



UNIVERSITY OF
GOTHENBURG

MEASURES OF STATE CAPACITY:

Same same, but different?

ANDREA VACCARO

WORKING PAPER SERIES 2020:9

QOG THE QUALITY OF GOVERNMENT INSTITUTE

Department of Political Science

University of Gothenburg

Box 711, SE 405 30 GÖTEBORG

September 2020

ISSN 1653-8919

© 2020 by Andrea Vaccaro. All rights reserved.

Measures of State Capacity: Same Same, but Different?
Andrea Vaccaro
QoG Working Paper Series 2020:9
September 2020
ISSN 1653-8919

ABSTRACT

This study provides a systematic comparative analysis of seven established cross-national measures of state capacity by focusing on three measurement issues: validity, interchangeability, and rating discrepancy. The author finds that the association and convergent validity of the measures is high, but the interchangeability of the measures is low. Through the weak external validity of three replicated longitudinal studies the author demonstrates that statistical differences in measures can have considerable consequences for empirical results. The cause of these somewhat counterpoising findings lies in strikingly high rating discrepancy within some individual countries. The author finds that this rating discrepancy depends systematically on the level of state capacity. No measure of state capacity seems to be clearly superior to others, but future studies should ensure that a given definition of state capacity matches with the chosen measure and should make clear whether the findings are generalizable or not.

Andrea Vaccaro

Department of Social Sciences and Economics
Sapienza University of Rome
andrea.vaccaro@uniroma1.it

Introduction

The concept of state capacity has begun to play a key role in many social science subfields. Despite some definitional disagreements, many scholars agree that state capacity has to do at minimum with the ability of the state to execute policies (e.g., Skocpol 1985; Fukuyama 2004; Dinecco 2017). Additionally, some scholars see the ability of the state to penetrate society (Mann 1984; Migdal 1988), to provide public goods (Norris 2012), to extract revenues (Levi 1988), to deliver well-being (Besley and Persson 2011), or to control economic resources (Evans 1985) as constituent characteristics of state capacity. Procedural definitions see impartiality (Rothstein and Teorell 2008), and efficiency and absence of corruption (Charron and Lapuente 2010, 2011) as fundamental features of state capacity.

Cross-national empirical work has associated state capacity to various social, political, and economic issues. To give some examples, as to the causes of state capacity, it has been shown that state capacity is affected by civil wars (Besley and Persson 2008), democracy (Bäck and Hadenius 2008; Charron and Lapuente 2010; Carbone and Memoli 2015; Memoli and Grassi 2016), and constraints on the executive (Ricciuti, Savoia, and Sen 2019). As to the consequences of state capacity, it has been shown that state capacity affects positively the provision of human rights (Englehart 2009), economic growth (Evans and Rauch 1999; Dinecco 2015), welfare state generosity (Rothstein, Samanni, and Teorell 2012), public goods (Hanson 2015; D'Arcy and Nistotskaya 2017), government stability (Walther, Hellström, and Bergman 2019), and Millennium Development Goals (Joshi 2011; Cingolani, Thomsson, and De Crombrugge 2015). Moreover, low capacity states are more likely to have civil wars (Fearon and Laitin 2003) and less likely to be democratic (Fortin 2012) or equal (Soifer 2013).

Despite proliferating quantitative work on the topic, the statistical analysis of measures of state capacity remains overlooked. Hendrix (2010), Cingolani (2013), and Savoia and Sen (2015) review some of the measures of state capacity, but measures of other broadly related concepts such as democracy (e.g., Knutsen 2010; Teorell and Lindstedt 2010; Högström 2014), and rule of law (Skaaning 2010; Møller and Skaaning 2011a, 2014) have been analysed and compared more comprehensively. The study at hand fills this gap in literature and provides a systematic statistical comparison of state capacity measures, by focusing mainly on three specific measurement issues: validity, interchangeability, and rating discrepancy.

A comparative statistical analysis of measures of state capacity is a valuable task per se, because our empirical knowledge about the similarities, divergencies, and possible shortcomings of these measures is limited. Anyhow, such an analysis has also major implications for the research agenda

on the topic. For example, if measures of state capacity are equally valid and interchangeable, scholars can be ensured that selecting one measure instead of another is not likely to cause major consequences for their research. However, if there are large dissimilarities among measures, our alarm bells should start ringing. If every measure tells a different story, it becomes well-founded to question the validity of frequently used measures as quantifications of state capacity, and even more, the validity of extant findings on the topic. Hence, ultimately, this study provides critical guidance for future quantitative work on state capacity.

Last, I want to emphasize that the overall aim of this study is not to contribute to the conceptual literature on state capacity. Without downplaying the importance of the conceptual debate on the topic, 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”. Since our knowledge about state capacity is affected by how it is measured, measurement issues are of primary importance.

Data and Methods

Selecting Measures of State Capacity

A plethora of measures have been used to quantify state capacity in cross-national comparative literature. Since a comprehensive analysis of all these measures is impossible, I select some of the most established ones for further analysis according to four criteria. First, selected measures must have been frequently used to measure state capacity in recent (≥ 2010) political research by many different scholars. This first criterion makes the original intended purpose of the measures trivial for aim of my study. Second, I focus on subjective measures of state capacity. All the selected measures are at least partially based on perception-based data from expert surveys and/or assessments. Third, selected measures have been coded on a yearly basis over time and across most of the countries in the world. Fourth and last, selected measures are publicly available free of charge. Selected measures and their main characteristics are presented in Table 1.

The Quality of Government Institute of the University of Gothenburg publishes the well-known *Quality of Government Index* (Teorell et al. 2019). The index conceives state capacity as a tri-dimensional concept and is based on three separate sub-indicators: bureaucracy quality, corruption, and law and order. QOG is computed as the average of these three sub-indicators, which are all coded by PRS Group’s country experts. The index provides data for almost 150 countries in the world since 1984.

Hanson and Sigman’s (2013) *State Capacity Index* has gained widespread popularity among political researchers because it is based on strong theoretical arguments. HSI focuses on three dimensions of

state capacity: extractive, coercive, and administrative. In turn, these three dimensions are captured by 24 different sub-indicators and synthesised to a single index with latent variable analysis. HSI provides annual data for up to 163 countries in 50 years (1960-2009). The index can be retrieved freely from several replication datasets.

Government Effectiveness is one of the six 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" (Kaufman, Kraay, and Mastruzzi 2011: 4). WGI is a composite index based on multiple sub-indicators (48 in 2018), it covers virtually all countries in the world, and it is available biannually from 1996 to 2002 and annually from 2002 onwards.

The *State Fragility Index* is produced and published by the Center for Systemic Peace. The index captures state capacity in a broad sense and measures the "capacity to manage conflict, make and implement public policy, and deliver essential services" (Marshall and Elzinga-Marshall 2017: 51). SFI is based on 14 sub-indicators related to political, social, economic, and security aspects of state effectiveness and legitimacy. The index provides annual scores for all countries in the world with a population of at least 500,000 since 1995.

The *Failed States Index*, produced by the US-based NGO Fund for Peace, is conceived to provide an entry point "to understand more about a state's capacities and pressures" (Fund for Peace 2019: 33). FSI scores are based on expert coding, content analysis of articles and reports, and quantitative secondary data concerning 12 domains such as security, rule of law, and public services. More than 100 sub-indicators are synthesised to get the final 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.

TABLE 1. MAIN CHARACTERISTICS OF SELECTED MEASURES OF STATE CAPACITY

Measure	Producer	Years	Countries	Scale	Underlying variables	Type of data	Employed in (e.g.)
Quality of Government Index (QOG)	Quality of Government Institute	1984-2018	147	0 to 1	3	Subjective	Charron and Lapuente (2010, 2011); Knutsen (2013); Rothstein, Samanni, and Teorell (2012); Walther, Hellström, and Bergman (2019).
State Capacity Index (HSI)	Hanson and Sigman (2013)	1960-2009	163	Mean 0, standard deviation 1	24	Subjective and objective	Grassi and Memoli (2016); Van Ham and Seim (2018); Kim and Kroeger (2018); Bizzarro et al. (2018).
Government Effectiveness (WGI)	World Bank Institute	1996-2018	193	Mean 0, standard deviation 1	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 (high) to 25 (low)	14	Subjective and objective	Besley and Persson (2011); Cingolani, Thomsson, and De Crombrughe (2015); Hiilamo and Glantz (2015).
Failed States Index (FSI)	Fund for Peace	2005-2019	178	0 (high) to 120 (low)	100+	Subjective and objective	Møller and Skaaning (2011b); Lee and Zhang (2017); D'Arcy and Nistotskaya (2017).
Impartial Public Administration (VDEM)	Varieties of Democracy	1789-2019	179	Mean 0, standard deviation 1	1	Subjective	Gjerlow et al. (2018); Bizzarro et al. (2018); Grundholm and Thorsen (2019); Cornell, Knutsen, and Teorell (2020).
Corruption Perceptions Index (CPI)	Transparency International	1995-2019	180	0 to 10 until 2011; 0 to 100 since 2012	14	Subjective	Joshi, Hughes, and Sisk (2015); Cingolani, Thomsson, and De Crombrughe (2015); Lin (2015).

Number of countries refers to the latest year of data.

Transparency International's *Corruption Perceptions Index* aggregates existing measurements of corruption and closely related issues. Since “measures of corruption may provide another way of measuring state capacity” (Englehart 2009: 46), the index has been employed as a proxy of state capacity in many cross-national studies. CPI has been published annually since 1995, it is based on secondary data from several expert surveys, and the 2018 edition covers 180 countries in the world.

V-Dem Institute's *Rigorous and Impartial Public Administration* provides information about “the extent to which public administration is characterized by arbitrariness and biases” (Coppedge et al. 2019: 162). Even if the indicator cannot capture state capacity as a whole, it has been used in several recent studies as a proxy of state capacity because the functioning of a bureaucracy is arguably the most critical aspect of capable states (e.g., Charron, Dahlström, and Lapuente 2012; Knutsen 2013). VDEM is based on assessments by multiple country experts and provides annual data since 1789 for nearly all countries in the world.

Research Strategy

Now that we have selected some of the most relevant measures of state capacity we can proceed to their statistical analysis. Unless otherwise stated, FSI and SFI are reversed so that a higher score indicates higher state capacity. First, the main statistical properties of the measures are examined and compared. Then, correlations are used to analyse bivariate similarity and association among the measures. Correlation analysis is a conventional tool to assess the convergent validity of instruments measuring the same construct. All correlations are computed with both Pearson's and Spearman's methods, but only Pearson's correlation coefficients are reported because the results are not considerably affected by the selected method. With principal component analysis (PCA) I explore the dimensionality and multivariate association of the measures. The results of the PCA suggest that measures of state capacity are strongly related among each other and capture a one-dimensional concept of state capacity.

Next, measures of state capacity are examined against external predictors with regressions. The interchangeability of our measures of state capacity is assessed by replicating a selection of studies on the effect of democracy on state capacity. The aim of these regressions is to assess whether different measures of state capacity lead to similar empirical findings (i.e., are interchangeable) and to answer the question *does the choice of a measure affect the conclusions of a given study?* Furthermore, as a by-product, we are able to assess the external validity of the replicated studies. Despite strong associations I find that the choice of the measurement of state capacity matters substantially for the conclusions to be drawn. This means that the replicated studies have weak external validity and their findings cannot be generalized.

In the last part of the paper, to understand better the similarities and differences of the measures of state capacity, I focus on individual country ratings. First, country ratings are analysed bivariately. Then, by creating an indicator of multivariate country-specific rating discrepancy, I determine which countries have highly similar or dissimilar scores across all measures and shed light on the causes of rating discrepancy. Last, by shifting back the level of analysis from individual countries to global, I show that rating discrepancy is systematically related to the level of state capacity.

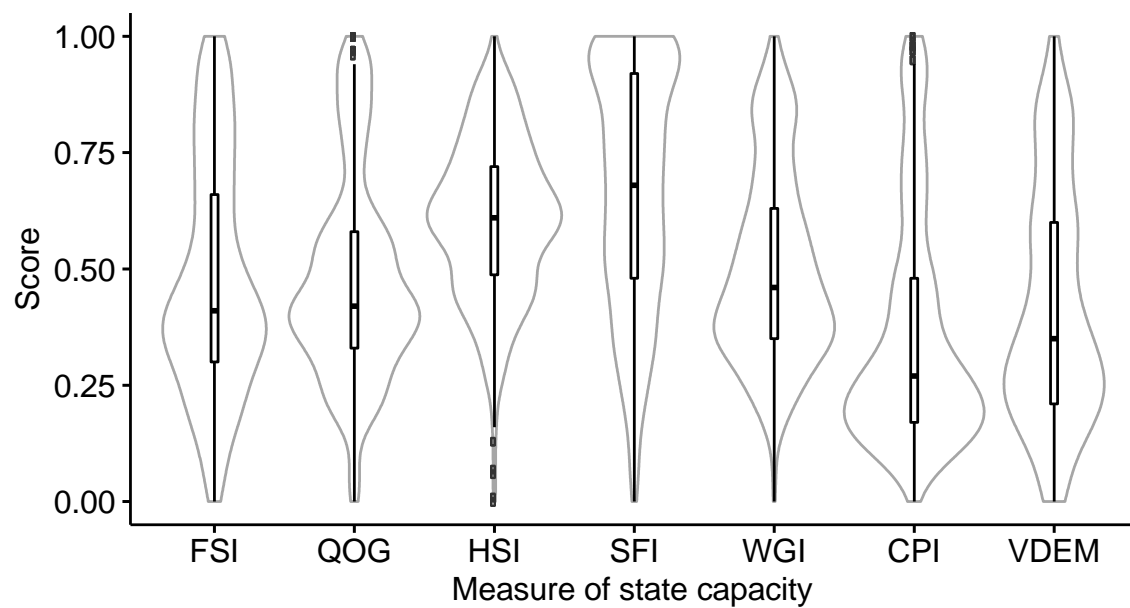
Results

Statistical Properties, Convergent Validity, and Dimensionality

Violin plots (Figure 1) reveal the main statistical features of the selected measures in years of common coverage. The outlines of the “violins” show the distributional characteristics 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 are represented by dots above or below the whiskers. The black dot inside the box represents the median value of each variable.

There are some interesting similarities and differences among our measures. Statistically, we would like to have more or less normally distributed variables, but the violins show that not all measures follow a normal distribution. CPI has particularly low modal and median values. This means that compared to the other measures of state capacity it compresses most observations are at the lower end of the scale. VDEM has a right-skewed distribution and relatively low modal and median values, as well. On the other side of the spectrum we have SFI, which has the highest modal and median values, and compresses most observations at the high capacity of the scale. FSI, QOG, and WGI have similar distributions with modal and median values slightly below the mid-point of the scale. QOG and CPI have some outliers at the high end of the scale, whereas HSI has outliers at the low end of the scale.

FIGURE 1. VIOLIN PLOTS OF MEASURES OF STATE CAPACITY (2005-2009).



Missing data handled with listwise deletion. Scores are normalized to range from 0 to 1.

A compression of observations at one of the two ends of the scale is likely to be problematic, because intervals and distances between observations become dependent on the level of state capacity. Out of 170 observations in CPI (2009), there are as many as 61 observations from 2 to 3 but only 44 observations from 5 to 10, causing unrealistic distances between observations. For example, according to the CPI scores (2009), the difference in state capacity between Liberia (3.0) and China (3.6) is smaller than the difference between Norway (8.6) and Denmark (9.3), and the difference between Austria (7.9) and the Netherlands (8.9) is twice the size of the difference between the Dem. Rep. of Congo (1.9) and Belarus (2.4).

SFI has similar limitations at the opposite end of the scale, since it rates only 8 countries from 20 to 25 (low capacity) but 61 countries from 0 to 5 (high capacity) in 2009. Additionally, according to SFI almost 20 countries have the maximum possible level of state capacity. This is extremely problematic for two reasons: First, if some of these countries improved their level of state capacity, SFI would not be able to capture the improvements. Second, since other measures of state capacity are able to distinguish between these countries almost without exceptions, we are induced to conclude that these countries do have some differences in state capacity, but SFI is not able to capture them.

Now we have some information about the main statistical properties of the measures, but we do not know if these measures are associated among each other. With bivariate correlations we can assess the strength of the relationship between two measures. Moreover, correlations against other measures of the same construct are conventionally used as a tool of measurement validation. Correlation coef-

ficients (Table 2) show that the measures are highly correlated among each other. The weakest correlations are between SFI and VDEM (0.70) and HSI and VDEM (0.72), while the strongest correlations are between CPI and WGI (0.94) and QOG and WGI (0.93). These findings indicate a high convergent validity of all the measures.

TABLE 2. PAIRWISE CORRELATION COEFFICIENTS OF MEASURES OF STATE CAPACITY (2005-2009).

