A New Indicator of Coalition Size: Tests Against Standard Regime-Type Indicators

Bruce Bueno de Mesquita[1] and Alastair Smith[2]

Wilf Family Department of Politics

New York University

19 West 4th St, New York NY 10012

[1]Bruce.buenodemesquita@nyu.edu

[2]Alastair.smith@nyu.edu

Janary 2022

Abstract

A new measure for selectorate's theory central concept, winning coalition size, is developed using indicators developed by the V-Dem project (Coppedge et al., 2021). To demonstrate the measure's efficacy, we examine its ability to account for government provision of public and private goods, a central prediction in selection theory. The winning coalition measure is compared to eight widely-used indicators of regime type in analyses of 30 measures of government policy provision. Based on Vuong tests and Akaike information criteria the new measure generally outperforms the alternative regime measures whether the assessment is across all regimes, within non-democratic regimes or within democratic settings. The new indicator provides an improved tool for testing the nuances of selectorate theory.

Word Count: 8947

The selectorate theory (Bueno de Mesquita et al., 2003) offers a conceptually simple, intuitive explanation of differences in organizational approaches to policy choices with particular empirical emphasis on these choices by governments. However, its empirical assessment has lacked a convincing indicator that estimates winning coalition size in governments. We propose a solution to that empirical impediment by offering a new, continuous indicator of coalition size based on indicators from the V-Dem Project (Coppedge et al., 2021). The new indicator's explanatory power is compared to eight widely-used indicators of regime type across thirty important dependent variables. Within fixed effects regression analyses all indicators perform well in accounting for policy and welfare outcomes. However, the results show that the new coalition indicator significantly outperforms the alternatives, providing an improved tool for testing the nuances of selectorate theory and its standing relative to alternative approaches to government performance.

In the selectorate perspective, policy choices and resource allocations by governments are logical consequences of the incentives that institutions impose on decision makers whose primary objective is to survive in power. Which policies best enhance leader survival depends on the institutional characteristics that determine the size of the selectorate (denoted as S) and the size of the winning coalition (denoted as W). The selectorate is the set or proportion of residents in a polity who have at least a nominal say in choosing leaders and W is the minimal subset or proportion of S whose support is essential for an incumbent to retain power or for a challenger to come to power. To make comparisons across polities, S and W are best thought of in terms of the similarity or differences in institutional configurations that encourage broader or narrow governmental accountability. That way, places such as Denmark and the United States should have highly similar selectorate and winning coalition sizes despite substantial differences in their respective populations (and, hence, absolute numbers in, for instance, W) while places like Denmark and Eritrea should have radically different coalition sizes despite having equivalently sized populations. Hence, S and W will be approximated based on each country's institutionalized mechanisms each year that act

to facilitate or restrict the number of people whose support is nominally required (S) or is essential (W) for securing and holding power.

A central result of selectorate theory is that as W increases governments increase their provision of public goods relative to private goods. To survive in office, leaders need to provide their supporters with greater rewards than their supporters can expect to receive if they defect, giving support to a rival politician. Leaders reward their supporters with a combination of public goods that benefit everyone and private goods that directly benefit only members of the winning coalition. Coalition size determines the efficient mix of public and private goods. If a unit of public goods costs p and, implicitly, the cost of private goods is proportional to the number of people who must receive them (W), then as coalition size increases private goods become relatively more expensive. This effective price increase induces leaders to shift their policy provision towards the relatively cheaper public goods (Bueno de Mesquita et al., 2003). Ascertaining the empirical veracity of these and other claims requires better measurement of winning coalition size.

