relative Political Capacity Dataset (Fisunoglu et al. 2020)
V4	Bureaucratic remuneration	Varieties of Democracy (Coppedge et al. 2020a)
V5	Bureaucratic quality	International Country Risk Guide (PRS Group 2019)
V6	Access to electricity	World Development Indicators (World Bank 2020)
V7	Urban population	World Development Indicators (World Bank 2020)
V8	State territorial control	Varieties of Democracy (Coppedge et al. 2020a)
V9	Political violence	Varieties of Democracy (Coppedge et al. 2020a)
V10	Internal conflicts	International Country Risk Guide (PRS Group 2019)
V11	Public sector corruption	Varieties of Democracy (Coppedge et al. 2020a)
V12	Corruption	International Country Risk Guide (PRS Group 2019)
V13	Judicial corruption	Varieties of Democracy (Coppedge et al. 2020a)
V14	Contract intensive money	Based on IMF data assembled by Graham and Tucker (2019)
V15	Property rights (V-Dem)	Varieties of Democracy (Coppedge et al. 2020a)
V16	Law and order	International Country Risk Guide (PRS Group 2019)
V17	Access to justice	Varieties of Democracy (Coppedge et al. 2020a)
V18	Meritocratic recruitment	Varieties of Democracy (Coppedge et al. 2020a)
V19	Impartial public officials	Varieties of Democracy (Coppedge et al. 2020a)
V20	Property rights (HF)	Heritage Foundation (2020)
	Access to state jobs 1	Varieties of Democracy (Coppedge et al. 2020a)
	Access to state jobs 2	Varieties of Democracy (Coppedge et al. 2020a)
	External conflicts	International Country Risk Guide (PRS Group 2019)
	GDP growth	World Development Indicators (World Bank 2020)
	GDP/capita	World Development Indicators (World Bank 2020)
	Government expenditures	World Development Indicators (World Bank 2020)
	Health equality	Varieties of Democracy (Coppedge et al. 2020a)
	Imports/exports	World Development Indicators (World Bank 2020)
	Income taxes	World Development Indicators (World Bank 2020)
	Infant mortality	World Development Indicators (World Bank 2020)
	Information capacity	Brambor et al. (2020)
	Internet	World Development Indicators (World Bank 2020)
	Land area	World Development Indicators (World Bank 2020)
	Law enforcement	Varieties of Democracy (Coppedge et al. 2020a)
	Military expenditures	Stockholm International Peace Research Institute (SIPRI 2020)
	Military personnel/capita	National Material Capabilities (Correlates of War 2020)
	Mobile subscriptions	World Development Indicators (World Bank 2020)
	Natural resource rents	World Development Indicators (World Bank 2020)
	Polity IV	Marshall, Gurr, and Jaggers (2019)
	Polyarchy	Varieties of Democracy (Teorell et al. 2019)
	Population density	World Development Indicators (World Bank 2020)
	Relative political extraction	Relative Political Capacity Dataset (Fisunoglu et al. 2020)
	Relative political reach	Relative Political Capacity Dataset (Fisunoglu et al. 2020)
	Total population	World Development Indicators (World Bank 2020)
	Years of education	Varieties of Democracy (Coppedge et al. 2020a)

Abbr. = Abbreviation. Access to state jobs 1 = *v2peasjoc* in original source; Access to state jobs 2 = *v2peasjsoecon* in original source.

Table C8. Largest/smallest rank discrepancies between Power and Procedures in 2018

| Country | Rank | |
	Power	Procedures
Morocco	59	59
Ghana	97	97
Estonia	20	20
South Africa	56	56
Congo DR	155	156
Iran	118	117
Guinea-Bissau	144	143
Latvia	31	30
Eswatini	93	94
Finland	4	6
Netherlands	10	12
Gabon	100	98
Tunisia	39	41
Oman	44	46
Romania	72	74
Sudan	151	149
Yemen	154	157
Sweden	1	4
Somalia	156	153
Vietnam	86	83
Niger	150	93
El Salvador	63	121
Russia	41	100
Sri Lanka	113	52
Bosnia-Herzegovina	30	91
Costa Rica	94	32
Malaysia	98	35
Honduras	53	116
Turkey	60	125
Bolivia	46	112
Ireland	74	8
Tajikistan	64	131
Tanzania	135	66
Burkina Faso	146	76
Gambia	129	53
Singapore	80	2
Benin	131	50
Ukraine	43	126
Rwanda	119	34
Cuba	13	120

Table C9. Summary statistics of variables used to replicate Hanson (2015)

Variable		Mean	SD	Min	Max	Observations
Infant mortality	Overall	39.505	35.776	1.700	176.500	N = 5280
	Between		32.145	3.391	130.160	n = 158
	Within		15.034	-19.926	101.403	T-bar = 33.418
State capacity	Overall	0.487	0.240	0.000	1.000	N = 5280
	Between		0.232	0.004	0.998	n = 158
	Within		0.058	0.141	0.754	T-bar = 33.418
Power	Overall	0.536	0.233	0.000	1.000	N = 5280
	Between		0.223	0.047	0.995	n = 158
	Within		0.068	0.187	0.834	T-bar = 33.418
Procedures	Overall	0.447	0.243	0.000	1.000	N = 5280
	Between		0.234	0.012	1.000	n = 158
	Within		0.061	0.034	0.740	T-bar = 33.418
Democracy	Overall	0.484	0.280	0.015	0.924	N = 5280
	Between		0.255	0.021	0.912	n = 158
	Within		0.117	-0.221	0.937	T-bar = 33.418
State capacity ×	Overall	0.287	0.268	0.000	0.921	N = 5280
Democracy	Between		0.257	0.001	0.903	n = 158
	Within		0.072	-0.300	0.613	T-bar = 33.418
Power ×	Overall	0.298	0.256	0.000	0.923	N = 5280
Democracy	Between		0.243	0.007	0.907	n = 158
	Within		0.077	-0.258	0.628	T-bar = 33.418
Procedures ×	Overall	0.268	0.261	0.000	0.920	N = 5280
Democracy	Between		0.250	0.002	0.904	n = 158
	Within		0.069	-0.298	0.591	T-bar = 33.418
Ln(GDP/capita)	Overall	8.244	1.515	5.101	11.431	N = 5280
	Between		1.472	5.541	11.245	n = 158
	Within		0.300	6.157	9.879	T-bar = 33.418
Population	Overall	148.234	514.099	1.204	7952.999	N = 5280
density	Between		491.917	1.592	5997.670	n = 158
	Within		112.979	-1771.495	2103.562	T-bar = 33.418

Entries are simple averages across the multiple imputation datasets used in the regression analysis. See Table C7 (Appendix C) for data sources. Democracy is measured with Polyarchy.

Bibliography

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. "The Colonial Origins of Comparative Development." *American Economic Review* 91(5): 1369–1401.

Acemoglu, Daron. 2008. 'Interactions between Governance and Growth: What World Bank Economists Need to Know.' In *Governance, Growth, and Development Decision-Making*, ed. D. Acemoglu, D. North, D, Rodrik, and F. Fukuyama. Washington, DC: World Bank.