	FSI	QOG	HSI	SFI	WGI	CPI	VDEM
FSI	1.00						
QOG	0.87 (671)	1.00					
HSI	0.85 (784)	0.86 (665)	1.00				
SFI	0.87 (796)	0.76 (675)	0.83 (809)	1.00			
WGI	0.90 (820)	0.93 (685)	0.90 (809)	0.81 (824)	1.00		
CPI	0.89 (794)	0.91 (674)	0.84 (782)	0.75 (795)	0.94 (824)	1.00	
VDEM	0.81 (820)	0.80 (685)	0.72 (809)	0.70 (824)	0.83 (864)	0.84 (824)	1.00

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

So far, we have examined measures of state capacity in years of common coverage. Despite some differences in main statistical properties, we have found that measures of state capacity are strongly related to each other from 2005 to 2009. Yet, the strong correlations hold also over a longer time period¹ (Tables A1-A9, Appendix A). Generally, correlation coefficients are high throughout the analysed period and it is astonishing how consistent the bivariate relationships are over time. Only in one case the strength of the correlation varies more than 0.1: the correlation between QOG and CPI ranges from 0.81 (1995) to 0.93 (multiple years). However, if we exclude 1995, the correlation between QOG and CPI is never lower than 0.90. Moreover, also the correlations between CPI and HSI, and CPI and VDEM take a pronounced leap from 1995 to 1996, suggesting that there could be

¹ 1995-2017; data before 1995 is not analysed because most of the measures do not cover earlier years.

something anomalous in the CPI scores of 1995. Certainly, the scarce amount of countries (39) rated by CPI in 1995 can affect its relationship with other measures.

Bivariate correlations provide information about the relationship between two given variables. However, we can analyse the relationship among our measures of state capacity with multivariate methods as well. PCA is often used as a variable-reduction technique but it can also help to understand better the common dimensionality and the multivariate association among multiple variables. The results of the PCA (Table 3) show that almost 87% of the common variance can be attributed to one single component. The second component explains only around 5% of the common variance. Since according to the Kaiser criterion components with eigenvalues under 1.0 should not be retained, the PCA indicates that the measures of state capacity are best represented by one single dimension and suggests that all the indicators measure the same concept. Only if we would have found the second component to explain a substantial amount of common variance, we could have questioned whether our measures capture the same concept at all. Robustness tests with extended year coverage do not change our conclusions and the bottom line remains the same: the instruments measure one and the same concept of state capacity.

TABLE 3. PRINCIPAL COMPONENT ANALYSIS OF MEASURES OF STATE CAPACITY (2005-2009).

Component	Eigenvalue	% of explained variance	Cumulative % of explained variance
1	6.084	86.91	86.91
2	0.332	4.74	91.66
3	0.267	3.82	95.47
4	0.113	1.61	97.09
5	0.095	1.35	98.44
6	0.066	0.95	99.39
7	0.043	0.61	100.00

Interchangeability and External Validity of Previous Studies

So far, we have found that our cross-national measurements of state capacity have high convergent validity, are strongly related to each other and measure the same concept. Nevertheless, high correlations do not always translate into high interchangeability, which can be assessed by analysing the measures against external predictors. To assess the empirical consequences of choosing one measure instead of another I replicate three regression models published in three studies on the effect of democracy on state capacity. The choice of replicating studies about this specific topic is not casual but determined by the fact that it constitutes one of the largest literatures where state capacity is

examined as an outcome. Besides providing information about the interchangeability of the measures, we are also able to assess the external validity of the replicated studies.

I have chosen to replicate three longitudinal models. Bäck and Hadenius' (2008) and Carbone and Memoli's (2015) studies are selected because of their influential contribution to the literature on the topic. Grassi and Memoli's (2016) study is selected because it is one of the most recent contributions on the topic and it covers a time span that is common almost to all our measurements of state capacity. Replication data is available only for the latter two studies but Bäck and Hadenius' (2008) study is replicated to the best of my ability by following scrupulously the procedure described by the authors. I want to stress that these replications are not intended to criticize any of the concerned studies. To ease the comparability of the estimations measures of state capacity are normalized to range from 0 to 1.

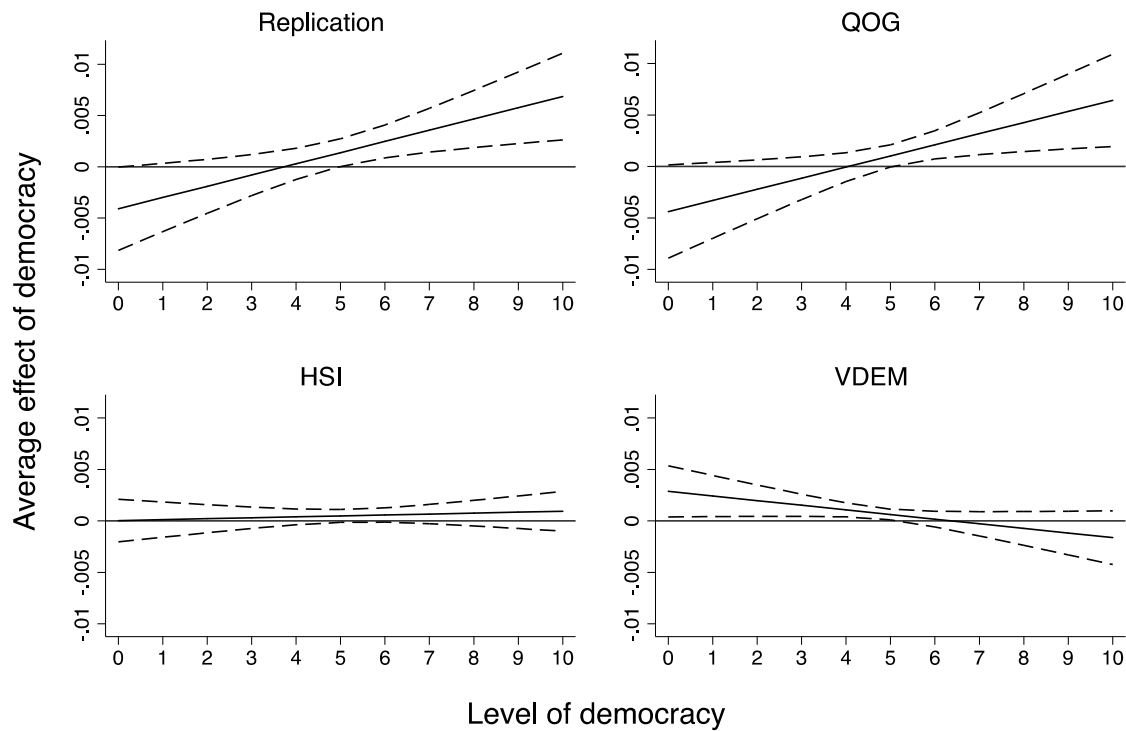
I start with Bäck and Hadenius' (2008) study, where the authors find evidence about a curvilinear effect of democracy on state capacity: at low levels of democracy the effect is negative, while at high levels of democracy the effect is positive. To operationalize state capacity the authors aggregate *Bureaucracy Efficiency* and *Corruption* from ICRG into an additive index that covers the period of time from 1984 to 2002. Only three of our seven measures cover the entire period of Bäck and Hadenius's study, and thus, the robustness of the original study is tested only with three "alternative" models. A summary of the regression results is presented in Table B1 (Appendix B).

The original model (1) confirms that democracy has a curvilinear effect on state capacity. As claimed by Bäck and Hadenius (2008), at low levels of democracy the effect is negative and at high levels of democracy it is positive. In Model 2 state capacity is measured with QOG. Now the predicted effect is similar but significant only at the 90% level. The strong equivalence between the two models is not surprising since the original measure of state capacity is based on almost the same sub-indicators than QOG. The curvilinear effect does not hold even closely in Model 3, in which state capacity is measured with HSI. Model 4, in which state capacity is measured with VDEM, provides some evidence about a curvilinear effect of democracy on state capacity (significant only at the 90% level) but in this case the curvilinearity is completely opposite than in the replicated model. While only at a lower level of significance, Model 4 suggests that the effect of democracy on state capacity is positive at low levels of democracy but negative at high levels of democracy.

Average marginal effect (AME) plots (Figure 2) show a more detailed picture of the consequences of choosing one measure over another. In the original model the effect of democracy is likely to be negative in countries with a complete absence of democracy. The effect of democracy is nonsignificant in countries with a low level of democracy but becomes significantly positive 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 and Nigeria the relationship between democracy and

state capacity is significantly positive. The results are similar when state capacity is measured with QOG. Anyhow, when state capacity is measured with VDEM the results are the opposite: from low to intermediate levels of democracy (<6) the relationship between democracy and state capacity is positive. 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 and Saudi Arabia and partially democratic countries like Russia and Nigeria. With VDEM the relationship becomes non-significant in more democratic countries. When state capacity is measured with HSI the results provide no evidence of a curvilinear association between democracy and state capacity.

FIGURE 2. AME OF DEMOCRACY ON STATE CAPACITY OF MODELS IN TABLE B1 (APPENDIX B).

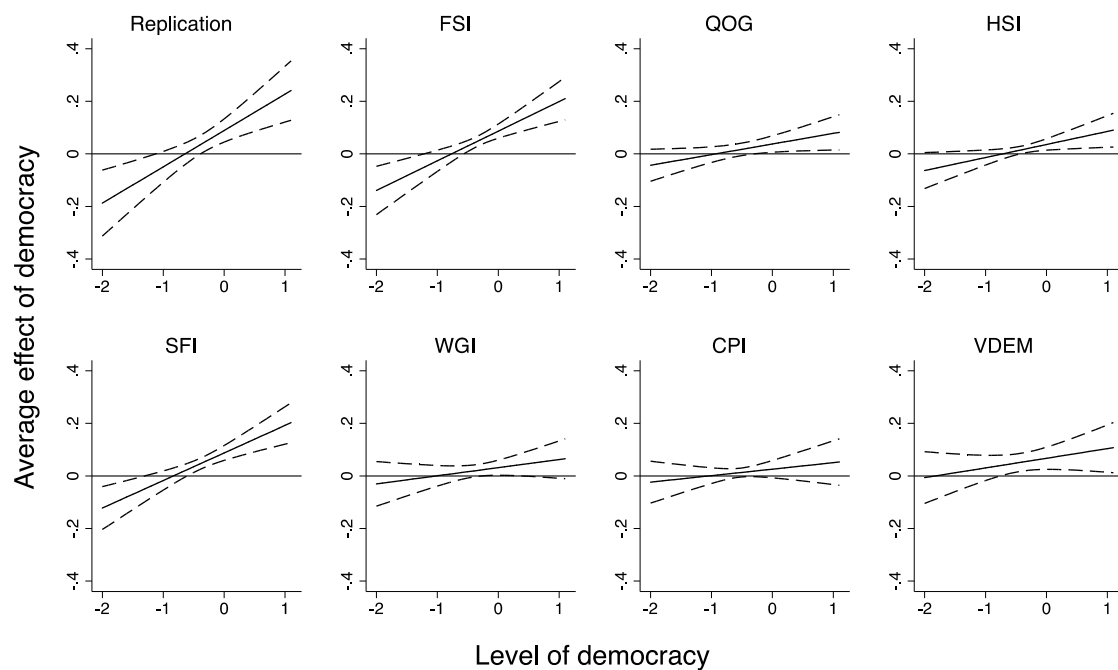


Second, I test whether Carbone and Memoli's (2015) findings are sensitive to the choice of the measure of state capacity (Table B2, Appendix B). Model 1 is replicated with the original measurement used in Carbone and Memoli's research, where *Monopoly on the Use of Force* and *Basic Administration* from Bertelsmann Stiftung are multiplicatively aggregated. The original model finds strong evidence about a curvilinear effect of democracy on state capacity. At extremely low levels the effect is negative, but the effect turns positive after a certain level of democracy has been reached. In models 2-8 the original measure is replaced with our alternative measures of state capacity. Surprisingly, as before, choosing one measure over another can lead to completely different results and interpretations. The strong curvilinear association between democracy and state capacity holds only with FSI or SFI.

When state capacity is measured with WGI, CPI, or VDEM there is no evidence of such a curvilinear association. With QOG or HSI the curvilinear relationship is substantially weaker compared to the replicated model and holds only at a lower level of statistical significance.

A more exhaustive analysis of the results reveals even further discrepancies among the models. AME plots (Figure 3) show that the main finding of the original model is hold only in two of the alternative models.

FIGURE 3. AME OF DEMOCRACY ON STATE CAPACITY OF MODELS IN TABLE B2.



Using FSI or SFI leads to similar findings compared to the original model, albeit with different magnitudes. Models with QOG and HSI suggest that the positive effect of democracy on state capacity begins only after a country has reached an intermediate level of democracy. In contrast with the original model neither of these two models find that democracy has a negative effect in extremely undemocratic countries. Models with WGI, CPI, and VDEM do not support any of these findings and according to these three models the effect of democracy on state capacity is not dependent on the level of democracy at all. Model 1 confirms that “democratic duration becomes a crucial factor when combined with the degree of democracy” (Carbone and Memoli 2015: 18), but this finding is not confirmed by any of the alternative models.

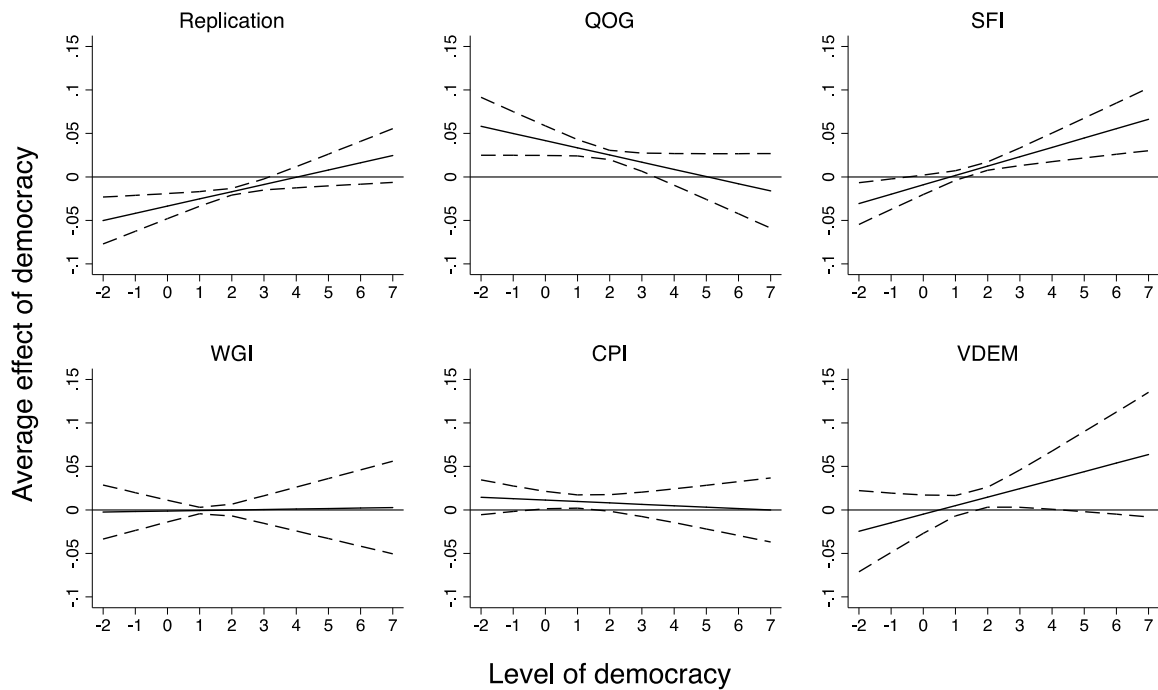
Third, I replicate Grassi and Memoli’s (2016) study and assess the external validity of its findings (Table B3, Appendix B). The discrepancies between the original model (with HSI) and the alternative models are even more pronounced than in the other two sets of longitudinal regressions. The original

model finds a significant non-linear effect of democracy on state capacity: this effect is negative in autocratic countries but fades out once a country reaches a certain level of democratization. Moreover, the original model finds that left-wing executives have fostered 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 significant at conventional levels. With QOG the main term is significant at the 99.9% level and the quadratic term is very close to conventional significance levels (i.e., significant at the 90% level), but the point estimates suggest a completely opposite story compared to the original model. With QOG it seems that democracy has a positive effect on state capacity in autocracies, but this effect gradually disappears once a certain level of democratization has been reached. With WGI or VDEM neither of the terms are significant. With SFI only the squared term is significant whereas with CPI only the main term is significant.

AME plots (Figure 4) show more in detail how the predicted impact of democracy is sensitive to the chosen measurement. In the original model the initially negative marginal effect of democracy disappears when the level of democracy increases. On the contrary with QOG the initially positive marginal effect of democracy disappears when a country becomes fully democratic. With WGI, CPI, and VDEM the average effect of democracy on state capacity does not depend on the level of democracy. In the model with SFI democracy increases state capacity only once a certain level of democracy has been reached.

FIGURE 4. AME OF DEMOCRACY ON STATE CAPACITY OF MODELS IN TABLE B3.