1   Calls for Better Measurement of Coalition Size

Tests of selectorate theory have shown strong consistency between theoretical expectations and outcomes across a wide array of dependent variables. However, the estimation of coalition size has, appropriately, been subject to important criticisms (Clarke and Stone, 2008; Gallagher and Hanson, 2015; Kennedy, 2009). The two major criticisms revolve around the nexus between theory and evidence. First, there are well-justified concerns about the translation of the concept of coalition size into an observable, measured quantity. Indeed, when the authors of the selectorate theory first presented indicators for coalition and selectorate size they wrote, "The indicators of W and S are much too crude. . . . We hope that [the theory] . . . will motivate the search for better ways to estimate the institutions with which we are concerned" (Bueno de Mesquita et al., 2003, p.133). The urge to find better

indicators has been echoed by critics but thusfar a better indicator of coalition size has not been forthcoming (Gallagher and Hanson, 2015). We believe that the demanding tests and evidence to follow will be persuasive that a better indicator of coalition size is now available with coverage that includes virtually all countries spanning as long as from 1789 to the present.

A second concern articulated by some critics is that the selectorate theory seems to work well in distinguishing policy choices between democracies and non-democracies but it appears to fail to distinguish choices within the range of non-democratic regimes (Gallagher and Hanson, 2015; Kennedy, 2009). This critique conflates the theory's logical implications with its original, crude measurements. To allay this concern, we present tests on the subset of governments defined by the Polity project as being autocracies (Polity2$\leq$-6) or anocracies (i.e., -5 $\leq$ Polity2 $\leq$ 5) and, as well, we replicate the tests just for those governments that are assessed to be democratic by the Polity2 cut-point of 6 or greater (Vreeland, 2008). These tests may mitigate previous concerns as they show that the new indicator of W outperforms common regime-type indicators both across and within regime categories.

The need to improve the fit between theoretical concept and empirical measure is undeniable and important. As Kevin Clarke and Randall Stone, two early critics, have noted, "If correct, the selectorate theory has important implications for the long-standing debate over whether democracy should be understood primarily in terms of attitudes and behavior . . . or in terms of institutions . . . . According to the theory, the differences between democratic and authoritarian politics and the differences in politics among democracies are reducible to institutional incentives; thus, institutions and not behaviors are the crux of the matter" (Clarke and Stone, 2008, p.387). That is why the new indicators for W and S emphasize institutional mechanisms rather than behavioral considerations.

2 Estimating Coalition Size

Recall that winning coalition size and selectorate size are defined as the proportion of people a leader is beholden to in order to retain power and the size of the group from which these supporters are drawn. W and S are institutional features and, as such, measures of these concepts should emphasize the (often de facto) rules and procedures that govern attaining and retaining political power. As we discuss below, some measures of democracies are defined, at least partially, in terms of the policies and behaviors these institutions induce.

The new indicator of coalition size, W, is a continuous index based on institutional variables in the V-Dem data set, version 11 (2021). The V-Dem project is a broad and encompassing endeavor that codes virtually all nations since 1789 using item response theory (IRT) (Pemstein et al., 2019; Marquardt and Pemstein, 2017) based on surveys of country experts. Of V-Dem's vast array of issues, including political competition, policy outcomes and freedoms, our indicator uses the following institutional dimensions:

1. Autonomy of election monitoring body (v2elembaut)

2. Opposition parties' autonomy (v2psoppaut)

3. Barriers to political party participation (v2psbars)

4. Closed Succession: Indicators of succession by heredity or within a military or single party setting (constructed from $v2x\_ex\_hereditary$, $v2x\_ex\_military$, and $v2x\_ex\_party$)

These four primarily institutional dimensions measure whether coalition membership is broadly accessible or is exceedingly restrictive or anywhere in between. The selected institutional indicators facilitate the estimation of variation in the credible, institutionalized competition for coalition membership across polities and across time.

The first element, autonomy of the election monitoring body, establishes the extent to which the administrative body responsible for monitoring elections is controlled by the government or other central authorities, like the military, or, instead, is independent and applies

election laws impartially. This first component in our indicator of W assesses the extent to which there is an institutionalized mechanism that allows a large class of selectors – the electorate – a say in choosing the government or, instead, restricts that say to a smaller class of selectors, such as the incumbent regime or the military. Clearly, the size of the prospective winning coalition increases as the autonomy of such a board is institutionalized, assuring its independent operation. Conversely, restrictions in the operation of such a board reduces the opportunity for some segments of the selectorate to compete for a place in the winning coalition. And, of course, it is possible that no such board exists, meaning that there is no electoral process or election monitoring that opens the prospective coalition beyond whatever the current leadership wants.