Acemoglu, Daron and James A. Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. New York, NY: Crown Publishers.

Acemoglu, Daron, Camilo García-Jimeno, and James A. Robinson. 2015. "State Capacity and Economic Development: A Network Approach." *American Economic Review* 105(8): 2364–2409.

Acemoglu, Daron and James A. Robinson. 2019. *The Narrow Corridor: States, Societies, and the Fate of Liberty*. New York, NY: Penguin Press.

Acemoglu, Daron, James A. Robinson, and Ragnar Torvik. 2020. "The Political Agenda Effect and State Centralization." *Journal of Comparative Economics* 48(4): 749–778.

Adcock, Robert and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95(3): 529–546.

Agnafors, Marcus. 2013. "Quality of Government: Toward a More Complex Definition." *American Political Science Review* 107(3): 433–444.

Albrecht, Peter and Paul Jackson. 2014. "State-Building through Security Sector Reform: The UK Intervention in Sierra Leone." *Peacebuilding* 2(1): 83–99.

Andersen, David, Jørgen Møller, Lasse Lykke Rørbæk, and Svend-Erik Skaaning. 2014. "State Capacity and Political Regime Stability." *Democratization* 21(7): 1305–1325.

Andersen, David and Jonathan Doucette. 2020. "State First? A Disaggregation and Empirical Interrogation." *British Journal of Political Science*. Advance online publication. doi: 10.1017/S0007123420000083.

Andrews, Matt. 2008. "The Good Governance Agenda: Beyond Indicators without Theory." *Oxford Development Studies* 36(4): 379–407.

Andrews, Matt, Roger Hay, and Jerrett Myers. 2010. "Can Governance Indicators Make Sense? Towards a New Approach to Sector-Specific Measures of Governance." *Oxford Development Studies* 38(4): 391–410.

Andrews, Matt, Lant Pritchett, and Michael Woolcock. 2017. *Building State Capability*. New York, NY: Oxford University Press.

Arbetman-Rabinowitz, Marina and Kristin Johnson. 2008. "Power Distribution and Oil in the Sudan: Will the Comprehensive Peace Agreement Turn the Oil Curse into a Blessing?" *International Interactions* 34(4): 382–401.

Arndt, Christiane and Charles Oman. 2006. *Uses and Abuses of Governance Indicators*. Paris: OECD Publishing.

Asadullah, M. Niaz and Antonio Savoia. 2018. "Poverty Reduction during 1990–2013: Did Millennium Development Goals Adoption and State Capacity Matter?" *World Development* 105: 70–82.

Asadullah, M. Niaz, Antonio Savoia, and Kunal Sen. 2020. "Will South Asia Achieve the Sustainable Development Goals by 2030? Learning from the MDGs Experience." *Social Indicators Research* 152(1): 165–189.

Bäck, Hanna and Axel Hadenius. 2008. "Democracy and State Capacity: Exploring a J-Shaped Relationship." *Governance* 21(1): 1–24.

Banfield, Edward C. 1958. *The Moral Basis of a Backward Society*. Glencoe, IL: The Free Press.

Bardhan, Pranab. 2016. "State and Development: The Need for a Reappraisal of the Current Literature." *Journal of Economic Literature* 54(3): 862–892.

Barsh, Russel L. 1993. "Measuring Human Rights: Problems of Methodology and Purpose." *Human Rights Quarterly* 15(1): 87–121.

Bauhr, Monika and Marcia Grimes. 2017. "Transparency to Curb Corruption? Concepts, Measures and Empirical Merit." *Crime, Law and Social Change* 68(4): 431–458.

Bell, Sam R., David Cingranelli, Amanda Murdie, and Alper Caglayan. 2013. "Coercion, Capacity, and Coordination: Predictors of Political Violence." *Conflict Management and Peace Science* 30(3): 240–262.

Beramendi, Pablo, Mark Dincecco, and Melissa Rogers. 2019. "Intra-Elite Competition and Long-Run Fiscal Development." *Journal of Politics* 81(1): 49–65.

Bersch, Katherine and Sandra Botero. 2014. "Measuring Governance: Implications of Conceptual Choices." *European Journal of Development Research* 26(1): 124–141.

Besley, Timothy and Torsten Persson. 2008. "Wars and State Capacity." *Journal of the European Economic Association* 6(2–3): 522–530.

Besley, Timothy and Torsten Persson. 2009. "The Origins of State Capacity: Property Rights, Taxation, and Politics." *American Economic Review* 99(4): 1218–1244.

Besley, Timothy and Torsten Persson. 2011. *Pillars of Prosperity: The Political Economics of Development Clusters*. Princeton, NJ: Princeton University Press.

Besley, Timothy and Torsten Persson. 2014. "The Causes and Consequences of Development Clusters: State Capacity, Peace, and Income." *Annual Review of Economics* 6: 927–949.

Bizzarro, Fernando, John Gerring, Carl Henrik Knutsen, Allen Hicken, Michael Bernhard, Svend Erik Skaaning, Michael Coppedge, and Staffan I. Lindberg. 2018. "Party Strength and Economic Growth." *World Politics* 70(2): 275–320.

Blackwell, Matthew, James Honaker, and Gary King. 2017. "A Unified Approach to Measurement Error and Missing Data: Details and Extensions." *Sociological Methods and Research* 46(3): 342–369.

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.

Boese, Vanessa A. 2019. "How (Not) to Measure Democracy." *International Area Studies Review* 22(2): 95–127.

Böhmelt, Tobias, Vincenzo Bove, and Kristian Skrede Gleditsch. 2019. "Blame the Victims? Refugees, State Capacity, and Non-State Actor Violence." *Journal of Peace Research* 56(1): 73–87.

Bollen, Kenneth A. 1980. "Issues in the Comparative Measurement of Political Democracy." *American Sociological Review* 45(3): 370–390.

Bollen, Kenneth A. 1986. "Political Rights and Political Liberties in Nations: An Evaluation of Human Rights Measures, 1950 to 1984." *Human Rights Quarterly* 8(4): 567–591.

Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York, NY: Wiley

Bollen, Kenneth A. 1993. "Liberal Democracy: Validity and Method Factors in Cross-National Measures." *American Journal of Political Science* 37(4): 1207–1230.

Bollen, Kenneth A. and Pamela Paxton. 2000. "Subjective Measures of Liberal Democracy." *Comparative Political Studies* 33(1): 58–86.

Bollen, Kenneth A. and Adamantios Diamantopoulos. 2017. "In Defense of Causal-Formative Indicators: A Minority Report." *Psychological Methods* 22(3): 581–596.

Bowman, Kirk, Fabrice Lehoucq, and James Mahoney. 2005. "Measuring Political Democracy: Case Expertise, Data Adequacy, and Central America." *Comparative Political Studies* 38(8): 939–970.

Braithwaite, Alex. 2010. "Resisting Infection: How State Capacity Conditions Conflict Contagion." *Journal of Peace Research* 47(3): 311–319.

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.