As to the partisan balance of the executive, the original model and the model with SFI find a positive impact of left-wing executives on state capacity, but instead with CPI it turns out that right-wing executives have a significantly positive impact on state capacity. The model with QOG supports the latter finding, although only weakly (at the 90% level). With WGI or VDEM, state capacity is not significantly affected by the partisan balance of the executive.

The replication of three studies with up to eight different measures of state capacity has shown that the choice of the measure plays a key role in the conclusions drawn from the replicated studies, undermining both the interchangeability of the measures and the external validity of these studies. Since measurements do not always cover the same sample of countries, my findings could be driven by different samples rather than different measures. To rule out selection bias I run all the previous sets of models with the same sample of observations. The results are not substantially affected by restricting the models within each set of replications to the same sample and the conclusions are not altered by using a set of common observations. Selection bias does not affect the interpretation of models in any of our sets of replications.

We have found strong evidence that our seven measures of state capacity are highly correlated among each other and represent the same one-dimensional concept. It is commonly thought that highly correlated variables are nearly equivalent to each other. Anyhow, our findings have shown 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 indicate that the interchangeability of measures of state capacity is low and the external validity of the replicated studies is weak. Not even one single pair of measures produces consistently similar results, but WGI and CPI seem to be the most interchangeable pair of measures. Overall, it is worrisome that previous findings on the nexus between democracy and state capacity are so sensitive to the chosen measure.

Country-Specific Rating Discrepancy

So far, we have mainly found that the interchangeability of measures of state capacity is low even if the measures are similar and strongly associated to each other. These contradictory findings require further investigation, and it is likely that we will better understand what causes our contradictory findings if we turn our attention to the country-level.

With bivariate scatter plots of state capacity measures (Figures C1-C21, Appendix C) we can grasp how similarly individual countries are rated in the most recent year of common observations. Overall, many countries are rated with high consistency by each pair of measures, as suggested previously by the correlation analysis. Somalia has an extremely low score in all measures, whereas the Nordic Countries, Switzerland, and New Zealand have an extremely high score in all measures. Yet, it becomes evident that there are also countries that are rated in a substantially different way by our measurements. Keeping in mind that the measures are normalized to range from 0 to 1, some of the rating divergencies are astonishing (Tables D1-D21, Appendix D).

As we already know, SFI tends to give countries higher and CPI lower scores than the other measures. Thus, we suspect to find large country-level discrepancies between SFI and CPI. There are as many as 45 countries that SFI rates more than 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.

SFI rates countries with considerable divergencies compared to most other measures as well. It rates Belarus 0.60 units higher than WGI and there are five other countries that are rated more than 0.40 units higher by SFI than by WGI. It rates Albania 0.55 units higher than QOG and there are 16 countries that are rated more than 0.40 units higher by SFI than by QOG. It 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. Instead, country-specific differences between SFI and HSI are relatively small. The most differently rated country is Argentina, which is rated 0.36 units higher by SFI.

Differences in country ratings between HSI and the other measures are substantial as well. Compared to CPI, HSI rates 16 countries at least 0.40 units higher, and six of these 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). Compared to VDEM, HSI rates seven countries at least 0.50 units higher: Egypt (0.66), Belarus (0.55), Kuwait (0.55), Malaysia (0.53), Tunisia (0.50), Azerbaijan (0.50), and Kazakhstan (0.50). As to HSI and FSI, Iran has the highest rating discrepancy. HSI rates Iran 0.44 units higher than FSI. As to HSI and QOG, Venezuela has the highest discrepancy: HSI rates it 0.50 units higher than QOG. As to HSI and WGI, Belarus is rated 0.44 units higher by HSI and it is the only country rated with a discrepancy larger than 0.40 between the two indices.

Due to CPI's comparatively low scores, it is not surprising to find that there are seven countries rated at least 0.30 units higher by FSI than CPI, but no countries rated at least 0.30 units higher by CPI than FSI. The country with the highest difference between the two measures is Argentina, rated 0.50 units higher by FSI. A similar pattern can be found when comparing the country ratings of CPI and VDEM. Five 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. Philippines is the country with the most diverging rating (0.33 units higher with WGI).

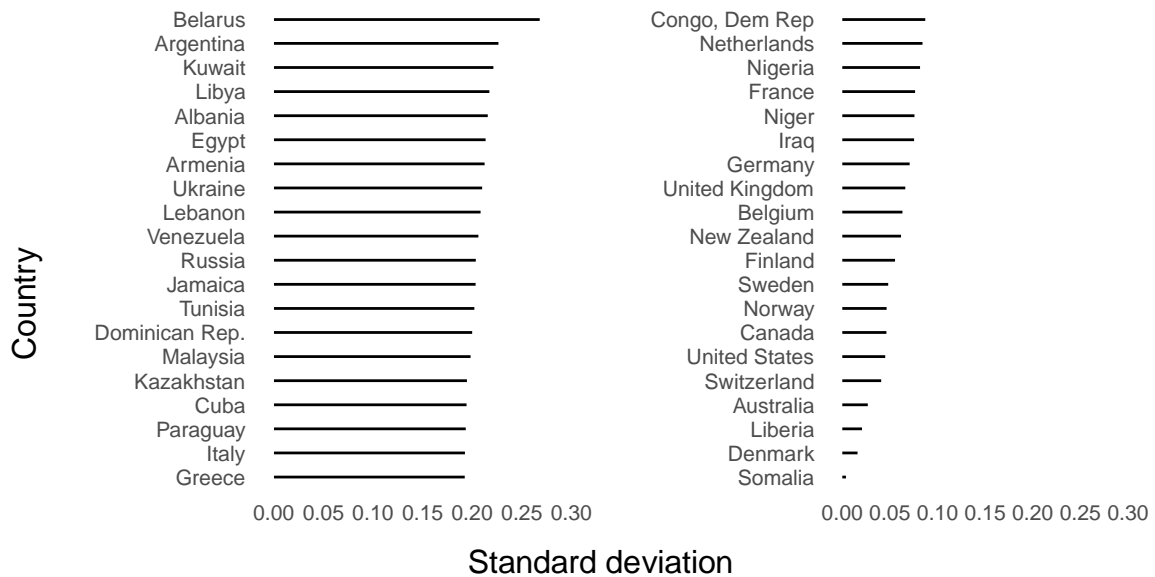
The country with the largest discrepancy between WGI and FSI is Cyprus, which is rated 0.81 by WGI and 0.48 by FSI. WGI and QOG tend to rate countries relatively similarly: there are no country scores with a discrepancy of more than 0.30. Differences in country scores between WGI and VDEM are slightly 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). 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 even less marked and 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.

These results have shown that measures of state capacity do not rate countries similarly. When differences in country scores between measures are so high, it is understandable that the interchangeability of measures is low. Overall, single observations with the largest discrepancies have relatively high scores with SFI or HSI and relatively low scores with CPI and VDEM. We can suspect that countries that are repeatedly among the most divergently rated ones bivariately, such as Belarus and Kuwait, stand out also in multivariate discrepancy. To determine multivariate rating discrepancy, I

compute the country-specific standard deviations of all country scores. 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.

As suspected, Belarus and Kuwait are among the countries with the largest multivariate rating discrepancy (Figure 5). This group of countries seems to have fairly heterogeneous characteristics. There are both developed and developing countries, and both democratic and authoritarian countries, but there are no Western liberal democracies besides Italy and Greece. Interestingly, not even one of the countries in the chart has full civil liberties according to Freedom House’s ratings of the same year. Politico-geographically, most of these countries are in Eastern Europe, the Middle East/North Africa, or Latin America/the Caribbean, whereas Sub-Saharan African countries are completely absent from the chart. Nearly half of the 20 most discrepantly rated countries have a Muslim-majority population. Countries with small rating discrepancy can be more straightforwardly categorized into two distinct groups: highly dysfunctional states (e.g., Somalia, Iraq, Liberia) and Western liberal democracies. These countries have either very low or very high capacity, and their scores are more or less equivalent across measures.

FIGURE 5. COUNTRIES WITH LARGEST/SMALLEST MULTIVARIATE RATING DISCREPANCY (2009).

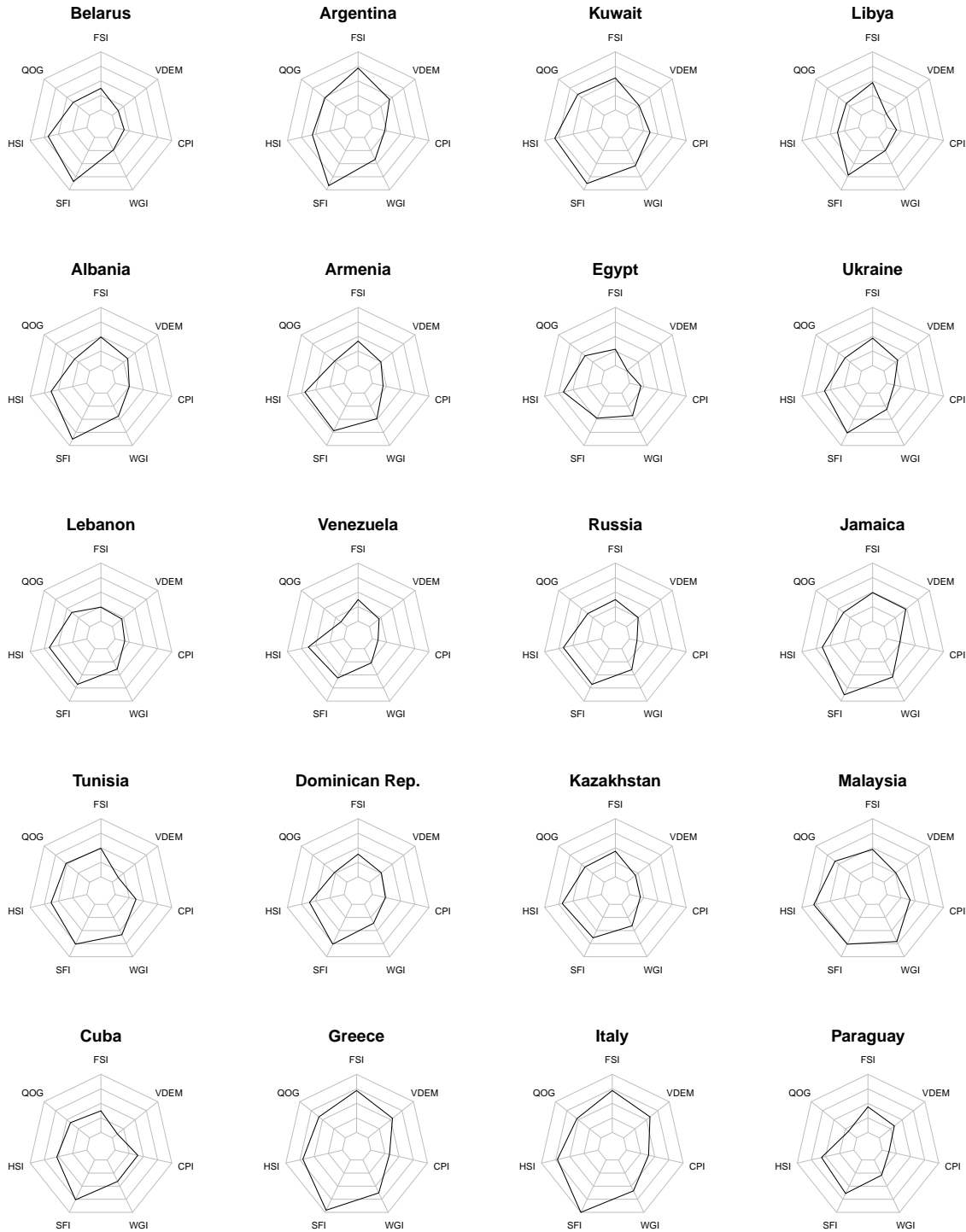


Countries with largest discrepancy on the left. Countries with smallest discrepancy on the right.

Figure 6 provides illustrative multivariate information about country scores in the most discrepantly rated countries and confirm a pattern that was previously suggested by bivariate comparisons: most of these countries have relatively high scores with SFI and HSI, but relatively low scores with CPI and VDEM. If the shapes of the “nets” are corresponding, different countries have multivariately equivalent scores. For instance, Italy, Greece, Albania, and Ukraine seem to be relatively similar: higher ratings with SFI, HSI, and FSI, but lower ratings with the other four measurements. Russia, Belarus, and Kazakhstan have some 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 rated are particularly low in VDEM. Paraguay and Venezuela are rated comparatively low in QOG and CPI.

Some of these discrepancies are likely to be determined by slight differences in the defining attributes of the measures. VDEM and CPI focus on corruption and related issues. SFI and HSI capture a broader set of dimensions, but in both the coercive dimension of the state plays a more important role than in the other measures, and both are based on several sub-indicators related to political institutionalization and security. WGI and QOG focus mainly on the quality of the bureaucracy, although the former emphasizes also the quality of public services, whereas the latter gives importance as well 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. Thus, with FSI state capacity is understood more broadly than with the other measures.

FIGURE 6. SPIDER CHARTS OF COUNTRIES WITH LARGEST MULTIVARIATE RATING DISCREPANCY.

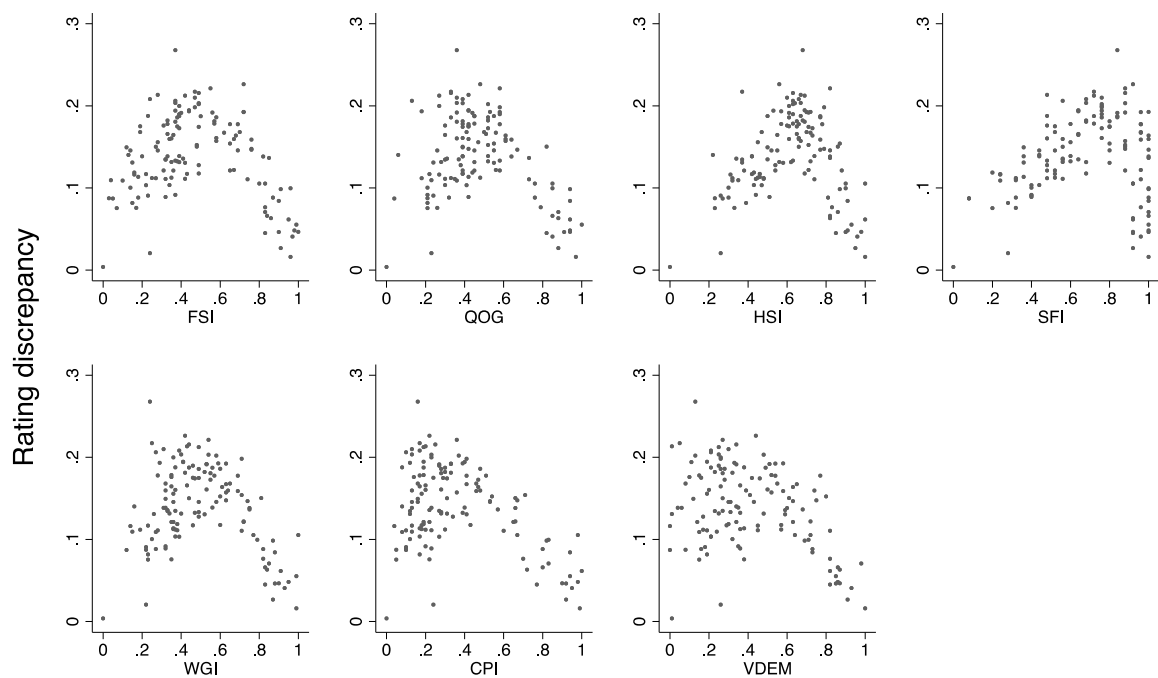


Scores are normalized to range from 0 to 1.

If we examine the ratings in relation to the aspects covered by each measure in individual countries, we can understand better some of the causes of the rating inconsistencies. For instance, it is not 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. It seems that many of the countries with high rating divergency are corrupted 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.

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 attributed to the different areas of state capacity quantified by each instrument, which is positive news. By rigorously matching a chosen definition of state capacity with a chosen measure, and by making these choices clear to the reader, scholars can push forward research on state capacity. Anyhow, it is less promising to find that rating discrepancy depends systematically on the level state capacity (Figure 7).

FIGURE 7. RATING DISCREPANCY AND LEVEL OF STATE CAPACITY (2009).



Regardless of the measure there 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. This is understandable, because survey experts and coders are more likely to agree about clear-cut cases on the extreme ends of the spectrum. Less clear cases are simply harder to code, and experts can be expected to have diverging perceptions about state capacity in these countries. Thus, systematic discrepancy can be attributed to the subjective nature of our measures, but it affects our knowledge on state capacity even when a given working definition matches perfectly with the selected measure.

Conclusions

This study has analysed and compared comprehensively seven of the most frequently used measures of state capacity and evaluated the validity, interchangeability, and rating discrepancy among the measures. The analysis has been predominantly statistical, but the possible causes of rating discrepancy have been also assessed in relation to the qualitative differences in the measured construct. The main findings of this paper are manifold. First and foremost, the study at hand provides one of the first systematic statistical comparisons of measures that have been frequently used to quantify state capacity in political research.