An autonomous election monitoring board is an important element in assuring the opportunity for a large number of selectors to determine leadership succession but it is not enough to assure a large coalition. After all, a one-party state, even if it has an autonomous election monitoring body, still restricts who is needed to form a winning coalition to those who endorse the one approved party. Hence, the second component of our indicator of W, opposition parties' autonomy, is V-Dem's measure of the degree to which political parties can operate independently of the incumbent regime. If a government has institutionalized a one-party state, such as Cuba has done, or if a government has institutionalized its authority to select parties (and their candidates) that are permitted to compete in elections, as Russia and Iran have done, then the government has restricted the size and composition of a winning coalition to members or supporters of authorized parties only, shrinking the subset of selectors (W) needed to sustain a leader in power. If, instead, there are many political parties competing for power, then more selectors will probably be attracted to participate in choosing a successor leadership and that means more support – a larger winning coalition – is likely to be needed to win and to sustain power. As Levitsky and Ziblatt (2018) note, democracy dies if there are not at least two political parties realistically competing for power.

The first two variables in the indicator for W help shape accountable government, but

they alone do not assure a large coalition. Hence, the third element in our indicator of W evaluates the degree of encumbrance in participating in the government-selection process. There can be a large winning coalition only if the political structure has institutionalized an unencumbered right to form competitive parties; if the system has institutionalized the right to have multiple parties; and if the competition takes place over an expansive pool of people that allows for the prospect of a large winning coalition.

The extent of institutional fulfillment of the three variables – autonomous monitoring, autonomous opposition party participation and unencumbered rights to form parties – collectively shape an electoral political system's winning coalition size. The more an electoral government falls down on one or more of these conditions, the smaller its coalition is likely to be. The more an electoral system has institutionalized the absence of encumbrances to party formation, the right to oppose the government and the right to have outcomes determined impartially and in light of the law, the larger the coalition is likely to be. But, not all political systems today, let alone over most of history, are electoral. Monarchies and military juntas, for instance, may not bother with even rigged elections. We take two steps in calculating the new indicator of W to reflect appropriate estimates for non-electoral political systems.

We know that political systems, like monarchies and military juntas, that do not hold elections also do not have election monitoring boards, political party autonomy and, indeed, they have full and effective barriers to the creation of opposition parties. While V-Dem has not coded such polities on these variables, we assign V-Dem scores that reflect that these polities do not meet any of these criteria. We code the missing estimates for non-electoral polities for the three variables we have discussed thus far as if they were at the 1st percentile in the distribution on each of these variables. That is, they are equivalent to rigged-election systems with the minimal allowance for opposition parties and for party and electoral autonomy. But, because these polities may, nevertheless, have considerable variation in their relatively small coalition size, we add a final set of conditions as part of our estimation of W.

7

We construct a final component to W from V-Dem's variables $v2x\_ex\_hereditary$, $v2x\_ex\_military$, and $v2x\_ex\_party$. As their labels imply, these last three variables assess the extent to which a person's opportunity to lead a government is limited and depends on their hereditary status, their military position or is determined by a ruling party even if they have widespread support outside of these eligibility constraints. Institutionalized succession that is determined by any one of these three criteria assures that a government has a small winning coalition, consisting, for instance, of a royal family (as in Saudi Arabia) or a group of generals or other military leaders (as in today's Egypt), or a small cadre of party leaders (as in North Korea).[1] We take the maximum value for each country-year from among hereditary, military and party and reverse its direction. That is, smaller values mean greater dependence on hereditary, military or party control and larger values indicate less reliance on such constraints on leadership selection.