Bratton, Michael and Eric C. C. Chang. 2006. "State Building and Democratization in Sub-Saharan Africa: Forwards, Backwards, or Together?" *Comparative Political Studies* 39(9): 1059–1083.

Buhaug, Halvard. 2010. "Dude, Where's My Conflict?: LSG, Relative Strength, and the Location of Civil War." *Conflict Management and Peace Science* 27(2): 107–128.

Bustikova, Lenka and Cristina Corduneanu-Huci. 2017. "Patronage, Trust, and State Capacity: The Historical Trajectories of Clientelism." *World Politics* 69(2): 277–326.

Cappelen, Christoffer and Jason Sorens. 2018. "Pre-Colonial Centralisation, Traditional Indirect Rule, and State Capacity in Africa." *Commonwealth and Comparative Politics* 56(2): 195–215.

Carbone, Giovanni. 2015. "Democratisation as a State-Building Mechanism: A Preliminary Discussion of an Understudied Relationship." *Political Studies Review* 13(1): 11–21.

Carbone, Giovanni and Vincenzo Memoli. 2015. "Does Democratization Foster State Consolidation? Democratic Rule, Political Order, and Administrative Capacity." *Governance* 28(1): 5–24.

Cárdenas, Mauricio. 2010. "State Capacity in Latin America." *Economía* 10(2): 1–45.

Carment, David, Joe Landry, Yiagadeesen Samy, and Scott Shaw. 2015. "Towards a Theory of Fragile State Transitions: Evidence from Yemen, Bangladesh and Laos." *Third World Quarterly* 36(7): 1316–1332.

Carothers, Thomas. 2007. "How Democracies Emerge." *Journal of Democracy* 18(1): 11–27.

Casper, Gretchen and Claudiu Tufis. 2003. "Correlation Versus Interchangeability: The Limited Robustness of Empirical Findings on Democracy Using Highly Correlated Data Sets." *Political Analysis* 11(2): 196–203.

Cassese, Sabino. 1986. "The Rise and Decline of the Notion of State." *International Political Science Review* 7(2): 120–130.

Centeno, Miguel A., Atul Kohli and Deborah J. Yashar. 2017. 'Unpacking States in the Developing World: Capacity, Performance, and Politics'. In *States in the Developing World*, ed. M. A. Centeno, A. Kohli, and D. J. Yashar. Cambridge, UK: Cambridge University Press.

Chabova, Kristyna. 2017. "Measuring Corruption in Europe: Public Opinion Surveys and Composite Indices." *Quality and Quantity* 51(4): 1877–1900.

Charron, Nicholas and Victor Lapuente. 2010. "Does Democracy Produce Quality of Government?" *European Journal of Political Research* 49(4): 443–470.

Charron, Nicholas and Victor Lapuente. 2011. "Which Dictators Produce Quality of Government?" *Studies in Comparative International Development* 46(4): 397–423.

Charron, Nicholas. 2016. "Do Corruption Measures Have a Perception Problem? Assessing the Relationship between Experiences and Perceptions of Corruption among Citizens and Experts." *European Political Science Review* 8(1): 147–171.

Child, Dennis. 1975. *The Essentials of Factor Analysis*. London, UK: Holt, Rinehart and Winston.

Cingolani, Luciana. 2013. "The State of State Capacity: A Review of Concepts, Evidence and Measures." *UNU-MERIT Working Paper Series* 2013-053.

Cingolani, Luciana, Kaj Thomsson, and Denis de Crombrugghe. 2015. "Minding Weber More than Ever? The Impacts of State Capacity and Bureaucratic Autonomy on Development Goals." *World Development* 72: 191–207.

Clague, Christopher, Philip Keefer, Stephen Knack, and Mancur Olson. 1999. "Contract-Intensive Money: Contract Enforcement, Property Rights, and Economic Performance." *Journal of Economic Growth* 4(2): 185–211.

Cole, Wade M. 2015. "Mind the Gap: State Capacity and the Implementation of Human Rights Treaties." *International Organization* 69(2): 405–441.

Cole, Wade M. 2016. "Managing to Mitigate Abuse: Bureaucracy, Democracy, and Human Rights, 1984 to 2010." *International Journal of Comparative Sociology* 57(1–2): 69–97.

Cole, Wade M. 2018. "Does Might Make Right or Fight? Coercive Capacity, Democracy, and Human Rights, 1975 to 2010." *Journal of Human Rights* 17(2): 147–162.

Coppedge, Michael, John Gerring, Staffan I. Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, … Jeffrey Staton. 2017. V-Dem Codebook v7. Varieties of Democracy (V-Dem) Project.

Coppedge, Michael, John Gerring, Carl H. Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, … Daniel Ziblatt. 2019. V-Dem Codebook v9. Varieties of Democracy (V-Dem) Project.

Coppedge, Michael, John Gerring, Carl H. Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, … Daniel Ziblatt. 2020. V-Dem Codebook v10. Varieties of Democracy (V-Dem) Project.

Cornell, Agnes, Carl Henrik Knutsen, and Jan Teorell. 2020. "Bureaucracy and Growth." *Comparative Political Studies* 53(14): 2246–2282.

Costello, Anna B. and Jason W. Osborne. 2005. "Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from Your Analysis." *Practical Assessment, Research and Evaluation* 10(7): 1–9.

Cronbach, Lee J. 1951. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika* 16(3): 297–334.

Cronert, Axel and Axel Hadenius. 2020. "Institutional Foundations of Global Well-Being." *International Political Science Review*. Advance online publication. doi:10.1177/0192512120917186.

Curley, Cali, Rachel M. Krause, Richard Feiock, and Christopher V. Hawkins. 2019. "Dealing with Missing Data: A Comparative Exploration of Approaches Using the Integrated City Sustainability Database." *Urban Affairs Review* 55(2): 591–615.

D'Arcy, Michelle and Marina Nistotskaya. 2017. "State First, Then Democracy: Using Cadastral Records to Explain Governmental Performance in Public Goods Provision." *Governance* 30(2): 193–209.

D'Arcy, Michelle and Marina Nistotskaya. 2018. "The Early Modern Origins of Contemporary European Tax Outcomes." *European Journal of Political Research* 57(1): 47–67.

D'Arcy, Michelle and Marina Nistotskaya. 2021. 'State Capacity, Quality of Government, Sequencing, and Development Outcomes.' In *Oxford Handbook of Quality of Government*, ed. B. Rothstein, M. Bauhr, A. Bågenholm, and M. Grimes. Oxford, UK: Oxford University Press.

Dahl, Robert A. 1971. *Polyarchy: Participation and Opposition*. New Haven, CT: Yale University Press.

Dahlström, Carl, Johannes Lindvall, and Bo Rothstein. 2013. "Corruption, Bureaucratic Failure and Social Policy Priorities." *Political Studies* 61(3): 523–542.

Dahlström, Carl, Jan Teorell, Stefan Dahlberg, Felix Hartmann, Annika Lindberg, and Marina Nistotskaya. 2015. The QoG Expert Survey Dataset II. University of Gothenburg: The Quality of Government Institute.

Di John, Jonathan. 2009. *From Windfall to Curse? Oil and Industrialisation in Venezuela, 1920 to the Present*. University Park, PA: Pennsylvania State University Press.