We have found that the convergent validity 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 unidimensionality of the measures is confirmed by a PCA. Qualitatively each measure captures slightly different aspects of state capacity, but the statistical analysis has shown that quantitatively they measure the same.

Despite a strong association between measures of state capacity, the set of replicated regression models has revealed that the interchangeability among these measures is low and the chosen measure influences the conclusions. In the most worrisome cases, we have found that two measures can lead to completely opposing interpretations. Scholars working on state capacity need to be aware that their research is not likely to be generalizable and should make clear that the external validity of their research is likely to be weak. Furthermore, the results of the replications cast doubt on the extant knowledge about the relationship between democracy and state capacity. How solid is our knowledge on the topic, if all replicated studies are so sensitive to the chosen measure?

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 differences in individual country scores among measures. By creating an indicator of 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. High rating discrepancy can generally be attributed at least to two factors: the different aspects

of state capacity that each measure captures and the systematic disagreement at intermediate levels of state capacity.

Despite high convergent validity, our findings have shown that the measures are not equivalent. For instance, SFI is not able to capture possible improvements in many high capacity countries. FSI covers such a broad understanding of state capacity that it undermines its analytical utility in causal research. SFI and HSI rate countries comparatively high, CPI and VDEM rate countries comparatively low, and the differences can be overwhelming. For instance, if we measure state capacity with HSI, Egypt ranks 57th in the world in 2009 (more or less like China and Russia), but if we measure state capacity with VDEM, Egypt ranks 169th in the world in 2009 and performs worse than Somalia and Madagascar. Scholars must be aware about these divergencies and the consequences of choosing one measure instead of another. The selected instrument must match the working definition of state capacity and make clear to the reader what the selected instrument is actually measuring.

Last, the findings of this study provide two methods-related implications. First, strong correlations should not be taken as a 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 be substantially different. Highly correlated variables do not necessarily portray the same picture. Second, the findings remind the importance of replication studies in our field. Replications are fundamental to evaluate the robustness of previous findings and foster our understanding on any given topic.

REFERENCES

- Adcock, Robert and David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95(3):529–46.
- Bäck, Hanna and Axel Hadenius. 2008. "Democracy and State Capacity: Exploring a J-Shaped Relationship." *Governance* 21(1):1–24.
- Besley, Timothy and Torsten Persson. 2008. "Wars and State Capacity." *Journal of the European Economic Association* 6(2–3):522–30.
- Besley, Timothy and Torsten Persson. 2011. *Pillars of Prosperity: The Political Economics of Development Clusters*. Princeton: Princeton University Press.
- 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.
- 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.
- Carbone, Giovanni and Vincenzo Memoli. 2015. "Does Democratization Foster State Consolidation? Democratic Rule, Political Order, and Administrative Capacity." *Governance* 28(1):5–24.
- Charron, Nicholas and Victor Lapuente. 2010. "Does Democracy Produce Quality of Government?" *European Journal of Political Research* 49(4):443–70.
- Charron, Nicholas and Victor Lapuente. 2011. "Which Dictators Produce Quality of Government?" *Studies in Comparative International Development* 46(4):397–423.
- Charron, Nicholas, Carl Dahlström, and Victor Lapuente. 2012. "No Law Without a State." *Journal of Comparative Economics* 40(2):176–193.
- 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 Crombrughe. 2015. "Minding Weber More than Ever? The Impacts of State Capacity and Bureaucratic Autonomy on Development Goals." *World Development* 72:191–207.

- 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.
- Cornell, Agnes, Carl H. Knutsen, and Jan Teorell. 2020. "Bureaucracy and Growth". *Comparative Political Studies* 1–37. doi:10.1177/0010414020912262.
- 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.
- Dincecco, Mark. 2015. "The Rise of Effective States in Europe." *Journal of Economic History* 75(3):901–18.
- Dinecco, Mark. 2017. *State Capacity and Economic Development. Present and Past*. Cambridge: Cambridge University Press.
- Englehart, Neil A. 2009. "State Capacity, State Failure, and Human Rights." *Journal of Peace Research* 46(2):163–80.
- Evans, Peter B. 1985. "Transnational Linkages and the Role of the State." In *Bringing the State Back In*, ed. P. Evans, D. Rueschemeyer, and T. Skocpol. Cambridge: Cambridge 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):748–65.
- Fearon, James D. and David D. Laitin. 2003. "Ethnicity and Civil War." *American Political Science Review* 97(1):75–90.
- Fortin, Jessica. 2012. "Is There a Necessary Condition for Democracy? The Role of State Capacity in Postcommunist Countries." *Comparative Political Studies* 45(7):903–30.
- 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.
- 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.
- Grassi, Davide and Vincenzo Memoli. 2016. "Democracy, Political Partisanship, and State Capacity in Latin America." *Rivista Italiana Di Scienza Politica* 46(1):47–69.

- 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.
- 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. 2015. "Democracy and State Capacity: Complements or Substitutes?" *Studies in Comparative International Development* 50(3):304–30.
- Hendrix, Cullen S. 2010. "Measuring State Capacity: Theoretical and Empirical Implications for the Study of Civil Conflict." *Journal of Peace Research* 47(3):273–85.
- Hilamo, 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* 124:241–45.
- Högström, John. 2014. "Does the Choice of Democracy Measure Matter? Comparisons between the Two Leading Democracy Indices, Freedom House and Polity IV." *Government and Opposition* 48(2):201–21.
- Joshi, Devin. 2011. "Good Governance, State Capacity, and the Millennium Development Goals." *Perspectives on Global Development and Technology* 10(2): 339–60.
- 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.
- 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):220–46.
- Kim, Nam K. and Alex M. Kroeger. 2018. "Do Multiparty Elections Improve Human Development in Autocracies?" *Democratization* 25(2):251–72.
- Knutsen, Carl H. 2010. "Measuring Effective Democracy." *International Political Science Review* 31(2):109–28.
- Knutsen, Carl H. 2013. "Democracy, State Capacity, and Economic Growth." *World Development* 43:1–18.

- Lee, Melissa M. and Nan Zhang. 2017. "Legibility and the Informational Foundations of State Capacity." *The Journal of Politics* 79(1):118–32.
- Levi, Margaret. 1988. *Of Rule and Revenue*. Berkeley: University of California Press.
- Lin, Thung Hong. 2015. "Governing Natural Disasters: State Capacity, Democracy, and Human Vulnerability." *Social Forces* 93(3):1267–1300.
- Mann, Michael. 1984. "The Autonomous Power of the State: Its Origins, Mechanisms and Results." *European Journal of Sociology* 25(2):185–213.
- Marshall, Monty G. and Gabrielle Elzinga-Marshall. 2017. *Global Report 2017: Conflict, Governance, and State Fragility*. Vienna, VA: Center for Systemic Peace.
- Migdal, Joel S. 1988. *Strong Societies and Weak States. State-Society Relations and State Capabilities in the Third World*. Princeton: Princeton University Press.
- Møller, Jørgen and Svend-Erik Skaaning. 2011a. "On the Limited Interchangeability of Rule of Law Measures." *European Political Science Review* 3(3):371–94.
- 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: Palgrave Macmillan.
- Norris, Pippa. 2012. *Making Democratic Governance Work: How Regimes Shape Prosperity, Welfare, and Peace*. Cambridge: Cambridge University Press.
- 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):351–80.
- Rothstein, Bo and Jan Teorell. 2008. "What Is Quality of Government? A Theory of Impartial Government Institutions." *Governance* 21(2):165–90.
- 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.
- 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):441–58.
- Seeberg, Merete Bech. 2018. "Electoral Authoritarianism and Economic Control." *International Political Science Review* 39(1):33–48.

- Skaaning, Svend-Erik. 2010. "Measuring the Rule of Law." *Political Research Quarterly* 63(2):449–60.
- Skocpol, Theda. 1985. 'Bringing the State Back In: Strategies of Analysis in Current Research.' In *Bringing the State Back In*, ed. P. Evans, D. Rueschemeyer, and T. Skocpol. Cambridge: Cambridge University Press.
- Soifer, Hillel D. 2013. "State Power and the Economic Origins of Democracy." *Studies in Comparative International Development* 48(1):1–22.
- Teorell, Jan and Catharina Lindstedt. 2010. "Measuring Electoral Systems Few Political Scientists." *Political Research Quarterly* 63(2):434–48.
- Teorell, Jan, Stefan Dahlberg, Sören Holmberg, Bo Rothstein, Natalia Alvarado Pachon, and Richard Svensson. 2019. "The QoG Standard Dataset 2019."
- 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.
- Walther, Daniel, Johan Hellström, and Torbjörn Bergman. 2019. "Government Instability and the State." *Political Science Research and Methods* 7(3):579–94.

APPENDIX

Appendix A: Year-by-year correlations of measures of state capacity.

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

Year	QOG	HSI	SFI	WGI	CPI	VDEM
2005	0.88 (123)	0.84 (144)	0.87 (144)	0.92 (144)	0.88 (137)	0.82 (144)
2006	0.87 (137)	0.85 (160)	0.85 (163)	0.90 (169)	0.89 (156)	0.81 (169)
2007	0.86 (137)	0.85 (160)	0.87 (163)	0.89 (169)	0.90 (167)	0.81 (169)
2008	0.86 (137)	0.84 (160)	0.87 (163)	0.89 (169)	0.90 (167)	0.81 (169)
2009	0.86 (137)	0.85 (160)	0.88 (163)	0.90 (169)	0.89 (167)	0.81 (169)
2010	0.87 (137)		0.89 (163)	0.90 (169)	0.88 (166)	0.81 (169)
2011	0.87 (137)		0.89 (163)	0.90 (170)	0.87 (168)	0.81 (170)
2012	0.88 (137)		0.89 (164)	0.91 (170)	0.87 (166)	0.81 (170)
2013	0.87 (137)		0.90 (164)	0.91 (170)	0.88 (167)	0.79 (170)
2014	0.87 (137)		0.90 (164)	0.92 (170)	0.88 (166)	0.80 (170)
2015	0.88 (137)		0.90 (164)	0.92 (170)	0.89 (164)	0.80 (170)
2016	0.89 (137)		0.89 (164)	0.92 (170)	0.90 (166)	0.79 (170)
2017	0.89 (137)		0.89 (164)	0.92 (170)	0.89 (169)	0.78 (170)

Pearson's correlation coefficients; n 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)		0.81 (39)	0.75 (126)
1996		0.85 (122)	0.82 (124)	0.88 (126)	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 (128)	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 (137)	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 (137)	0.92 (98)	0.79 (137)

2003		0.84 (133)	0.77 (135)	0.92 (137)	0.91 (123)	0.80 (137)
2004		0.85 (133)	0.79 (135)	0.93 (137)	0.91 (129)	0.80 (137)
2005	0.88 (123)	0.87 (133)	0.79 (135)	0.93 (137)	0.91 (133)	0.81 (137)
2006	0.87 (137)	0.88 (133)	0.77 (135)	0.93 (137)	0.91 (133)	0.80 (137)
2007	0.86 (137)	0.88 (133)	0.76 (135)	0.93 (137)	0.91 (136)	0.79 (137)
2008	0.86 (137)	0.85 (133)	0.74 (135)	0.93 (137)	0.91 (136)	0.79 (137)
2009	0.86 (137)	0.84 (133)	0.75 (135)	0.93 (137)	0.91 (136)	0.79 (137)
2010	0.87 (137)		0.75 (135)	0.94 (137)	0.91 (135)	0.80 (137)
2011	0.87 (137)		0.75 (135)	0.94 (137)	0.92 (137)	0.79 (137)
2012	0.88 (137)		0.75 (135)	0.94 (137)	0.92 (137)	0.80 (137)
2013	0.87 (137)		0.74 (135)	0.93 (137)	0.92 (137)	0.78 (137)
2014	0.87 (137)		0.73 (135)	0.93 (137)	0.92 (137)	0.78 (137)
2015	0.88 (137)		0.75 (135)	0.93 (137)	0.93 (137)	0.78 (137)
2016	0.89 (137)		0.75 (135)	0.93 (137)	0.93 (137)	0.78 (137)
2017	0.89 (137)		0.75 (135)	0.93 (137)	0.93 (137)	0.78 (137)

Pearson's correlation coefficients; N 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)		0.80 (39)	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 (144)	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)	0.84 (152)	0.72 (162)
2007	0.85 (160)	0.88 (133)	0.83 (162)	0.90 (162)	0.84 (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)	0.82 (162)	0.90 (162)	0.83 (160)	0.71 (162)

Pearson's correlation coefficients; n 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 (39)	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 (144)	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 (165)	0.76 (163)	0.69 (165)
2009	0.88 (163)	0.75 (135)	0.82 (162)	0.81 (165)	0.76 (163)	0.70 (165)
2010	0.89 (163)	0.75 (135)		0.81 (165)	0.73 (162)	0.68 (165)
2011	0.89 (163)	0.75 (135)		0.81 (165)	0.72 (164)	0.69 (165)
2012	0.89 (164)	0.75 (135)		0.81 (166)	0.74 (163)	0.68 (166)
2013	0.90 (164)	0.74 (135)		0.81 (166)	0.74 (164)	0.69 (166)
2014	0.90 (164)	0.73 (135)		0.83 (166)	0.74 (163)	0.68 (166)
2015	0.90 (164)	0.75 (135)		0.83 (166)	0.75 (162)	0.69 (166)
2016	0.89 (164)	0.75 (135)		0.83 (166)	0.75 (163)	0.70 (166)
2017	0.89 (164)	0.75 (135)		0.81 (166)	0.75 (165)	0.69 (166)

Pearson's correlation coefficients; n 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 (126)	0.84 (159)	0.82(161)	0.94 (53)	0.83 (167)
1998		0.92 (128)	0.87 (160)	0.82 (163)	0.93 (84)	0.83 (170)
2000		0.92 (137)	0.85 (160)	0.82 (163)	0.94 (89)	0.84 (170)
2002		0.92 (137)	0.87 (161)	0.82 (164)	0.92 (101)	0.83 (171)
2003		0.92 (137)	0.88 (161)	0.82 (164)	0.93 (130)	0.83 (171)
2004		0.93 (137)	0.89 (161)	0.83 (164)	0.94 (143)	0.84 (171)
2005	0.92 (144)	0.93 (137)	0.89 (161)	0.83 (164)	0.93 (156)	0.84 (172)
2006	0.90 (169)	0.93 (137)	0.90 (162)	0.82 (165)	0.93 (158)	0.83 (173)
2007	0.89 (169)	0.93 (137)	0.90 (162)	0.80 (165)	0.94 (170)	0.83 (173)
2008	0.89 (169)	0.93 (137)	0.90 (162)	0.80 (165)	0.94 (170)	0.84 (173)
2009	0.90 (169)	0.93 (137)	0.90 (162)	0.81 (165)	0.94 (170)	0.83 (173)
2010	0.90 (169)	0.94 (137)		0.81 (165)	0.93 (170)	0.84 (173)
2011	0.90 (170)	0.94 (137)		0.81 (165)	0.93 (172)	0.83 (174)
2012	0.91 (170)	0.94 (137)		0.81 (166)	0.94 (169)	0.83 (174)
2013	0.91 (170)	0.93 (137)		0.81 (166)	0.94 (170)	0.82 (174)
2014	0.92 (170)	0.93 (137)		0.83 (166)	0.93 (169)	0.80 (174)
2015	0.92 (170)	0.93 (137)		0.83 (166)	0.93 (167)	0.80 (174)
2016	0.92 (170)	0.93 (137)		0.83 (166)	0.93 (169)	0.81 (174)
2017	0.92 (170)	0.93 (137)		0.81 (166)	0.93 (173)	0.79 (174)

Pearson's correlation coefficients; n 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		0.81 (39)	0.80 (39)	0.73 (39)		0.81 (39)
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 (130)	0.81 (130)
2004		0.91 (129)	0.85 (137)	0.75 (139)	0.94 (143)	0.81 (143)
2005	0.88 (137)	0.91 (133)	0.83 (150)	0.73 (152)	0.93 (156)	0.82 (156)
2006	0.89 (156)	0.91 (133)	0.84 (152)	0.73 (154)	0.93 (158)	0.82 (158)
2007	0.90 (167)	0.91 (136)	0.84 (160)	0.76 (163)	0.94 (170)	0.84 (170)
2008	0.90 (167)	0.91 (136)	0.84 (160)	0.76 (163)	0.94 (170)	0.84 (170)
2009	0.89 (167)	0.91 (136)	0.83 (160)	0.76 (163)	0.94 (170)	0.84 (170)
2010	0.88 (166)	0.91 (135)		0.73 (162)	0.93 (170)	0.84 (170)
2011	0.87 (168)	0.92 (137)		0.72 (164)	0.93 (172)	0.83 (172)
2012	0.87 (166)	0.92 (137)		0.74 (163)	0.94 (169)	0.86 (169)
2013	0.88 (167)	0.92 (137)		0.74 (164)	0.94 (170)	0.84 (170)
2014	0.88 (166)	0.92 (137)		0.74 (163)	0.93 (169)	0.85 (169)
2015	0.89 (164)	0.93 (137)		0.75 (162)	0.93 (167)	0.85 (167)
2016	0.90 (166)	0.93 (137)		0.75 (163)	0.93 (169)	0.86 (169)
2017	0.89 (169)	0.93 (137)		0.75 (165)	0.93 (173)	0.86 (173)

Pearson's correlation coefficients; n 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 (39)
1996		0.77 (126)	0.71 (160)	0.76 (163)	0.83 (167)	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 (170)	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)	0.84 (170)	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 (171)	0.82 (101)
2003		0.80 (137)	0.67 (161)	0.71 (164)	0.83 (171)	0.81 (130)
2004		0.80 (137)	0.69 (161)	0.71 (164)	0.84 (171)	0.81 (143)
2005	0.82 (144)	0.81 (137)	0.73 (161)	0.71 (164)	0.84 (172)	0.82 (156)
2006	0.81 (169)	0.80 (137)	0.72 (162)	0.70 (165)	0.83 (173)	0.82 (158)
2007	0.81 (169)	0.79 (137)	0.72 (162)	0.70 (165)	0.83 (173)	0.84 (170)

2008	0.81 (169)	0.79 (137)	0.70 (162)	0.69 (165)	0.84 (173)	0.84 (170)
2009	0.81 (169)	0.79 (137)	0.71 (162)	0.70 (165)	0.83 (173)	0.84 (170)
2010	0.81 (169)	0.80 (137)		0.68 (165)	0.84 (173)	0.84 (170)
2011	0.81 (170)	0.79 (137)		0.69 (165)	0.83 (174)	0.83 (172)
2012	0.81 (170)	0.80 (137)		0.68 (166)	0.83 (174)	0.86 (169)
2013	0.79 (170)	0.78 (137)		0.69 (166)	0.82 (174)	0.84 (170)
2014	0.80 (170)	0.78 (137)		0.68 (166)	0.80 (174)	0.85 (169)
2015	0.80 (170)	0.78 (137)		0.69 (166)	0.80 (174)	0.85 (167)
2016	0.79 (170)	0.78 (137)		0.70 (166)	0.81 (174)	0.86 (169)
2017	0.78 (170)	0.78 (137)		0.69 (166)	0.79 (174)	0.86 (173)

Pearson's correlation coefficients; n in parentheses; all coefficients significant at the $p < 0.001$ level.