Winning coalition size is constructed by summing the 4 standardized parts and then dividing by 4 to take their average standardized score. Specifically, let j index the component, i index nation and t index year.

$$w_{i,t} = \frac{1}{4} \sum_{j=1}^{4} \frac{x_{j,i,t} - \bar{x}_j}{sd(x_j)},$$ (1)

where $x_1 = v2elembaut$, $x_2 = v2psoppaut$, $x_3 = v2psbars$ and $x_4 = -max\{v2x\_ex\_hereditary,$ v2x_ex_military, $v2x\_ex\_party\}$ and $\bar{x}_j$ and $sd(x_j)$ are the mean and standard deviation of x_j across all nation-year observations.

For ease of comparison to other measures of government type (which we normalize to fall between 0 and 1), we normalize our measure on coalition size:

$$W_{i,t} = \frac{w_{i,t} - min(w_{i,t})}{max(w_{i,t}) - min(w_{i,t})}$$ (2)

---

[1]Even these governments have held rigged elections although not for their national leader in the cases of Saudi Arabia and North Korea.
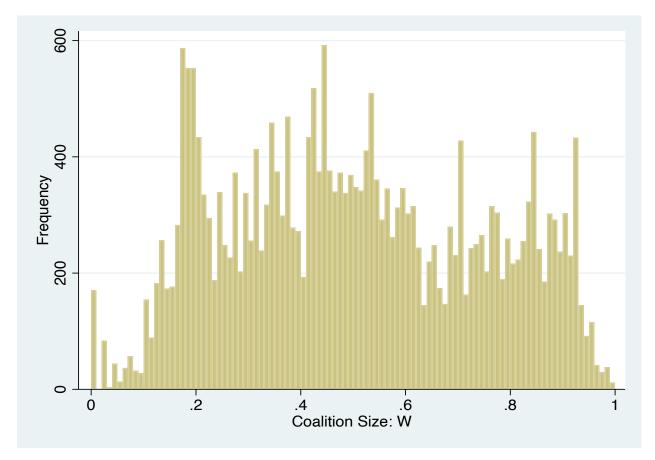
Figure 1: Distribution of Coalition Size, 1800-2018

The normalization of $W_{i,t}$ (we suppress the indexing moving forward) insures that the coalition size indicator falls between 0 and 1. Whereas the original indicator of coalition size, now referred to as $W(old)$, included only 5 values, we observe thousands of unique values for W in the data spanning the years from 1789 to 2018. Figure 1 displays the distribution of W between 1800 and 2018, divided into units of width 0.01. As the figure makes clear, W was widely distributed over the more than two-centuries that are plotted and it is essentially continuous although grouped into 100 blocs for ease of presentation.

The new indicator of W is designed to evaluate the extent to which institutions shape the relative size of any government's winning coalition by assessing how unconstrained (autonomous) or restricted institutions are that address opportunities to compete with the incumbent regime and its leadership selection. The theory predicts the proportionate allo-

9

cation of the budget between public and private goods based on coalition size for any size budget. This aspect of the selectorate theory is readily tested with W. Our data are available at .XXXXXXXXXXXXXXXXXXXX

3 Alternative Regime-Type Indicators

Numerous efforts have been made to estimate regime types with each effort tailored to the specific needs of the research to which the indicator was applied. Such efforts have generally focused, in one way or another, on whether particular governments were democratic or not or on the degree of democracy or authoritarianism manifested by different governments (Boix, Miller and Rosato, 2013; Bollen and Jackman, 1989; Collier and Adcock, 1999; Coppedge et al., 2011; Elkins, 2000; Gleditsch and Ward, 1997; Marshall, Gurr and Jaggers, 2016; Przeworski and Limongi, 1997; Teorell et al., 2019). Selectorate theory's predictions sometimes overlap with other approaches and sometimes diverge. Here the focus is on the key selectorate prediction that even small variations in coalition size influence the division of resource allocations between public goods that benefit everyone and private goods that specifically benefit coalition members. It is important to emphasize that we are making no claims about W being a better institutional measure in some global sense. Rather we argue that W is the appropriate measure for assessing the central selectorate proposition concerning the focus between public and private goods. That is what is testeed here.