Diamond, Larry and Leonardo Morlino. 2004. "The Quality of Democracy: An Overview." *Journal of Democracy* 15(4): 20–31.

Dimitrova, Antoaneta, Honorata Mazepus, Dimiter Toshkov, Tatsiana Chulitskaya, Natallia Rabava, and Ina Ramasheuskaya. 2020. "The Dual Role of State Capacity in Opening Socio-Political Orders: Assessment of Different Elements of State Capacity in Belarus and Ukraine." *East European Politics*. Advance online publication. doi: 10.1080/21599165.2020.1756783.

Dincecco, Mark and Mauricio Prado. 2012. "Warfare, Fiscal Capacity, and Performance." *Journal of Economic Growth* 17(3): 171–203.

Dincecco, Mark. 2015. "The Rise of Effective States in Europe." *Journal of Economic History* 75(3): 901–918.

Dincecco, Mark and Gabriel Katz. 2016. "State Capacity and Long-Run Economic Performance." *Economic Journal* 126(590): 189–218.

Dincecco, Mark. 2018. *State Capacity and Economic Development. Present and Past.* Cambridge, UK: Cambridge University Press.

Donchev, Dilyan and Gergely Ujhelyi. 2014. "What Do Corruption Indices Measure?" *Economics and Politics* 26(2): 309–331.

Edwards, Jeffrey R. 2011. "The Fallacy of Formative Measurement." *Organizational Research Methods* 14(2): 370–388.

Elff, Martin and Sebastian Ziaja. 2018. "Method Factors in Democracy Indicators." *Politics and Governance* 6(1): 92–104.

Elkins, Zachary. 2000. "Gradations of Democracy? Empirical Tests of Alternative Conceptualizations." *American Journal of Political Science* 44(2): 293–300.

Englebert, Pierre and Denis M. Tull. 2008. "Postconflict Reconstruction in Africa: Flawed Ideas about Failed States." *International Security* 32(4): 106–139.

Englehart, Neil A. 2009. "State Capacity, State Failure, and Human Rights." *Journal of Peace Research* 46(2): 163–180.

Evans, Peter B. 1985. 'Transnational Linkages and the Role of the State.' In *Bringing the State Back In*, ed. P. B. Evans, D. Rueschemeyer, and T. Skocpol. Cambridge, UK: Cambridge University Press.

Evans, Peter. 1995. *Embedded Autonomy: States and Industrial Transformation*. Princeton, NJ: Princeton University Press.

Evans, Peter and James E. Rauch. 1999. "Bureaucracy and Growth: A Cross-National Analysis of the Effects of 'Weberian' State Structures on Economic Growth." *American Sociological Review* 64(5): 748–765.

Fabra Mata, J. and S. Ziaja. 2009. *Users' Guide on Measuring Fragility*. Bonn: German Development Institute.

Fabrigar, Leandre R. and Duane T. Wegener. 2012. *Exploratory Factor Analysis*. New York, NY: Oxford University Press.

Fearon, James D. and David D. Laitin. 2003. "Ethnicity and Civil War." *American Political Science Review* 97(1): 75–90.

Fisunoglu, Ali, Kyungkook Kang, Marina Arbetman-Rabinowitz, and Jacek Kugler. 2011. "Relative Political Capacity Dataset (Version 2.4) August 2020". Harvard Dataverse V7.

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.

Fortin, Jessica. 2010. "A Tool to Evaluate State Capacity in Post-Communist Countries, 1989-2006." *European Journal of Political Research* 49(5): 654–686.

Fortin, Jessica. 2012. "Is There a Necessary Condition for Democracy? The Role of State Capacity in Postcommunist Countries." *Comparative Political Studies* 45(7): 903–930.

Fortin-Rittberger, Jessica. 2014a. "Exploring the Relationship between Infrastructural and Coercive State Capacity." *Democratization* 21(7): 1244–1264.

Fortin-Rittberger, Jessica. 2014b. "The Role of Infrastructural and Coercive State Capacity in Explaining Different Types of Electoral Fraud." *Democratization* 21(1): 95–117.

Fund for Peace. 2019. *Fragile States Index Annual Report 2019*. Washington, DC: Fund for Peace.

Fukuyama, Francis. 2004. "The Imperative of State-Building." *Journal of Democracy* 15(2): 17–31.

Fukuyama, Francis. 2005. "'Stateness' First." *Journal of Democracy* 16(1): 84–88.

Fukuyama, Francis. 2013. "What Is Governance?" *Governance* 26(3): 347–368.

Fukuyama, Francis. 2014. "States and Democracy." *Democratization* 21(7): 124–138.

Fukuyama, Francis. 2020. "The Pandemic and Political Order." *Foreign Affairs* 99(4): 26–32.

Gan, Natalia. 2020. "Rule of Law Crisis, Militarization of Citizen Security, and Effects on Human Rights in Venezuela." *European Review of Latin American and Caribbean Studies* 109(2020): 67–86.

Geddes, Barbara. 1994. *Politician's Dilemma: Building State Capacity in Latin America.* Berkeley, CA: University of California Press.

Geloso, Vincent J. and Alexander W. Salter. 2020. "State Capacity and Economic Development: Causal Mechanism or Correlative Filter?" *Journal of Economic Behavior and Organization* 170: 372–385.

Gibler, Douglas M. and Steven V. Miller. 2014. "External Territorial Threat, State Capacity, and Civil War." *Journal of Peace Research* 51(5): 634–646.

Giebler, Heiko. 2012. "Bringing Methodology (Back) in: Some Remarks on Contemporary Democracy Measurements." *European Political Science* 11(4): 509–518.

Gisselquist, Rachel M. 2014. "Developing and Evaluating Governance Indexes: 10 Questions." *Policy Studies* 35(5): 513–531.

Gjerlow, Haakon, Carl H. Knutsen, Tore Wig, Matthew C. Wilson. 2018. "Stairways to Denmark: Does the Sequence of State-Building and Democratization Matter for Economic Development?" *The Varieties of Democracy Institute Working Paper Series* 2018-72.

Gleditsch, Kristian S. and Michael D. Ward. 1997. "Double Take: A Reexamination of Democracy and Autocracy in Modern Polities." *Journal of Conflict Resolution* 41(3): 361–383.

Goertz, Gary. 2020. *Social Science Concepts and Measurement: New and Completely Revised Edition*. Princeton, NJ: Princeton University Press.

Goretzko, David, Christian Heumann, and Markus Bühner. 2020. "Investigating Parallel Analysis in the Context of Missing Data: A Simulation Study Comparing Six Missing Data Methods." *Educational and Psychological Measurement* 80(4): 756–774.

Graham, Benjamin A. T. and Jacob R. Tucker. 2019. "The International Political Economy Data Resource." *Review of International Organizations* 14(1): 149–161.

Grassi, Davide and Vincenzo Memoli. 2016a. "Political Determinants of State Capacity in Latin America." *World Development* 88: 94–106.