APPENDIX B. Regression results summaries of replicated studies.

Table B1. Replication estimates of Bäck and Hadenius' (2008) study.

	Replication	QOG	HSI	VDEM
	(1)	(2)	(3)	(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

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 B2. Replication estimates of Carbone and Memoli's (2015) study.

	Replication	FSI	QOG	HSI	SFI	WGI	CPI	VDEM
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Democracy	0.089*** (0.022)	0.086*** (0.014)	0.038* (0.016)	0.036** (0.011)	0.087*** (0.014)	0.031* (0.015)	0.026 (0.017)	0.067** (0.022)
Democracy ²	0.069*** (0.019)	0.056*** (0.014)	0.020* (0.010)	0.025* (0.011)	0.052*** (0.012)	0.015 (0.013)	0.012 (0.014)	0.018 (0.015)
Duration of democracy	-0.003 (0.002)	-0.001 (0.001)	-0.0004 (0.001)	-0.0001 (0.001)	0.001 (0.001)	-0.0001 (0.001)	0.0001 (0.001)	0.0002 (0.002)
Duration of democracy ²	-0.00001 (0.00003)	-0.00001 (0.00001)	0.00001 (0.00002)	0.000003 (0.00001)	-0.00001 (0.00002)	0.00001 (0.00002)	-0.00001 (0.00002)	-0.000004 (0.00002)
Democracy*Duration of democracy	0.003* (0.001)	0.001 (0.001)	-0.0003 (0.001)	-0.001 (0.001)	-0.001 (0.001)	0.00001 (0.001)	0.001 (0.001)	0.001 (0.001)
Ethnic fractionalization	-0.093 (0.069)	-0.120** (0.045)	-0.061 (0.049)	-0.100** (0.036)	-0.172*** (0.045)	-0.074 (0.040)	0.021 (0.038)	0.064 (0.040)
Log(GDP/capita)	0.099*** (0.012)	0.032*** (0.006)	0.031*** (0.007)	0.054*** (0.006)	0.087*** (0.006)	0.040*** (0.006)	0.046*** (0.008)	0.030*** (0.007)
Log(Area)	-0.025** (0.008)	-0.017* (0.007)	-0.007 (0.009)	-0.005 (0.006)	-0.025*** (0.005)	-0.017* (0.007)	-0.026** (0.009)	-0.013* (0.007)

Constant	0.063 (0.152)	0.356*** (0.098)	0.265* (0.121)	0.214** (0.080)	0.278** (0.097)	0.352*** (0.089)	0.201* (0.095)	0.247* (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	0.80	0.58	0.55	0.49
Wald chi-square (8)	248.69	207.60	72.37	284.75	730.80	175.32	96.22	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	121	122	122	122
Observations	344	344	284	344	344	345	343	345

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 B3. Replication estimates of Grassi and Memoli's (2016) study.

	Replication	QOG	SFI	WGI	CPI	VDEM
	(1)	(2)	(3)	(4)	(5)	(6)
Lagged stateness	0.169*** (0.021)	-0.036 (0.044)	-0.016 (0.035)	0.009 (0.012)	0.017 (0.025)	0.017 (0.015)

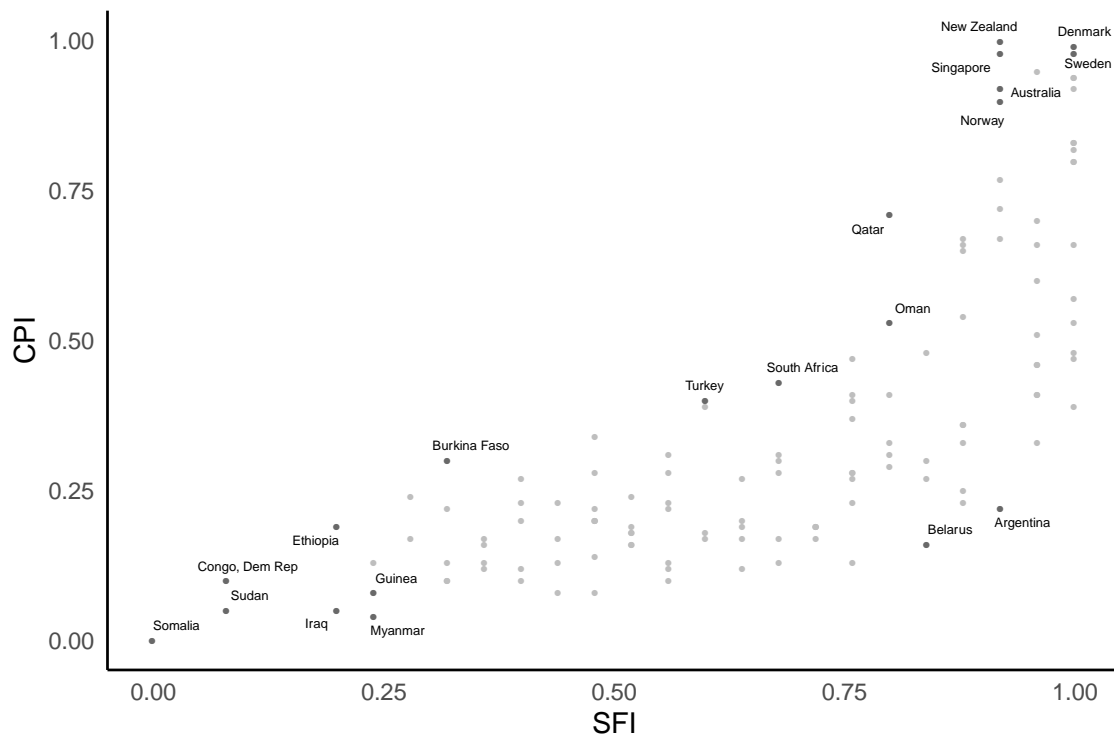
Democracy	-0.034***	0.042***	-0.009	-0.001	0.011*	-0.005
	(0.007)	(0.009)	(0.006)	(0.006)	(0.005)	(0.011)
Democracy ²	0.004*	-0.004	0.005**	0.0003	-0.001	0.005
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
Executive partisan balance	0.133**	-0.063	0.116***	-0.041	-0.104*	0.066
	(0.042)	(0.035)	(0.031)	(0.062)	(0.049)	(0.072)
Level of economic development (1994)	0.095	0.068	0.286***	0.165***	0.143*	0.175**
	(0.059)	(0.061)	(0.048)	(0.045)	(0.072)	(0.066)
Land size (km ²)	0.018	0.007	-0.029	-0.005	-0.002	-0.018
	(0.020)	(0.021)	(0.016)	(0.016)	(0.024)	(0.021)
Oil rents	-0.130	-0.065	-0.142	-0.219**	-0.141*	-0.085
	(0.163)	(0.114)	(0.146)	(0.078)	(0.066)	(0.153)
Log(Ethnic fractionalization)	-0.068	-0.049	-0.155*	-0.016	-0.100	-0.078
	(0.128)	(0.141)	(0.071)	(0.106)	(0.150)	(0.085)
Constant	-0.681	-0.241	-1.616***	-0.773*	-0.840	-0.979
	(0.467)	(0.343)	(0.400)	(0.375)	(0.604)	(0.533)
Sigma_u	0.15	0.17	0.15	0.12	0.17	0.18
Sigma_e	0.05	0.08	0.08	0.04	0.06	0.04
Rho	0.91	0.80	0.79	0.92	0.90	0.94

Between R ²	0.30	0.15	0.76	0.89	0.38	0.49
Wald chi-square (8)	1393.54	228.57	170.67	52.03	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

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.

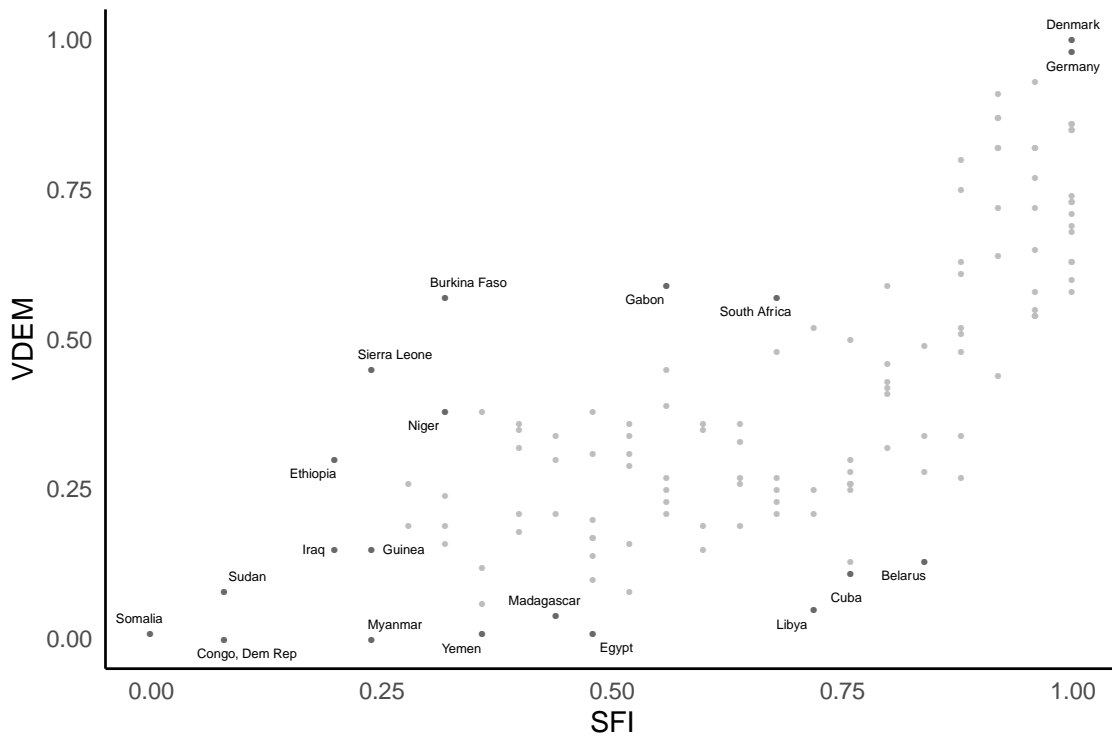
APPENDIX C. Scatter plots of measures of state capacity.

Figure C1. Level of state capacity (2009), CPI and SFI.



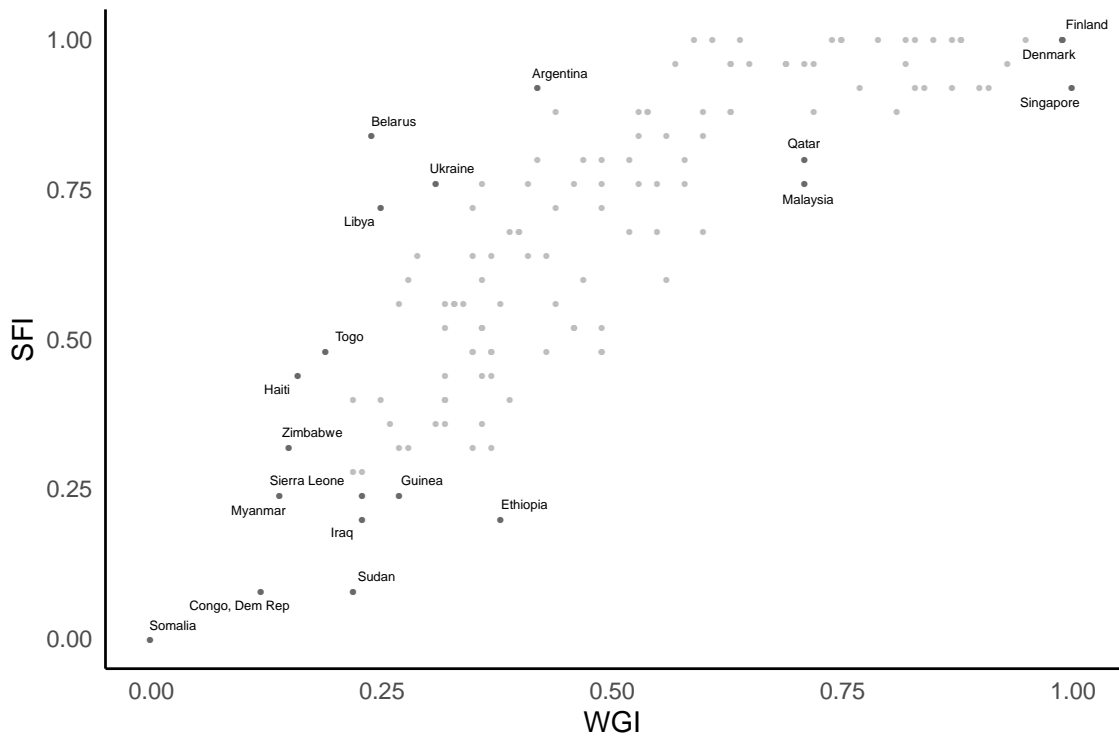
Scores are normalized to range from 0 to 1.

Figure C2. Level of state capacity (2009), VDEM and SFI.



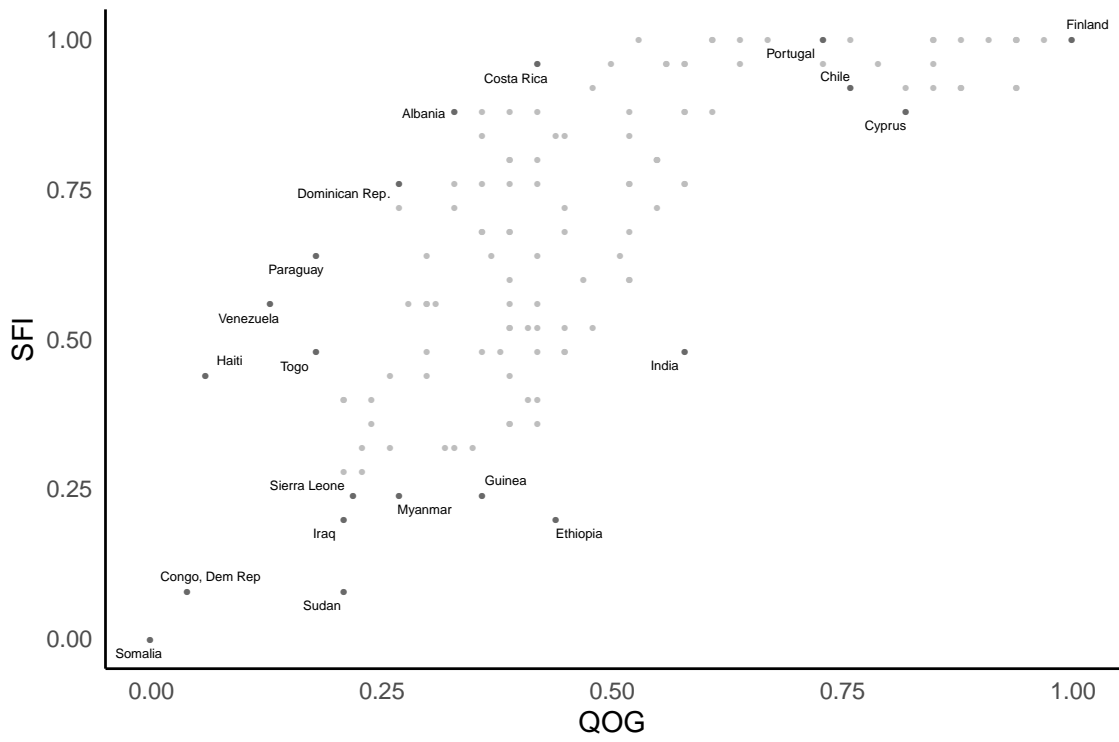
Scores are normalized to range from 0 to 1.