A common characteristic of regime-type measurements is to define a cut point above which governments are judged to be democratic and below which they are determined to be non-democratic (Przeworski and Limongi, 1997; Boix, Miller and Rosato, 2013) or in other cases, below which they are said to be autocratic (Marshall, Gurr and Jaggers, 2016). These assessments are often determined by certain behavioral characteristics of governments.

The Polity 21-point scale, for instance, is based on an evaluation of each state's "elections for competitiveness and openness, the nature of political participation in general, and

the extent of checks on executive authority."[2] Some of these characteristics, such as some checks on executive authority, are institutionalized constraints while other characteristics of the Polity democracy-autocracy score, such as political participation and the evaluation of elections, measure behavioral or policy outcomes rather than institutional constraints. The use of behavioral outcomes, such as the freeness or competitiveness of elections or the defeat of an incumbent government (Przeworski and Limongi, 1997), serve the research purposes for which such indicators were developed but make them conceptually problematic from a selectorate perspective. The theoretical structure of the selectorate approach draws attention to how institutions incentivize leaders to emphasize public goods – like freeness and fairness, whether in elections or in other aspects of life– or to emphasize rigged systems that assure access to private benefits for the incumbent's winning coalition. From a selectorate perspective, the behavioral outcomes that are components of many regime-type indicators are things to be explained rather than to be used as measures of governance structures.

The selectorate perspective is not concerned with labeling or categorizing some governments as democratic and others as not democratic. In its two-dimensional institutional space – selectorate size and coalition size – government "types" do not fall into discrete categories. Monarchies and military juntas tend to have relatively small coalitions and small selectorates; rigged-election autocracies also have relatively small coalitions but relatively large selectorates; democracies have relatively large coalitions and large selectorates. While nominally democratic governments tend to have large winning coalitions, the size of their coalitions can vary markedly depending upon such factors as whether the system operates under proportional representation or majoritarian rules.

Rather than treating governance as categorical, such as "democratic" or "autocratic" (Przeworski and Limongi, 1997; Boix, Miller and Rosato, 2013), or as a mix of behavioral, attitudinal and institutional characteristics (Coppedge et al., 2011; Marshall, Gurr and Jaggers, 2016; Teorell et al., 2019), the selectorate approach focuses attention on governance

---

[2]See https://www.systemicpeace.org/csprandd.html

only in terms of its conceptually continuous institutional dimensions. Those dimensions serve as a positive way to explain a wide array of actions and behavioral outcomes that encompass domestic and foreign policy choices. Thus, it seeks to explain what other approaches treat as defining the normative and behavioral characteristics of governance. Still, if selectorate logic is correct and if, indeed, "democracies" tend to have larger winning coalitions than non-democratic regimes, standard regime-type indicators that divide governments into democracies and other types should fare well in the empirical tests we perform. However, because selectorate theory indicates that even small differences in relative coalition size matter, if selectorate logic is borne out, W should better explain the set of public and private goods allocations we investigate than do other regime-type estimates.

In addition to our new measure for W, we examine BDM2S2 old measure of coalition size ($W(old)$) based on the Polity measures of Competitiveness of Executive Recruitment, Openness of Executive Recruitment and Competitiveness of Participation and the Banks and Wilson (2019) measure of whether the government was a military regime. V-Dem contains a measure of *Support* for the government (v2regsupgroupssize), that assesses "how large is the percentage share of the domestic adult (18+) population that belongs to the political regime's supporting groups?". While at first glance this measure might appear to capture the concept of winning coalition, it is better thought of as an upper bound for W, a concept referred to in selectorate theory as the support coalition (Morrow et al., 2008). In particular, *Support* measures the proportion of the population in groups that support the government and does not account for the proportion of the people within these group that are critical for government survival.