Grassi, Davide and Vincenzo Memoli. 2016b. "Democracy, Political Partisanship, and State Capacity in Latin America." *Rivista Italiana Di Scienza Politica* 46(1): 47–69.

Grävingholt, Jörn, Sebastian Ziaja, and Merle Kreibaum. 2015. "Disaggregating State Fragility: A Method to Establish a Multidimensional Empirical Typology." *Third World Quarterly* 36(7): 1281–1298.

Grindle, Merilee S. and Mary E. Hilderbrand. 1995. "Building Sustainable Capacity in the Public Sector: What Can Be Done?" *Public Administration and Development* 15(5): 441–463.

Grundholm, Alexander Taaning and Matilde Thorsen. 2019. "Motivated and Able to Make a Difference? The Reinforcing Effects of Democracy and State Capacity on Human Development." *Studies in Comparative International Development* 54(3): 381–414.

Ha, Eunyoung and Melissa Rogers. 2017. "What's Left to Tax? Partisan Reallocation of Trade Taxation in Less Developed Countries." *Political Research Quarterly* 70(3): 495–508.

Hadenius, Axel and Jan Teorell. 2005. "Assessing Alternative Indices of Democracy." *IPSA Committee on Concepts and Methods Working Papers* 6 (August 2005), www.concepts-methods.org/WorkingPapers/PoliticalConcepts.

Halleröd, Björn, Bo Rothstein, Adel Daoud, and Shailen Nandy. 2013. "Bad Governance and Poor Children: A Comparative Analysis of Government Efficiency and Severe Child Deprivation in 68 Low- and Middle-Income Countries." *World Development* 48: 19–31.

Hanson, Jonathan K., and Rachel Sigman. 2013. "Leviathan's Latent Dimensions: Measuring State Capacity for Comparative Political Research." Unpublished Manuscript. Version: September 2013.

Hanson, Jonathan K. 2015. "Democracy and State Capacity: Complements or Substitutes?" *Studies in Comparative International Development* 50(3): 304–330.

Hanson, Jonathan K. 2018. "State Capacity and the Resilience of Electoral Authoritarianism: Conceptualizing and Measuring the Institutional Underpinnings of Autocratic Power." *International Political Science Review* 39(1): 17–32.

Hanson, Jonathan K., and Rachel Sigman. 2020. "Leviathan's Latent Dimensions: Measuring State Capacity for Comparative Political Research." *Journal of Politics*. Advance online publication. doi:10.1086/715066.

Harman, Harry H. and Wayne H. Jones. 1966. "Factor Analysis by Minimizing Residuals (Minres)." *Psychometrika* 31(3): 351–368.

Hellmann, Olli. 2018. "High Capacity, Low Resilience: The 'Developmental' State and Military–Bureaucratic Authoritarianism in South Korea." *International Political Science Review* 39(1): 67–82.

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.

Hendrix, Cullen S. and Joseph K. Young. 2014. "State Capacity and Terrorism: A Two-Dimensional Approach." *Security Studies* 23(2): 329–363.

Hiilamo, Heikki and Stanton A. Glantz. 2015. "Implementation of Effective Cigarette Health Warning Labels among Low and Middle Income Countries: State Capacity, Path-Dependency and Tobacco Industry Activity." *Social Science and Medicine* 124: 241–245.

Högström, John. 2013. "Does the Choice of Democracy Measure Matter? Comparisons between the Two Leading Democracy Indices, Freedom House and Polity IV." *Government and Opposition* 48(2): 201–221.

Holmberg, Sören, Bo Rothstein, and Naghmeh Nasiritousi. 2009. "Quality of Government: What You Get." *Annual Review of Political Science* 12(1): 135–161.

Holmberg, Sören and Bo Rothstein. 2011. "Dying of Corruption." *Health Economics, Policy, and Law* 6(4): 529–547.

Honaker, J. and G. King. 2010. "What to Do About Missing Values in Time-Series Cross-Section Data." *American Journal of Political Science* 54(2): 561–581.

Honaker, J., G. King, and M. Blackwell. 2011. "Amelia II: A Program for Missing Data." *Journal of Statistical Software* 45(7): 1–47.

Hulme, David, Antonio Savoia, and Kunal Sen. 2015. "Governance as a Global Development Goal? Setting, Measuring and Monitoring the Post-2015 Development Agenda." *Global Policy* 6(2): 85–96.

Huntington, Samuel P. 1968. *Political Order in Changing Societies*. New Haven, CT: Yale University Press.

Ikenberry, G. John. 1986. "The Irony of State Strength: Comparative Responses to the Oil Shocks in the 1970s." *International Organization* 40(1): 105–137.

Jerven, Morten. 2013. *Poor Numbers*. Ithaca, NY: Cornell University Press.

Johnson, Noel D. and Mark Koyama. 2017. "States and Economic Growth: Capacity and Constraints." *Explorations in Economic History* 64(2016): 1–20.

Jolliffe, Ian T. 2002. *Principal Component Analysis, Second Edition*. New York, NY: Springer.

Joshi, Devin. 2011. "Good Governance, State Capacity, and the Millennium Development Goals." *Perspectives on Global Development and Technology* 10(2): 339–360.

Joshi, Devin, Barry B. Hughes, and Timothy D. Sisk. 2015. "Improving Governance for the Post-2015 Sustainable Development Goals: Scenario Forecasting the Next 50 Years." *World Development* 70: 286–302.

Kaiser, Henry F. 1974. "An Index of Factorial Simplicity." *Psychometrika* 39(1): 31–36.

Kalinowski, Thomas. 2008. "Korea's Recovery since the 1997/98 Financial Crisis: The Last Stage of the Developmental State." *New Political Economy* 13(4): 447–462.

Kalinowski, Thomas. 2016. "Trends and Mechanisms of Corruption in South Korea." *Pacific Review* 29(4): 625–645.

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2011. "The Worldwide Governance Indicators: Methodology and Analytical Issues." *Hague Journal on the Rule of Law* 3(2): 220–246.

Kim, Nam K. and Alex M. Kroeger. 2018. "Do Multiparty Elections Improve Human Development in Autocracies?" *Democratization* 25(2): 251–272.

King, Gary. 1995. "Replication, Replication." *PS: Political Science and Politics* 28(3): 444–452.

Knack, Stephen, Mark Kugler, and Nick Manning. 2003. "Second-Generation Governance Indicators." *International Review of Administrative Sciences* 69(3): 345–364.

Knack, Stephen. 2007. "Measuring Corruption: A Critique of Indicators in Eastern Europe and Central Asia." *Journal of Public Policy* 27(3): 255–291.

Knutsen, Carl H. 2010. "Measuring Effective Democracy." *International Political Science Review* 31(2): 109–128.

Knutsen, Carl H. 2013. "Democracy, State Capacity, and Economic Growth." *World Development* 43(2012): 1–18.

Knutsen, Carl Henrik and Hanne Fjelde. 2013. "Property Rights in Dictatorships: Kings Protect Property Better than Generals or Party Bosses." *Contemporary Politics* 19(1): 94–114.

Koren, Ore and Anoop K. Sarbahi. 2018. "State Capacity, Insurgency, and Civil War: A Disaggregated Analysis." *International Studies Quarterly* 62(2): 274–888.