Figure C3. Level of state capacity (2009), SFI and WGI.



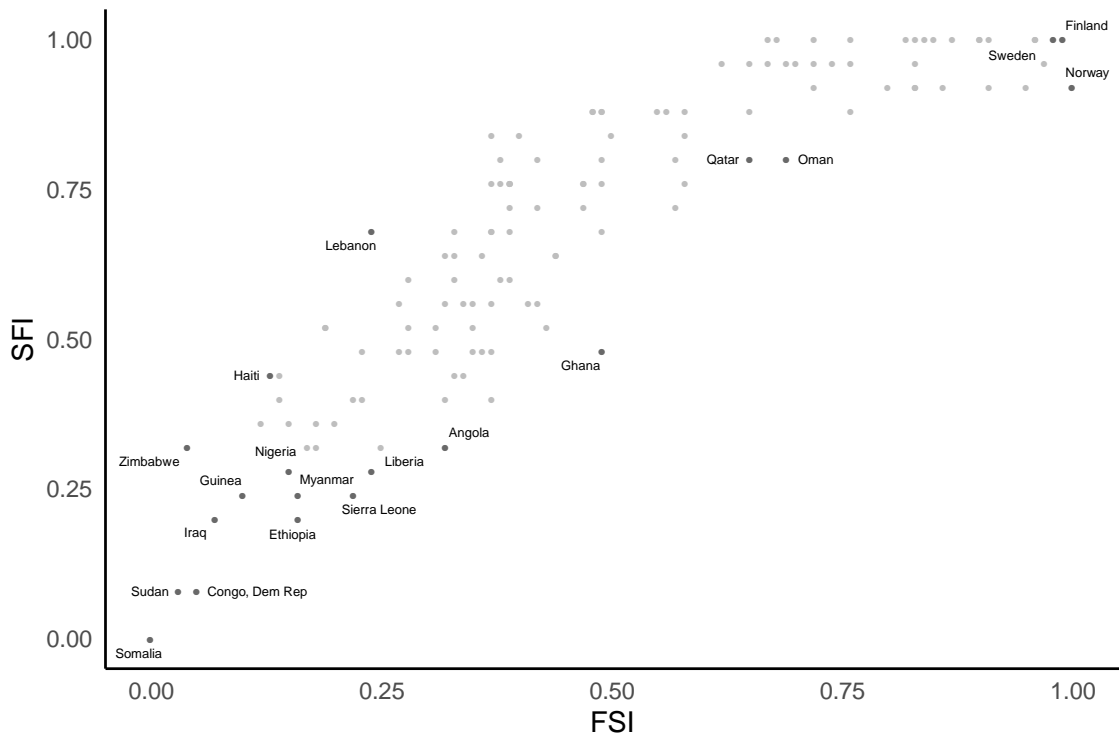
Scores are normalized to range from 0 to 1.

Figure C4. Level of state capacity (2009), SFI and QOG.



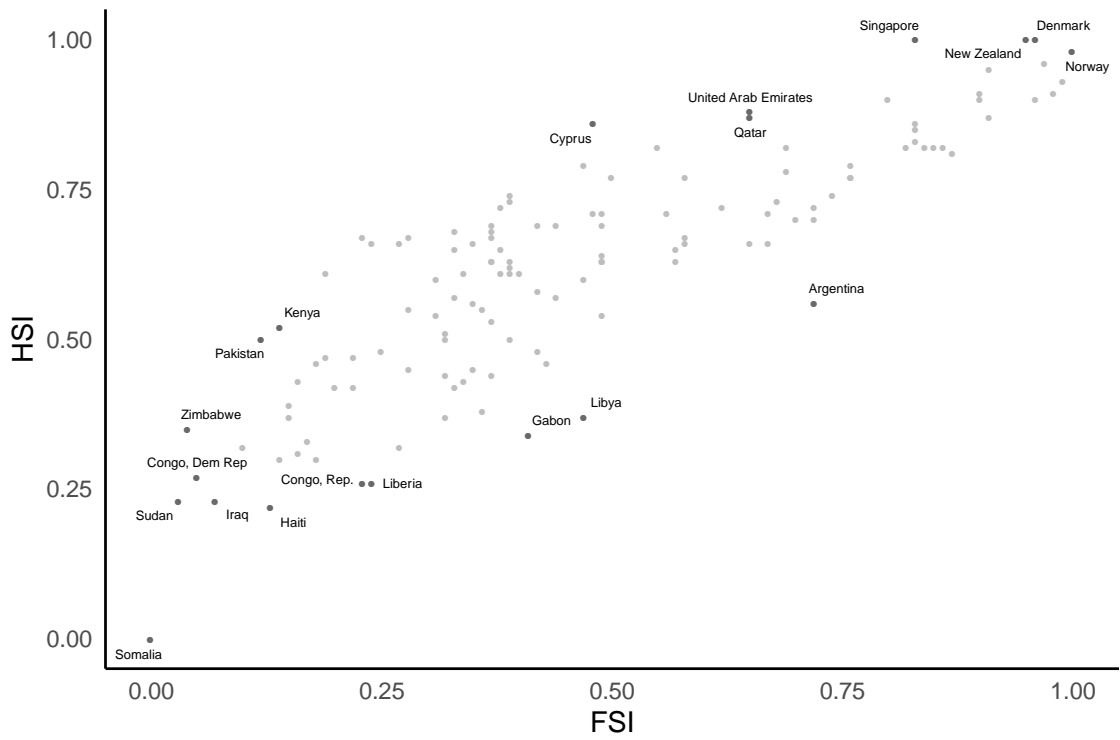
Scores are normalized to range from 0 to 1.

Figure C5. Level of state capacity (2009), SFI and FSI.



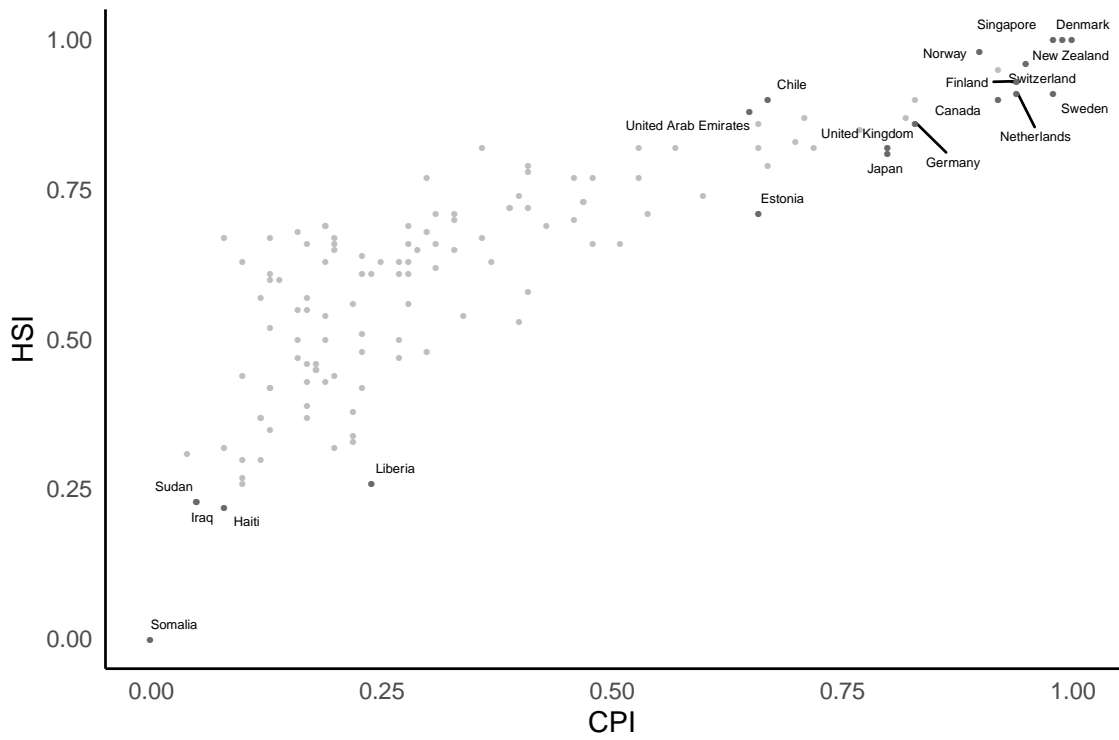
Scores are normalized to range from 0 to 1.

Figure C6. Level of state capacity (2009), HSI and SFI.



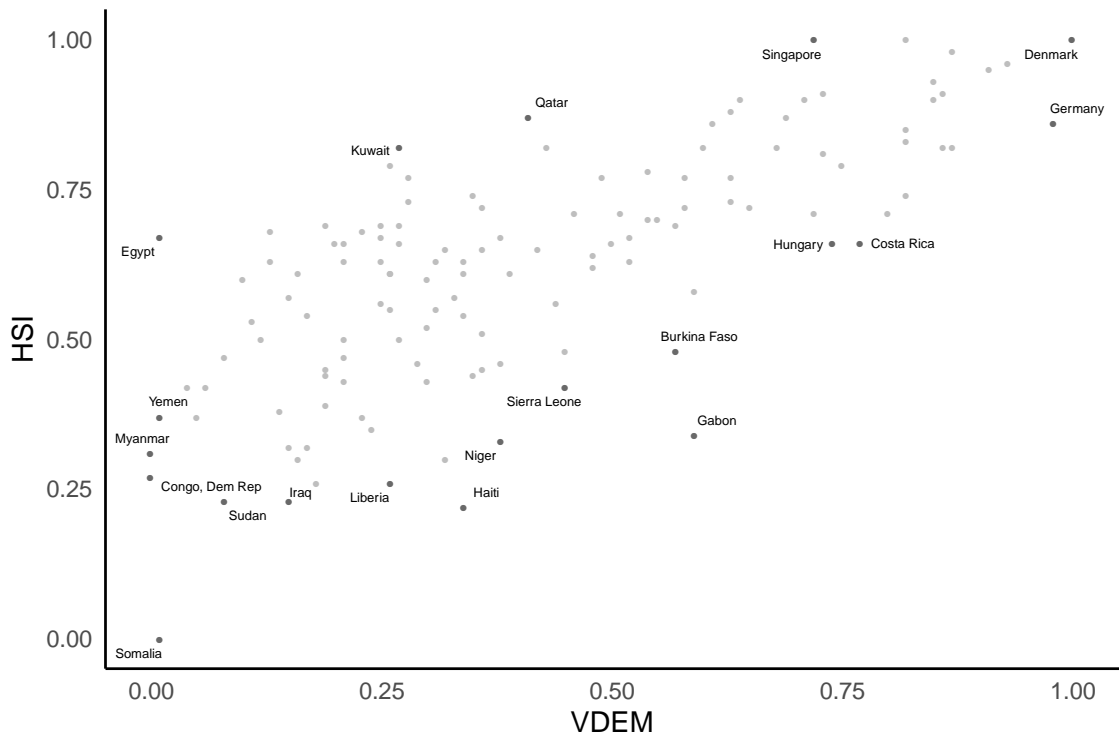
Scores are normalized to range from 0 to 1.

Figure C7. Level of state capacity (2009), HSI and CPI.



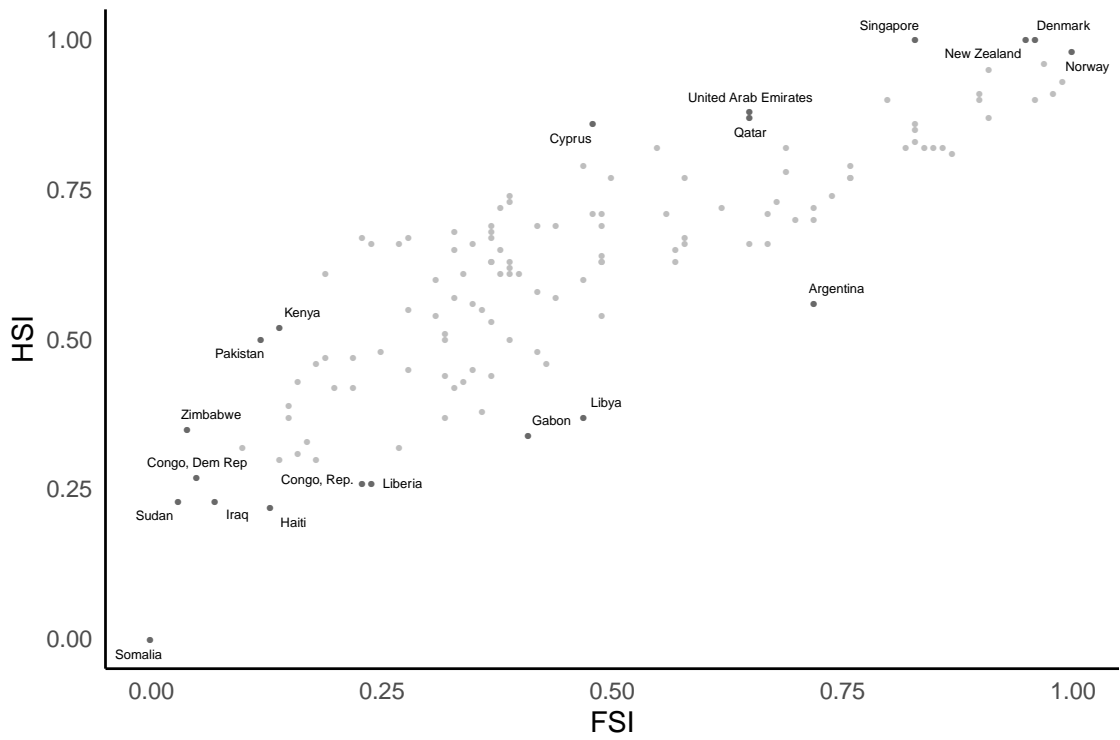
Scores are normalized to range from 0 to 1.

Figure C8. Level of state capacity (2009), HSI and VDEM.



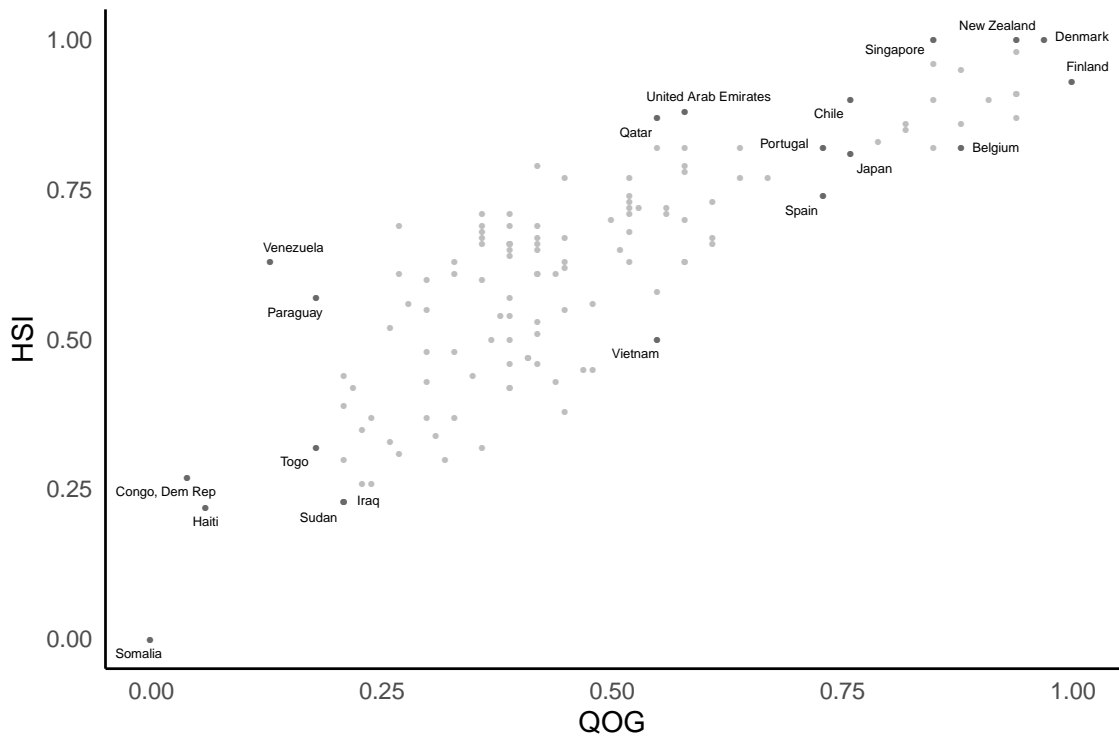
Scores are normalized to range from 0 to 1.