Thus far, our discussion has focused on measures of democracy. However, recently scholars have generated measures of autocratic governance. We focus on two particularly good examples, Geddes, Wright and Frantz (2014); Geddes et al. (2018) (hereafter GWF) and Teorell and Lindberg (2019) (hereafter TL). In additional to a democracy indicator, GWF code whether a regime is a monarchy, based around the military, based on a single party or

is a personalistic regime. Likewise TL classifies regimes on five dimensions: direct election, presence of a confidence vote to remove leader, hereditary succession, military regime and single party regime. We use the latter three dimensions to form one of the components in our W measure. When we include GWF and TL in our comparative analyses, we include all five indicators, meaning that the resultant regression analyses are not quite comparable to those conducted to test W or any of the other single dimensional measures.

We assess the ability of the new W measure to account for the public/private focus of government policy compared to a host of dichotomous or graded indicators. In particular our comparisons include $W(old)$; *Support*, the 21-point Polity2 scale; its commonly-used dichotomous variant in which scores of 6 or higher indicate a democratic government and scores below 6 indicate a non-democratic regime (hereafter Dem6); the dichotomous indicator of democracy developed by Boix et al (hereafter Boix) (Boix, Miller and Rosato, 2013); the dichotomous indicator developed by Przeworski et al (hereafter Przeworski) (Alvarez et al., 1996); the set of five dichotomous variables – democracy, one-party, military, monarchy, and personalist – used by Geddes et al to categorize regime types (Geddes, Wright and Frantz, 2014; Geddes et al., 2018); and the five dimensions of governance proposed by Teorell and Lindberg (2019). Further when focusing on autocratic systems we differentiate between anocracies and autocracies, a common distinction made in the Polity scale (Vreeland, 2008).

One consequential indicator, the polyarchy index (Dahl, 1971; Coppedge et al., 2011), is excluded from our analysis. The polyarchy index, though closest conceptually to our notion of winning coalition size, conflates institutional drivers of policy choices and variations in behavioral outcomes under different governments. As a consequence, many of its components are the outputs of government policy, the very things that selectorate theory tries to account for.

Of course, at the end of the day, which institutional measure is most appropriate is governed by the questions being asked. Our analyses focus on comparing how W performs relative to other prominent measures of institutions at accounting for public and private

goods policy provisions.

Before turning to the testing procedures, we should establish some key statistical properties of the regime-type indicators we use. Table 1 displays summary statistics for each of the regime-type indicators. The new indicator for coalition size, W, has a correlation of 0.474 with $Support$, 0.606 with $W(old)$, 0.829 with the polity indicator; 0.737 with the dichotomous variant of polity; 0.761 with $Boix$ and 0.799 with $Przeworski$. The correlations between W and democracy, one-party, military, monarchy and personalist measures of GWF are 0.803, -0.432, -0.191, -0.286 and -0.216, respectively. Likewise the correlations between W and the direct election, confidence votes, heredity, military and single party dimensions of TL are 0.194, .557, -0.477, -0.263 and -0.096, respectively. Clearly the indicators overlap but have sufficient differences in how they vary that comparative tests of how well they do at explaining the dependent variables we investigate should be informative about their relative merits at least as means to test selectorate theory and, perhaps, to resolve the debate between an institutional and a behavioral approach to regime-type.

Table 1: Summary Statistics:

Variable	Observations	Mean	Median	Std. Dev.	Years
W	26360	0.50	0.48	0.24	1789 - 2020
$Support$	25266	-0.04	-0.00	1.49	1789 - 2020
W (old)	28026	0.50	0.50	0.27	1789 - 2020
Polity2	17325	0.48	0.35	0.35	1800 - 2018
Democracy6	17325	0.30	0.00	0.46	1800 - 2018
Boix	17124	0.32	0.00	0.47	1800 - 2015
Przeworski	9115	0.44	0.00	0.50	1946 - 2008
GWF: One Party	7954	0.28	0.00	0.45	1946 - 2010
GWF: Military	7954	0.07	0.00	0.26	1946 - 2010
GWF: Monarchy	7954	0.07	0.00	0.26	1946 - 2010
GWF: Personalist	7954	0.14	0.00	0.35	1946 - 2010
GWF: Democracy	7954	0.40	0.00	0.49	1946 - 2010
TL: Direct Election	25336	0.20	0.00	0.40	1789 - 2020
TL: Confidence Vote	25972	0.26	0.00	0.35	1789 - 2020
TL: Hereditary	26185	0.13	0.00	0.25	1789 - 2020
TL: Military	26185	0.14	0.00	0.22	1789 - 2020
TL: Party	26185	0.09	0.00	0.18	1789 - 2020