Krasner, Stephen D. 1984. "Review: Approaches to the State: Alternative Conceptions and Historical Dynamics." *Comparative Politics* 15(2): 223–246.

Kugler, Jacek and Ronald L. Tammen (Eds.). 2012. *The Performance of Nations*. Lanham, MD: Rowman and Littlefield Publishers.

Kurtz, Marcus J. and Andrew Schrank. 2012. "Capturing State Strength." *Revista De Ciencia Política* 32(3): 613–621.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny. 1999. "The Quality of Government." *Journal of Law, Economics, and Organization* 15(1): 222–279.

Landman, Todd. 2002. "Comparative Politics and Human Rights." *Human Rights Quarterly* 24(4): 890–923.

Landman, Todd. 2004. "Measuring Human Rights: Principle, Practice, and Policy." *Human Rights Quarterly* 26(4): 906–931.

Larsson, Tomas. 2013. "The Strong and the Weak: Ups and Downs of State Capacity in Southeast Asia." *Asian Politics and Policy* 5(3): 337–358.

Lee, Melissa M. and Nan Zhang. 2017. "Legibility and the Informational Foundations of State Capacity." *The Journal of Politics* 79(1): 118–132.

Levi, Margaret and Douglass C. North. 1982. "Toward a Property-Rights Theory of Exploitation." *Politics & Society* 11(3): 315–320.

Levi, Margaret. 1988. *Of Rule and Revenue*. Berkeley, CA: University of California Press.

Lin, Thung Hong. 2015. "Governing Natural Disasters: State Capacity, Democracy, and Human Vulnerability." *Social Forces* 93(3): 1267–1300.

Lindvall, Johannes and Jan Teorell. 2016. "State Capacity as Power: A Conceptual Framework." *STANCE Working Paper Series* 2016(1): 1–32.

Linz, Juan J., and Alfred Stepan. 1996. *Democratic transition and consolidation: Southern Europe, South America, and Post-Communist Europe*. Baltimore, MD: The Johns Hopkins University Press.

Little, Roderick J. A. and Donald B. Rubin. 2002. *Statistical Analysis with Missing Data.* Hoboken, NJ: Wiley.

Lorenzo-Seva, Urbano and Joost R. Van Ginkel. 2016. "Multiple Imputation of Missing Values in Exploratory Factor Analysis of Multidimensional Scales: Estimating Latent Trait Scores." *Anales de Psicología* 32(2): 596–608.

Luiz, J. M. 1999. "The State and Industrialisation in Korea and Brazil." *South African Journal of Economics* 67(2): 91–102.

Luna, Juan Pablo and Hillel David Soifer. 2017. "Capturing Sub-National Variation in State Capacity: A Survey-Based Approach." *American Behavioral Scientist* 61(8): 887–907.

Maggino, Filomena. 2017. 'Developing Indicators and Managing the Complexity.' In *Complexity in Society: From Indicators Construction to their Synthesis*, ed. F. Maggino. Cham: Springer.

Mazzuca, Sebastián L. and Gerardo L. Munck. 2014. "State or Democracy First? Alternative Perspectives on the State-Democracy Nexus." *Democratization* 21(7): 1221–1243.

Mann, Michael. 1984. "The Autonomous Power of the State: Its Origins, Mechanisms and Results." *European Journal of Sociology* 25(2): 185–213.

Mann, Michael. 1986. *The Sources of Social Power Volume I: A History of Power from the Beginning to A.D. 1760.* New York, NY: Cambridge University Press.

Mann, Michael. 2008. "Infrastructural Power Revisited." *Studies in Comparative International Development* 43(3–4): 355–365.

Mansour, Mario. 2014. "A Tax Revenue Dataset for Sub-Saharan Africa: 1980-2010." *Ferdi Working Papers* I19(July 2014): 1–26.

Marshall, Monty G. and Gabrielle Elzinga-Marshall. 2017. *Global Report 2017: Conflict, Governance, and State Fragility.* Vienna, VA: Center for Systemic Peace.

McAdam, Doug, Sidney Tarrow, and Charles Tilly. 2001. *Dynamics of Contention.* New York, NY: Cambridge University Press.

McNeish, Daniel. 2017. "Exploratory Factor Analysis with Small Samples and Missing Data." *Journal of Personality Assessment* 99(6): 637–652.

Melville, Andrei and Mikhail Mironyuk. 2016. "Bad Enough Governance: State Capacity and Quality of Institutions in Post-Soviet Autocracies." *Post-Soviet Affairs* 32(2): 132–151.

Migdal, Joel S. 1988a. *Strong Societies and Weak States. State-Society Relations and State Capabilities in the Third World.* Princeton, NJ: Princeton University Press.

Migdal, Joel S. 1988b. "Vision and Practice: The Leader, the State, and the Transformation of Society." *International Political Science Review* 9(1): 23–41.

Møller, Jørgen and Svend-Erik Skaaning. 2011a. "On the Limited Interchangeability of Rule of Law Measures." *European Political Science Review* 3(3): 371–394.

Møller, Jørgen and Svend Erik Skaaning. 2011b. "Stateness First?" *Democratization* 18(1): 1–24.

Møller, Jørgen and Svend-Erik Skaaning. 2014. The Rule of Law: Definitions, Measures, Patterns and Causes. New York, NY: Palgrave Macmillian.

Morlino, Leonardo. 2012. *Changes for Democracy: Actors, Structures, Processes*. Oxford, UK: Oxford University Press.

Mudde, Cas and Andreas Schedler. 2010. "Introduction: Rational Data Choice." *Political Research Quarterly* 63(2): 410–416.

Munck, Gerardo L. and Jay Verkuilen. 2002. "Conceptualizing and Measuring Democracy." *Comparative Political Studies* 35(1): 5–34.

Mungiu-Pippidi, Alina. 2006. "Corruption: Diagnosis and Treatment." *Journal of Democracy* 17(3): 86–99.

Murphy, Ryan H. 2019. "Liberalizing, State Building, and Getting to Denmark: Analyzing 21st-Century Institutional Change." *Journal of Institutional Economics* 15(6): 933–950.

Nassiri, Vahid, Anikó Lovik, Geert Molenberghs, and Geert Verbeke. 2018. "On Using Multiple Imputation for Exploratory Factor Analysis of Incomplete Data." *Behavior Research Methods* 50(2): 501–517.

Nettl, John P. 1968. "The State as a Conceptual Variable." *World Politics* 20(4): 559–592.

Neumayer, Eric, and Thomas Plümper. 2017. *Robustness Tests for Quantitative Research*. Cambridge, UK: Cambridge University Press.

Newman, Daniel A. 2014. "Missing Data: Five Practical Guidelines." *Organizational Research Methods* 17(4): 372–411.

Nistotskaya, Marina and Luciana Cingolani. 2016. "Bureaucratic Structure, Regulatory Quality, and Entrepreneurship in a Comparative Perspective: Cross-Sectional and Panel Data Evidence." *Journal of Public Administration Research and Theory* 26(3): 519–534.