Figure C9. Level of state capacity (2009), HSI and FSI.



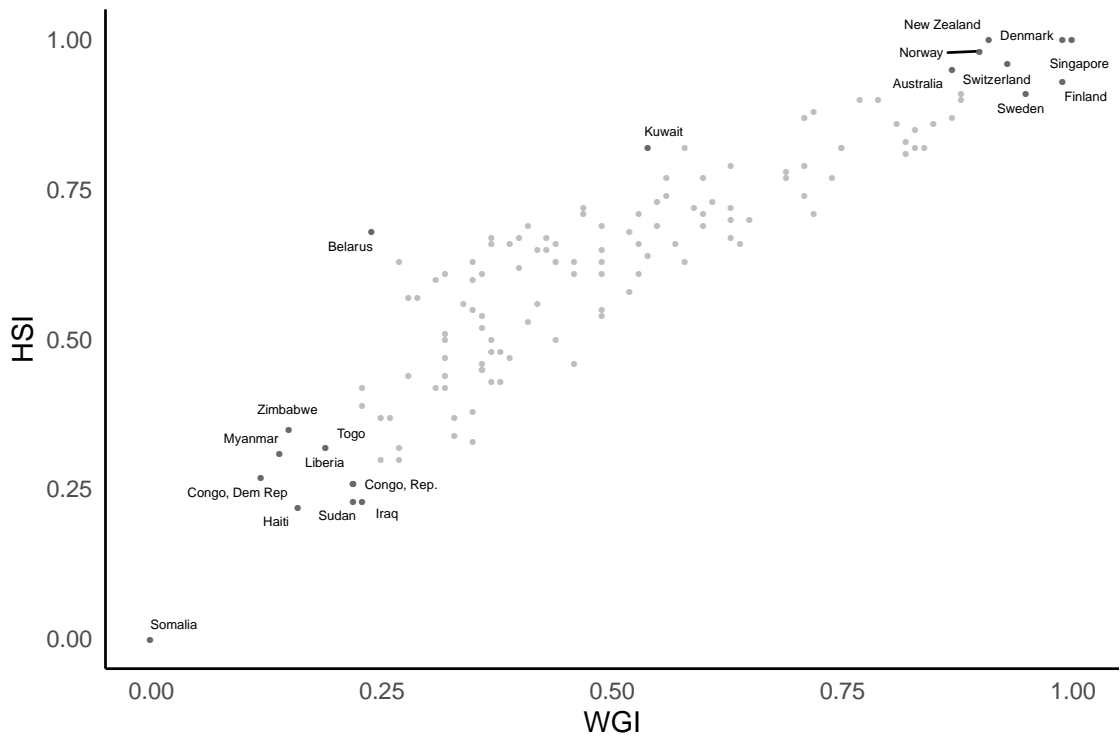
Scores are normalized to range from 0 to 1.

Figure C10. Level of state capacity (2009), HSI and QOG.



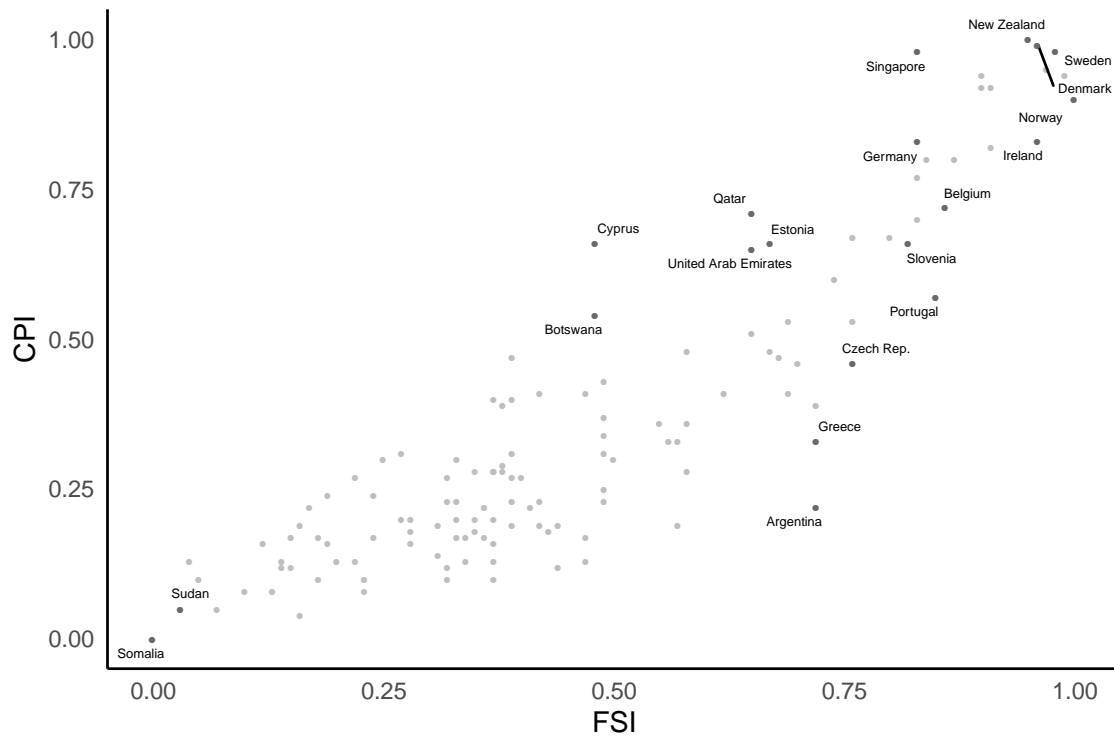
Scores are normalized to range from 0 to 1.

Figure C11. Level of state capacity (2009), HSI and WGI.



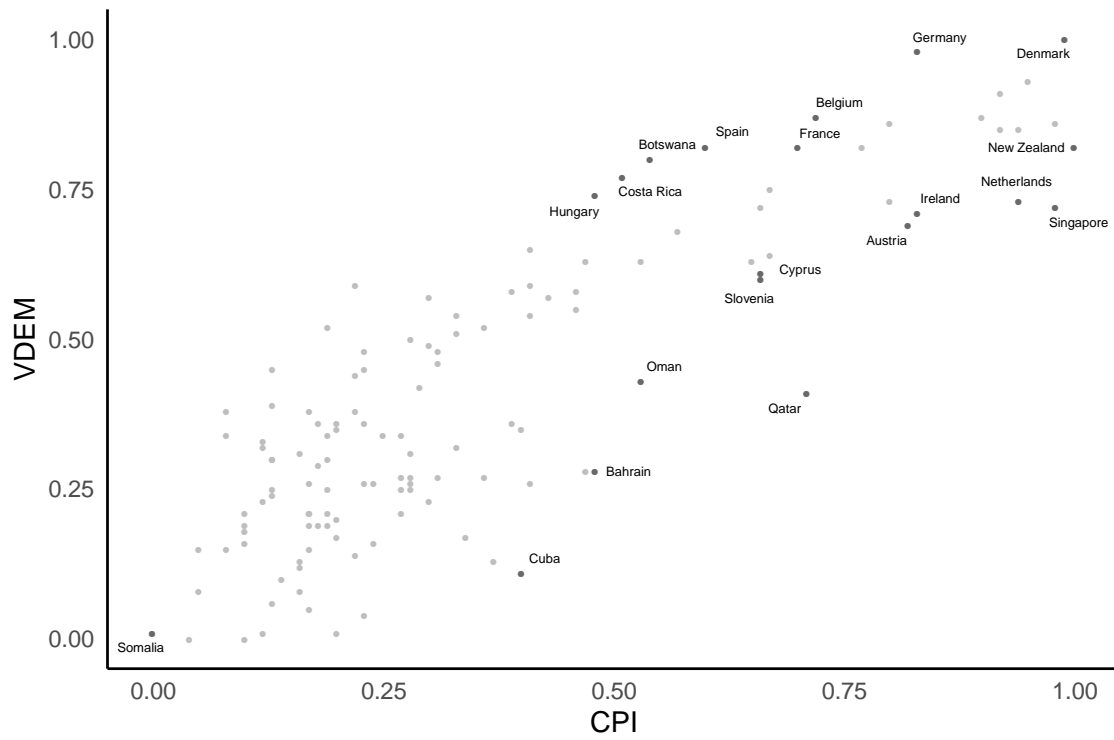
Scores are normalized to range from 0 to 1.

Figure C12. Level of state capacity (2009), CPI and FSI.



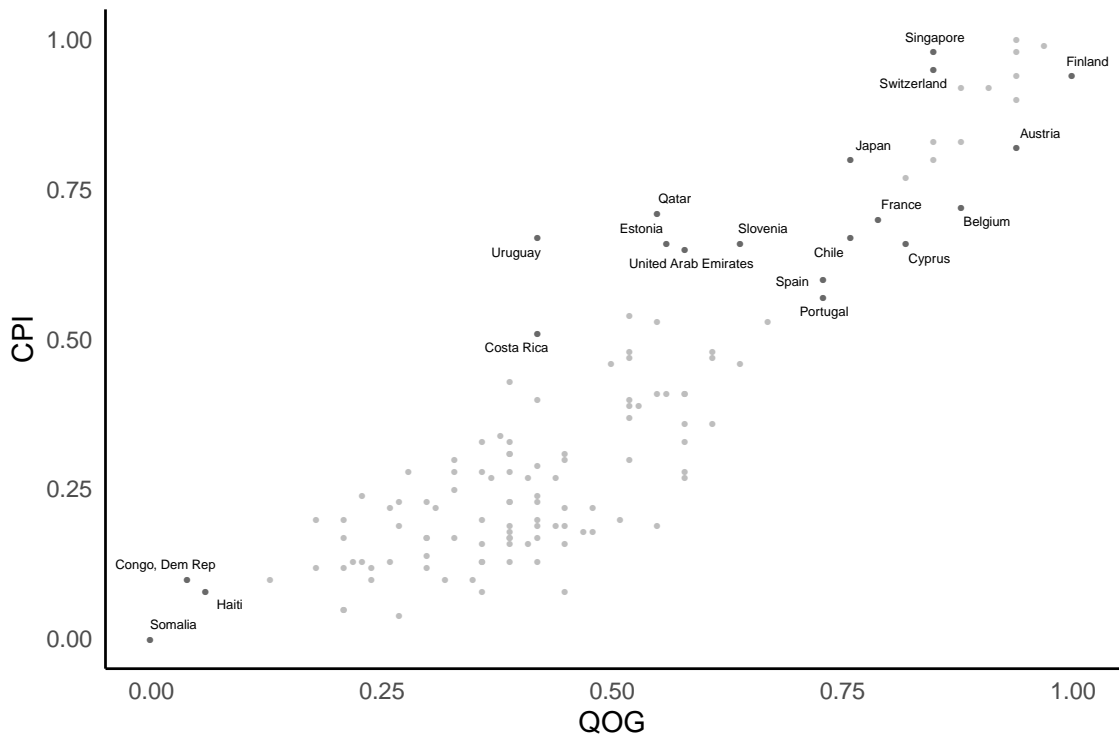
Scores are normalized to range from 0 to 1.

Figure C13. Level of state capacity (2009), VDEM and CPI.



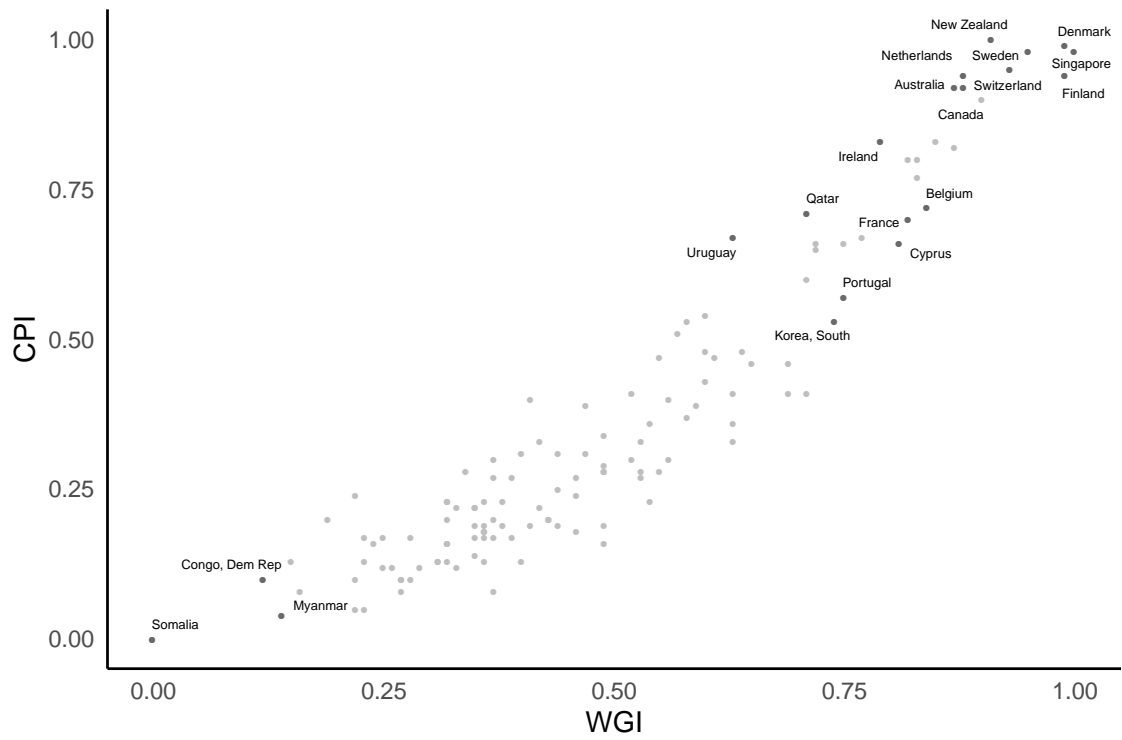
Scores are normalized to range from 0 to 1.

Figure C14. Level of state capacity (2009), CPI and QOG.



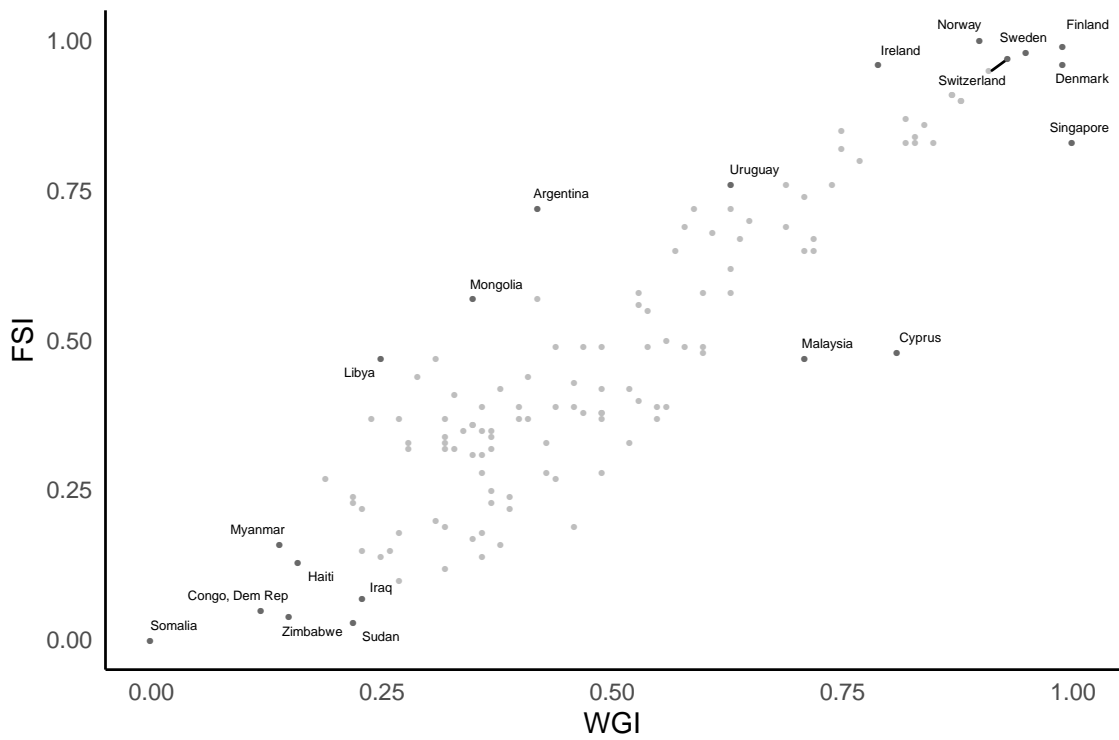
Scores are normalized to range from 0 to 1.