4 Testing Regime Indicators and W

Our first analysis examines the relationship between Public Goods provision (V-Dem's v2peapsol: "access to basic public services, such as order and security, primary education, clean water, and healthcare, distributed equally across political groups") and six different institutional measure using nation and year fixed effects regression on a common sample of 9,892 nation-years. The analyses include a measure of wealth (Ln(GDP per capita), V-Dem's e_migdppcln) and population size (Ln(population), V-Dem's e_mipopula).[3] As is clearly seen by the highly statistically significant coefficient estimates, each institutional measure is strongly associated with Public Goods. To determine whether W or an alternative institutional measure better accounts for policies and outcomes we rely on two non-nested tests: Akaike information criteria (AIC) and the Vuong test (Akaike, 1974; Vuong, 1989). Both tests are grounded in information theory and are discussed in detail in the appendix.

The AIC test provides a means of ranking alternative non-nested model specifications in terms of goodness of fit. To avoid over-fitting a model, the log-likelihood is discounted according to the number of model parameters included. The preferred model is the one with the smallest AIC statistic. Given our focus on W, we report ΔAIC as the difference between the AIC for W and the AIC for each of the alternative measures, and following standard practices we regard one model as superior if its AIC is at least 10 less than the AIC's for the competing models (Burnham and Anderson, 1998, p. 128). A limitation with the AIC approach is that it requires a common sample to compare across models. In contrast the Vuong test provides a pairwise test which we use to compare each institutional measure against W (Specifically, we use the cluster robust Vuong test that Woolridge (2010, section 13.11.2) recommends for panel data).

Table 2 reports results both for the AIC and Vuong tests for the general public goods dependent variable. On the basis of the AIC criteria, the W specification appears superior

---

[3]We report standard errors corrected for the nation level clustering, although the model comparison statistics reported are not affected by the form of standard error corrections.

to specifications with alternative measures. In terms of the Vuong test, the Vuong statistics suggest that W is the preferred model in comparisons to $W(old)$ and Polity2. However the Vuong statistics for the comparisons to *Support*, Dem6 and Boix fail to reject the null hypothesis that the models have the same likelihood at the 5% level. Of course showing that the new measure of W better accounts for the distribution of a single public goods measure is by itself hardly compelling. We examine a series of 30 measures of government policy that for presentational convenience we have broken into five categories: Key Pubic Goods, Key Private Goods, Fundamental Freedoms, Abuses and Health and Education.

Table 2: Public Goods Provisions and Institutional Indicators: Dependent Variable is Public Goods (`V-DEM:v2peapspol`)

	(1)	(2)	(3)	(4)	(5)	(6)
	W	Support	W (old)	Polity2	Dem6	Boix
Institution Measure	0.250***	0.306***	0.0926***	0.127***	0.0838***	0.0868***
	(8.45)	(7.18)	(5.90)	(7.91)	(7.99)	(7.37)
Ln(GDPpc)	0.0147	0.0174$^+$	0.0142	0.0207*	0.0202*	0.0179$^+$
	(1.56)	(1.81)	(1.49)	(2.18)	(2.14)	(1.90)
Ln(population)	-0.0212	-0.0288$^+$	-0.0106	-0.0196	-0.0213	-0.0260$^+$
	(-1.40)	(-1.80)	(-0.67)	(-1.24)	(-1.33)	(-1.77)
Constant	0.560***	0.538***	0.565***	0.574***	0.631***	0.689***
	(3.71)	(3.56)	(3.76)	(3.89)	(4.22)	(4.95)
Observations	9892	9892	9892	9892	9892	9892
FEs	270	270	270	270	270	270
Loglike.	14217.856	13887.482	13376.026	13817.614	13785.271	13863.948
Δ AIC	0.000	660.748	1683.660	800.485	865.171	707.816
Vuong	0.000	1.008	3.534	2.155	1.642	1.691
Pr(Vuong)	0.000	0.315	0.001	0.033	0.103	0.093

t statistics in parentheses
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