Norris, Pippa. 2012. *Making Democratic Governance Work: How Regimes Shape Prosperity, Welfare, and Peace*. New York, NY: Cambridge University Press.

North, Douglass C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge, UK: Cambridge University Press.

O'Donnell, Guillermo. 1993. "On the State, Democratization and Some Conceptual Problems." *World Development* 21(8): 1355–1369.

OECD. 2008. *Handbook on Constructing Composite Indicators: Methodology and User Guide*. Paris: OECD Publishing.

Olson, Mancur. 1993. "Dictatorship, Democracy, and Development." *The American Political Science Review* 87(3): 567–576.

Osborne, Jason W. 2015. "What Is Rotating in Exploratory Factor Analysis?" *Practical Assessment, Research and Evaluation* 20(2): 1–7.

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.

Pemstein, Daniel, Kyle L. Marquardt, Eitan Tzelgov, Yi-ting Wang, Juraj Medzihorsky, Joshua Krusell, Farhad Miri, and Johannes von Römer. 2019. "The V-Dem Measurement Model: Latent Variable Analysis for Cross-National and Cross-Temporal Expert-Coded Data." *The Varieties of Democracy Institute Working Paper Series* 2019-21 (4th edition).

Persson, Tove Ahlbom and Marina Povitkina. 2017. "'Gimme Shelter': The Role of Democracy and Institutional Quality in Disaster Preparedness." *Political Research Quarterly* 70(4): 833–847.

Pickel, Susanne, Toralf Stark, and Wiebke Breustedt. 2015. "Assessing the Quality of Quality Measures of Democracy: A Theoretical Framework and Its Empirical Application." *European Political Science* 14(4): 496–520.

Popper, Karl. 2005. *The Logic of Scientific Discovery*. London, UK: Routledge. (Original work published in 1959).

Povitkina, Marina and Ketevan Bolkvadze. 2019. "Fresh Pipes with Dirty Water: How Quality of Government Shapes the Provision of Public Goods in Democracies." *European Journal of Political Research* 58(4): 1191–1212.

Prichard, Wilson. 2016. "Reassessing Tax and Development Research: A New Dataset, New Findings, and Lessons for Research." *World Development* 80: 48–60.

Putnam, Robert D. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.

Qu, Guangjun, Bob Slagter, Kevin Sylwester, and Kyle Doiron. 2019. "Explaining the Standard Errors of Corruption Perception Indices." *Journal of Comparative Economics* 47(4): 907–920.

Reed, W. Robert and Min Zhu. 2017. "On Estimating Long-Run Effects in Models with Lagged Dependent Variables." *Economic Modelling* 64(2017): 302–311.

Ricciuti, Roberto, Antonio Savoia, and Kunal Sen. 2019. "How Do Political Institutions Affect Fiscal Capacity? Explaining Taxation in Developing Economies." *Journal of Institutional Economics* 15(2): 351–380.

Rice, Susan and Stewart Patrick. 2008. *Index of State Weakness in the Developing World*. Washington, DC: Brookings Institution.

Rogers, Melissa Ziegler and Nicholas Weller. 2014. "Income Taxation and the Validity of State Capacity Indicators." *Journal of Public Policy* 34(2): 183–206.

Rosales, Antulio. 2019. "Statization and Denationalization Dynamics in Venezuela's Artisanal and Small Scale-Large-Scale Mining Interface." *Resources Policy* 63(2019): 101422.

Rossignoli, Domenico. 2016. "Democracy, State Capacity and Civil Wars: A New Perspective." *Peace Economics, Peace Science and Public Policy* 22(4): 427–437.

Rotberg, Robert I. 2014. "Good Governance Means Performance and Results." *Governance* 27(3): 511–518.

Rothstein, Bo and Jan Teorell. 2008. "What Is Quality of Government? A Theory of Impartial Government Institutions." *Governance* 21(2): 165–190.

Rothstein, Bo. 2011. *The Quality of Government: Corruption, Social Trust, and Inequality in International Perspective*. Chicago, IL: University of Chicago Press.

Rothstein, Bo and Jan Teorell. 2012. 'Defining and Measuring Quality of Government.' In *Good Government*, ed. S. Holmberg and B. Rothstein. Cheltenham, UK: Edward Elgar.

Rothstein, Bo. 2014. "What Is the Opposite of Corruption?" *Third World Quarterly* 35(5): 737–752.

Rothstein, Bo, Marcus Samanni, and Jan Teorell. 2012. "Explaining the Welfare State: Power Resources vs. the Quality of Government." *European Political Science Review* 4(1): 1–28.

Rouyer, Alwyn R. 1987. "Political Capacity and the Decline of Fertility in India." *American Political Science Review* 81(2): 453–470.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York, NY: Wiley.

Rudolfsen, Ida. 2017. "State Capacity, Inequality and Inter-Group Violence in Sub-Saharan Africa: 1989–2011." *Civil Wars* 19(2): 118–145.

Savoia, Antonio and Kunal Sen. 2015. "Measurement, Evolution, Determinants, and Consequences of State Capacity: A Review of Recent Research." *Journal of Economic Surveys* 29(3): 441–458.

Saylor, Ryan. 2013. "Concepts, Measures, and Measuring Well: An Alternative Outlook." *Sociological Methods and Research* 42(3): 354–391.

Schafer, Joseph L. and John W. Graham. 2002. "Missing Data: Our View of the State of the Art." *Psychological Methods* 7(2): 147–177.

Schlenkrich, Oliver, Lukas Lemm, and Christoph Mohamad-Klotzbach. 2016. "The Contextualized Index of Statehood (CIS): Assessing the Interaction between Contextual Challenges and the Organizational Capacities of States." *Zeitschrift Fur Vergleichende Politikwissenschaft* 10(3–4): 241–272.

Schmidt, Stefan. 2009. "Shall We Really Do It Again? The Powerful Concept of Replication Is Neglected in the Social Sciences." *Review of General Psychology* 13(2): 90–100.

Scott, James C. 1998. *Seeing Like a State*. New Haven, CT: Yale University Press.

Seawright, Jason and David Collier. 2014. "Rival Strategies of Validation: Tools for Evaluating Measures of Democracy." *Comparative Political Studies* 47(1): 111–138.

Seeberg, Merete Bech. 2018. "Electoral Authoritarianism and Economic Control." *International Political Science Review* 39(1): 33–48.

Serikbayeva, Balzhan, Kanat Abdulla, and Yessengali Oskenbayev. 2020. "State Capacity in Responding to COVID-19." *International Journal of Public Administration*. Advance online publication. doi:10.1080/01900692.2020.1850778.

Shih, Victor, Christopher Adolph, and Mingxing Liu. 2012. "Getting Ahead in the Communist Party: Explaining the Advancement of Central Committee Members in China." *American Political Science Review* 106(1): 166–187.

Sikkink, Kathryn. 1991. *Ideas and Institutions: Developmentalism in Brazil and Argentina*. Ithaca, NY: Cornell University Press.

Skaaning, Svend-Erik. 2009. "Measuring Civil Liberty: An Assessment of Standards-Based Data Sets." *Revista de Ciencia Politica* 29(3): 721–740.