Figure C15. Level of state capacity (2009), CPI and WGI.



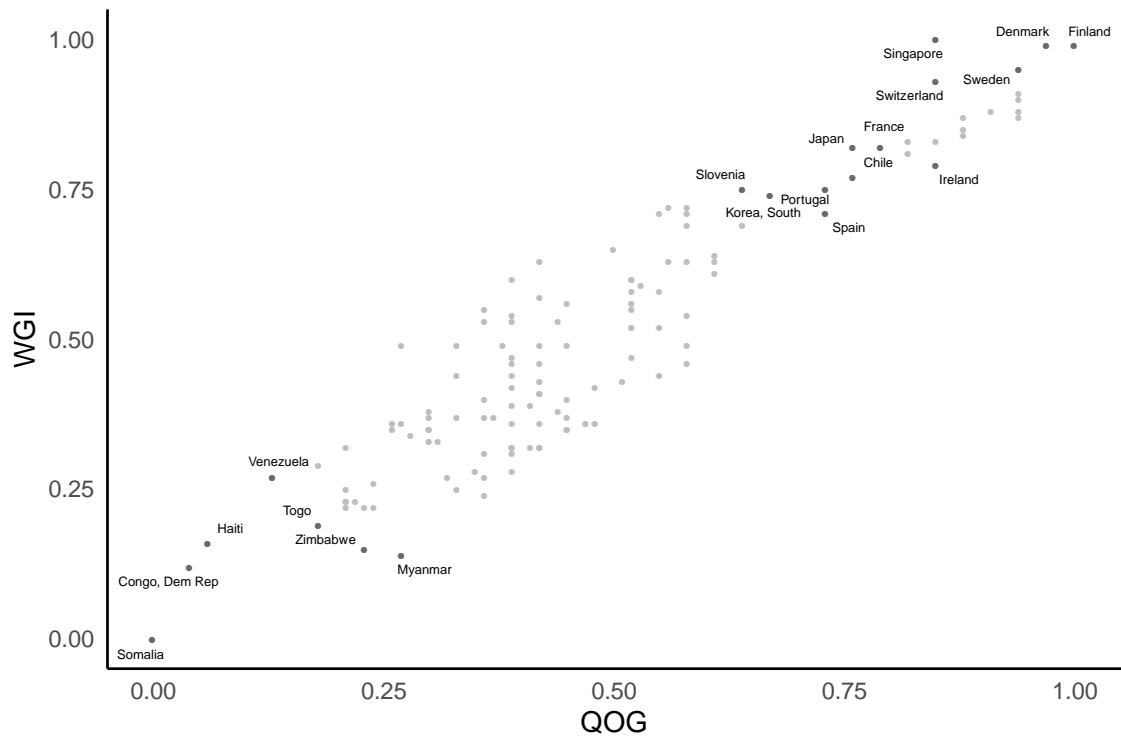
Scores are normalized to range from 0 to 1.

Figure C16. Level of state capacity (2009), FSI and WGI.



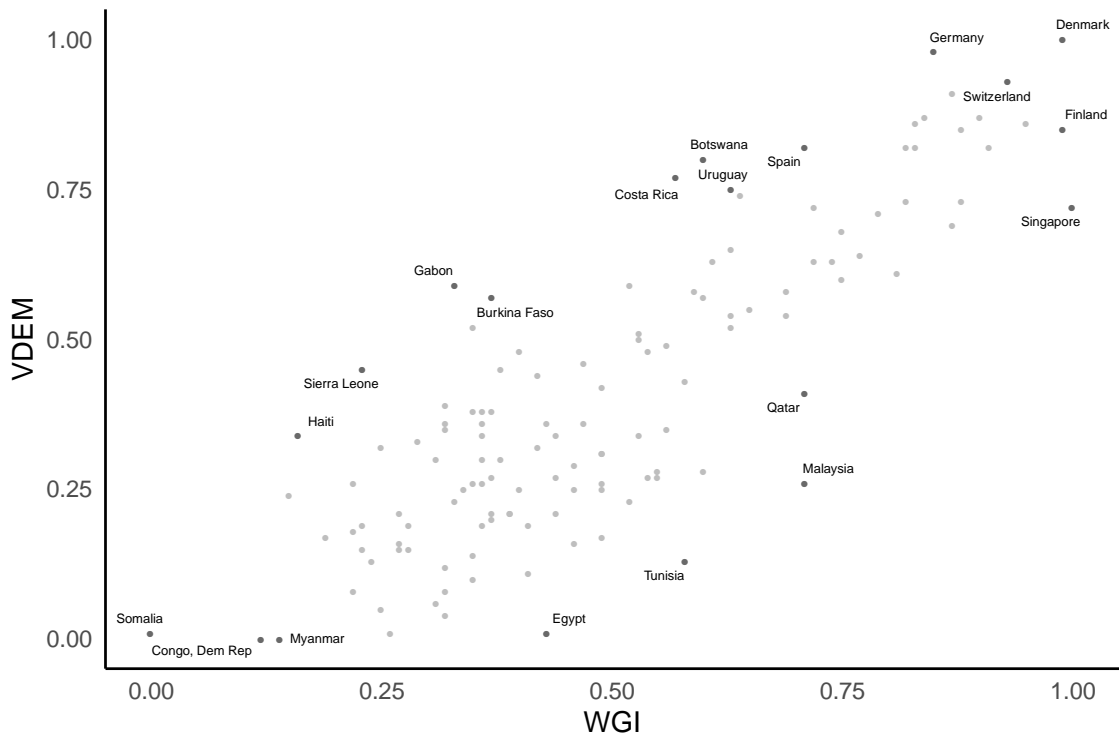
Scores are normalized to range from 0 to 1.

Figure C17. Level of state capacity (2009), WGI and QOG.



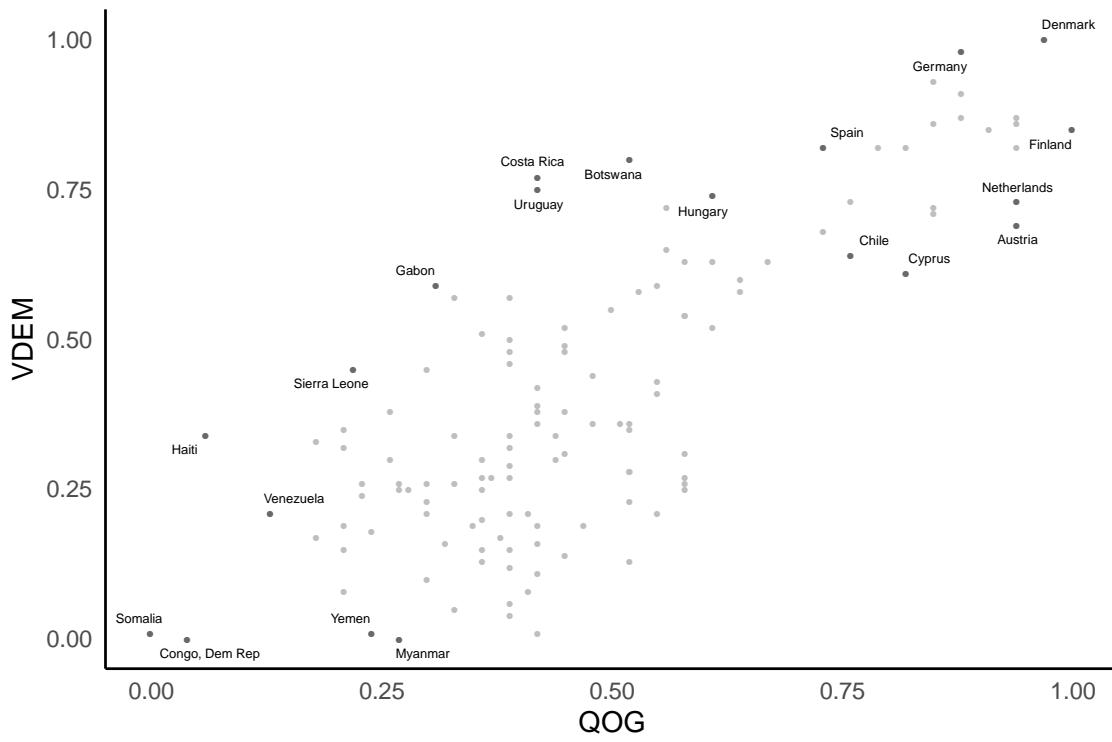
Scores are normalized to range from 0 to 1.

Figure C18. Level of state capacity (2009), VDEM and WGI.



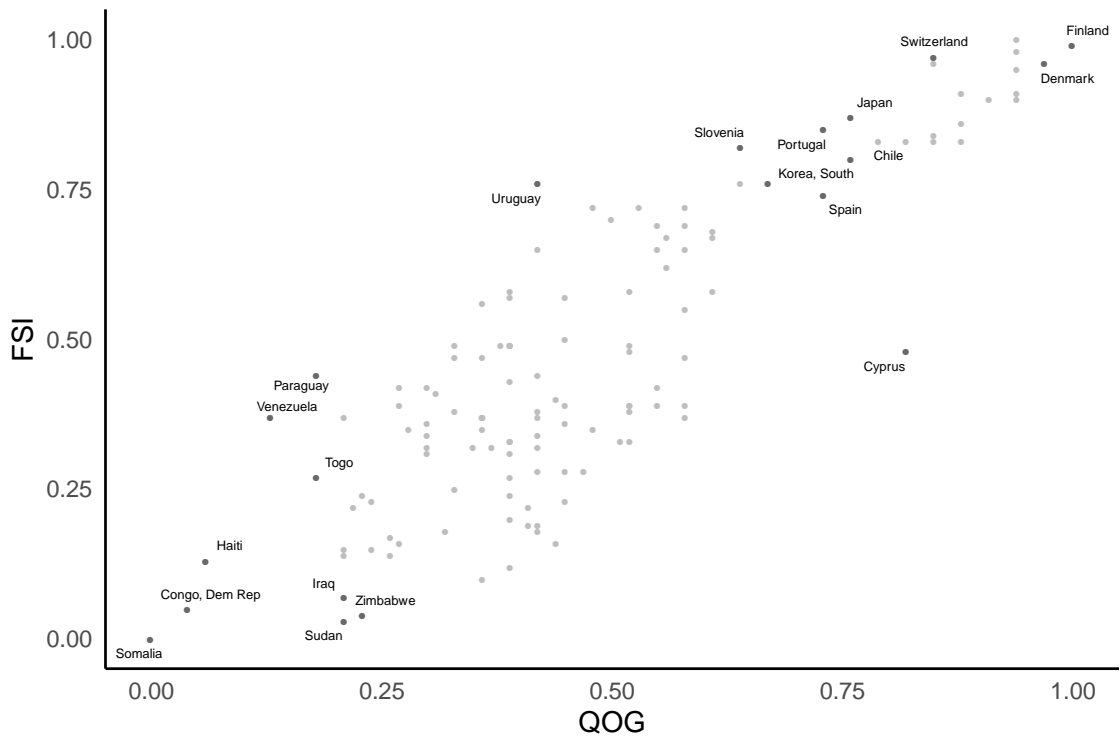
Scores are normalized to range from 0 to 1.

Figure C19. Level of state capacity (2009), VDEM and QOG.



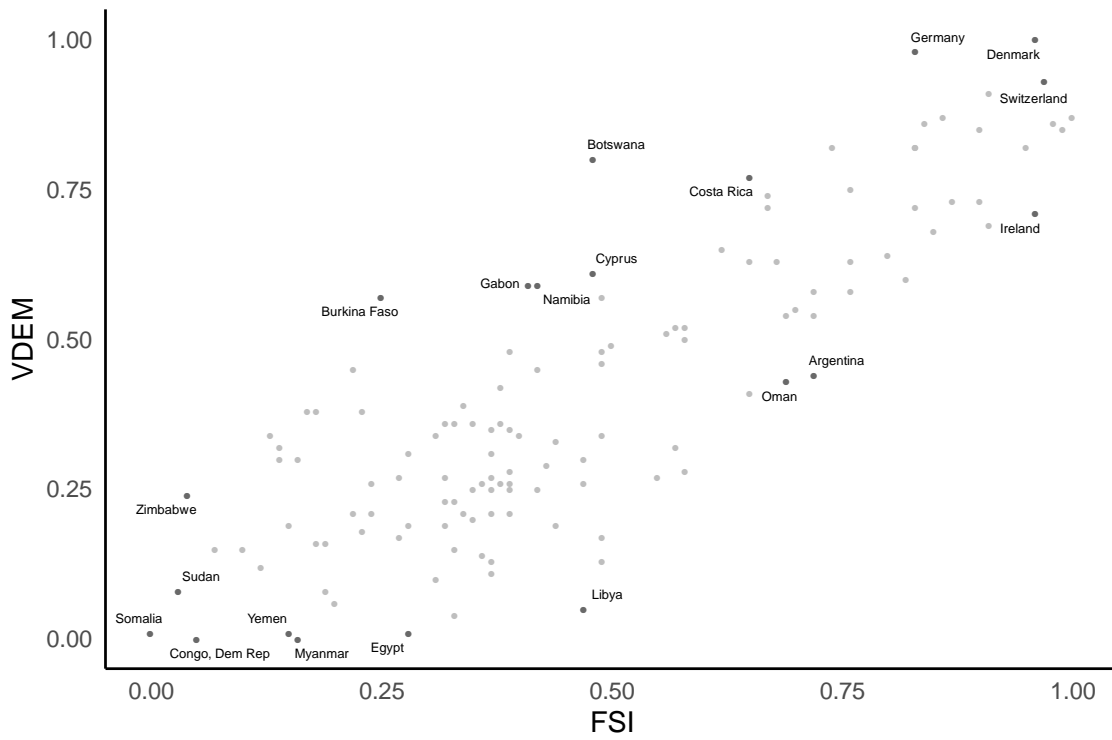
Scores are normalized to range from 0 to 1.

Figure C20. Level of state capacity (2009), FSI and QOG.



Scores are normalized to range from 0 to 1.

Figure C21. Level of state capacity (2009), VDEM and FSI.



Scores are normalized to range from 0 to 1.

APPENDIX D. Country-specific rating divergencies among measures of state capacity.

Table D1. Obs. with the highest rating discrepancy between SFI and CPI in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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

Scores are normalized to range from 0 to 1.

Table D2. Obs. with the highest rating discrepancy between SFI and VDEM in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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

Scores are normalized to range from 0 to 1.

Table D3. Obs. with the highest rating discrepancy between SFI and WGI in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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

Scores are normalized to range from 0 to 1.

Table D4. Obs. with the highest rating discrepancy between SFI and QOG in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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

Scores are normalized to range from 0 to 1.

Table D5. Obs. with the highest rating discrepancy between SFI and FSI in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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

Scores are normalized to range from 0 to 1.

Table D6. Obs. with the highest rating discrepancy between SFI and HSI in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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

Scores are normalized to range from 0 to 1.

Table D7. Obs. with the highest rating discrepancy between HSI and CPI in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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

Scores are normalized to range from 0 to 1.

Table D8. Obs. with the highest rating discrepancy between HSI and VDEM in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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

Scores are normalized to range from 0 to 1.

Table D9. Obs. with the highest rating discrepancy between HSI and FSI in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
Iran	0.44	0.67 with HSI	0.23 with FSI
Sri Lanka	0.43	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

Scores are normalized to range from 0 to 1.

Table D10. Obs. with the highest rating discrepancy between HSI and QOG in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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

Scores are normalized to range from 0 to 1.

Table D11. Obs. with the highest rating discrepancy between HSI and WGI in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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	0.38 with WGI

Scores are normalized to range from 0 to 1.

Table D12. Obs. with the highest rating discrepancy between CPI and FSI in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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

Scores are normalized to range from 0 to 1.

Table D13. Obs. with the highest rating discrepancy between CPI and VDEM in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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

Scores are normalized to range from 0 to 1.

Table D14. Obs. with the highest rating discrepancy between CPI and QOG in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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

Scores are normalized to range from 0 to 1.

Table D15. Obs. with the highest rating discrepancy between WGI and CPI in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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

Scores are normalized to range from 0 to 1.

Table D16. Obs. with the highest rating discrepancy between WGI and FSI in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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

Scores are normalized to range from 0 to 1.

Table D17. Obs. with the highest rating discrepancy between WGI and QOG in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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

Scores are normalized to range from 0 to 1.

Table D18. Obs. with the highest rating discrepancy between WGI and VDEM in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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

Scores are normalized to range from 0 to 1.

Table D19. Obs. with the highest rating discrepancy between QOG and VDEM in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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

Scores are normalized to range from 0 to 1.

Table D20. Obs. with the highest rating discrepancy between QOG and FSI in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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	0.13 with QOG
Guinea	0.26	0.36 with QOG	0.10 with FSI

Scores are normalized to range from 0 to 1.

Table D21. Obs. with the highest rating discrepancy between FSI and VDEM in 2009.

Observation (country)	Discrepancy (in units)	Higher score	Lower score
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

Scores are normalized to range from 0 to 1.