Skaaning, Svend-Erik. 2010. "Measuring the Rule of Law." *Political Research Quarterly* 63(2): 449–460.

Skaaning, Svend-Erik. 2018. "Different Types of Data and the Validity of Democracy Measures." *Politics and Governance* 6(1): 105–116.

Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. New York, NY: Cambridge University Press.

Skocpol, Theda. 1985. 'Bringing the State Back In: Strategies of Analysis in Current Research.' In *Bringing the State Back In*, ed. P. B. Evans, D. Rueschemeyer, and T. Skocpol. Cambridge, UK: Cambridge University Press.

Skocpol, Theda. 2008. "Bringing the State Back In: Retrospect and Prospect." *Scandinavian Political Studies* 31(2): 109–124.

Soifer, Hillel D. 2012. "The State of the State in Latin America." *Revista De Ciencia Política* 32(3): 585–598.

Soifer, Hillel D. 2013. "State Power and the Economic Origins of Democracy." *Studies in Comparative International Development* 48(1): 1–22.

Sriram, Chandra L. 2011. '(Re)building the Rule of Law in Sierra Leone: Beyond the Formal Sector?' In *Peacebuilding and Rule of Law in Africa*, ed. C. L. Sriram, O. Martin-Ortega and J. Herman. New York, NY: Routledge.

Stockemer, Daniel. 2014. "Regime Type and Good Governance in Low and High Income States: What Is the Empirical Link?" *Democratization* 21(1): 118–136.

Tabachnick, Barbara G. and Linda S. Fidell. 2013. *Using Multivariate Statistics*. Boston: Pearson.

Tabellini, Guido. 2008. "Institutions and Culture." *Journal of the European Economic Association* 16(2–3): 255–294.

Ten Berge, Jos M. F., Wim P. Krijnen, Tom Wansbeek, and Alexander Shapiro. 1999. "Some New Results on Correlation-Preserving Factor Scores Prediction Methods." *Linear Algebra and Its Applications* 289(1–3): 311–318.

Teorell, Jan, Sören Holmberg, and Bo Rothstein. 2008. The Quality of Government Dataset, version 15May08. University of Gothenburg: The Quality of Government Institute.

Teorell, Jan, Stefan Dahlberg, Sören Holmberg, Bo Rothstein, Natalia Alvarado Pachon, and Richard Svensson. 2019. The QoG Standard Dataset 2019. University of Gothenburg: The Quality of Government Institute.

Teorell, Jan, Michael Coppedge, Staffan Lindberg, and Svend Erik Skaaning. 2019. "Measuring Polyarchy Across the Globe, 1900–2017." *Studies in Comparative International Development* 54(1): 71–95.

Thomas, Melissa A. 2010. "What Do the Worldwide Governance Indicators Measure." *European Journal of Development Research* 22(1): 31–54.

Tikuisis, Peter, David Carment, Yiagadeesen Samy, and Joseph Landry. 2015. "Typology of State Types: Persistence and Transition." *International Interactions* 41(3): 565–582.

Tilly, Charles. 1975. 'Reflections on the History of European State-Making.' In *The Formation of National States in Western Europe*. Ed. C. Tilly. Princeton, NJ: Princeton University Press.

Tilly, Charles. 1992. *Coercion, Capital, and European States, AD 990-1992*. Cambridge, MA: Blackwell.

Tilly, Charles. 2007. *Democracy*. Cambridge, UK: Cambridge University Press.

Transparency International. 2019. "Corruption Perceptions Index 2019: Frequently Asked Questions." https://www.transparency.org/en/cpi.

Trapnell, Stephanie E. 2011. "Actionable Governance Indicators: Turning Measurement into Reform." *Hague Journal on the Rule of Law* 3(2): 317–348.

Treisman, Daniel. 2007. "What Have We Learned About the Causes of Corruption from Ten Years of Cross-National Empirical Research?" *Annual Review of Political Science* 10(1): 211–244.

United Nations. 2019. Report of the Secretary-General on the Peacebuilding Fund. UN General Assembly A/73/829.

Vaccaro, Andrea. 2021. "Comparing Measures of Democracy: Statistical Properties, Convergence, and Interchangeability." *European Political Science*. Advance online publication. doi:10.1057/s41304-021-00328-8.

Van Ginkel, Joost R., Marielle Linting, Ralph C. A. Rippe, and Anja van der Voort. 2020. "Rebutting Existing Misconceptions About Multiple Imputation as a Method for Handling Missing Data." *Journal of Personality Assessment* 102(3): 297–308.

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.

Versteeg, Mila and Tom Ginsburg. 2017. "Measuring the Rule of Law: A Comparison of Indicators." *Law and Social Inquiry* 42(1): 100–137.

Walther, Daniel, Johan Hellström, and Torbjörn Bergman. 2019. "Government Instability and the State." *Political Science Research and Methods* 7(3): 579–594.

Wang, Erik H. and Yiqing Xu. 2018. "Awakening Leviathan: The Effect of Democracy on State Capacity." *Research and Politics* 5(2): 1–7.

Weber, Max. 1946. "Politics as a vocation." In *From Max Weber: Essays in Sociology*, ed. H. H. Gerth and C. W. Mills. New York, NY: Oxford University Press.

Weber, Max. 1978. *Economy and Society*. Berkeley, CA: University of California Press.

Wilkins, Arjun S. 2017. "To Lag or Not to Lag? Re-Evaluating the Use of Lagged Dependent Variables in Regression Analysis." *Political Science Research and Methods* 6(2): 1–19.

Williams, Martin J. 2021. "Beyond State Capacity: Bureaucratic Performance, Policy Implementation and Reform." *Journal of Institutional Economics* 17(2): 339–357.

Woo, J. J. 2020. "Policy Capacity and Singapore's Response to the COVID-19 Pandemic." *Policy and Society* 39(3): 345–362.

World Bank. 1992. *Governance and Development*. Washington, DC: The World Bank.

World Bank. 1994. *Governance: The World Bank's Experience*. Washington, DC: The World Bank.

World Bank. 1997. *World Development Report 1997: The State in a Changing World*. New York, NY: Oxford University Press.

World Bank. 2019. *World Development Indicators [January 2019]*. The World Bank. https://datatopics.worldbank.org/world-development-indicators.

World Bank. 2020. *World Development Indicators [October 2020]*. The World Bank. https://datatopics.worldbank.org/world-development-indicators.

You, Jong-Sung. 2017. 'South Korea: The Odissey to Corruption Control.' In *Transitions to Good Governance: Creating Virtuous Circles of Anti-Corruption*, ed. A. Mungiu-Pippidi and M. Johnston. Cheltenham, UK: Elgar.

Zeller, Richard A. and Edward G. Carmines. 1980. *Measurement in the Social Sciences*. New York, NY: Cambridge University Press.

Ziaja, Sebastian, Jörn Grävingholt, and Merle Kreibaum. 2019. "Constellations of Fragility: An Empirical Typology of States." *Studies in Comparative International Development* 54(2): 299–321.