It would be too repetitious to fully describe each of the 30 dependent variables (see appendix) and show the full regression results for each dependent variable. Instead Table 3 reports tests comparing the performance of W to equivalently normalized versions of $W(old)$, Polity2, Dem6 and Boix on a common sample for seven key public goods. Each cell in table 3 reports four statistics: 1) estimates of the coefficient associated with the institutional measure, $\hat{\beta}$, 2) t-statistic as to whether the coefficient estimate is significantly different from zero, $\hat{\beta}/\hat{se}$, 3) Akaike information criteria (AIC) test, and 4) Vuong test. These two non-

nested test statistics are as described for Table 2. The analogous results for the remaining 23 dependent variables in the other four categories are reported in the appendix as tables A4, A5, A6 and A7.

Table 3: Model Comparisons: Key Public Goods

Variable	Stats	W	Support	W (old)	Polity2	Dem6	Boix
Public Goods	$\hat{\beta}$	0.250	0.306	0.093	0.127	0.084	0.087
obs = 9892	$t = \hat{\beta}/\hat{se}$	8.445	7.178	5.901	7.907	7.995	7.366
v2peapspol	Δ AIC	0.00	660.75	1683.66	800.48	865.17	707.82
	Vuong	0	1.008	3.534	2.155	1.642	1.691
Public Goods Ratio	$\hat{\beta}$	0.124	0.216	0.046	0.065	0.051	0.051
obs = 9892	$t = \hat{\beta}/\hat{se}$	4.431	5.945	3.903	4.500	6.131	5.089
ratio of v2peapspol and v2excrptps	Δ AIC	0.00	-493.20	659.65	263.20	22.65	2.19
	Vuong	0	-0.892	1.950	1.119	0.066	0.008
PublicGoods or clientelism	$\hat{\beta}$	0.719	2.055	0.464	0.346	0.359	0.315
obs = 11446	$t = \hat{\beta}/\hat{se}$	2.214	5.850	3.152	2.276	3.578	3.244
v2dlencmps	Δ AIC	0.00	-808.30	-1.89	73.88	-109.85	-38.53
	Vuong	0	-2.479	-0.011	0.625	-0.696	-0.279
TransparentLaws	$\hat{\beta}$	3.182	2.565	1.381	1.604	0.910	0.992
obs = 11443	$t = \hat{\beta}/\hat{se}$	17.345	7.498	10.328	12.477	10.599	12.558
v2cltrnslw	Δ AIC	0.00	3398.78	3024.51	1720.61	2805.65	2290.89
	Vuong	0	4.644	5.329	3.359	5.013	4.341
Elites Want Common Good	$\hat{\beta}$	1.773	3.062	0.815	0.971	0.585	0.641
obs = 9976	$t = \hat{\beta}/\hat{se}$	4.945	8.449	4.836	5.037	5.338	5.540
v2dlcommon	Δ AIC	0.00	-633.91	724.07	291.85	511.90	321.92
	Vuong	0	-1.235	1.833	1.062	1.340	0.998
Public Want Common Good	$\hat{\beta}$	4.165	3.117	1.568	2.205	1.070	1.219
obs = 11448	$t = \hat{\beta}/\hat{se}$	21.076	9.437	10.411	15.580	10.430	12.500
v2dlengage	Δ AIC	0.00	4986.33	4880.38	1869.77	4509.92	3690.40
	Vuong	0	5.256	6.396	3.741	6.715	6.367
FreeExpression	$\hat{\beta}$	1.150	0.681	0.437	0.615	0.312	0.329
obs = 11424	$t = \hat{\beta}/\hat{se}$	28.358	6.702	11.297	16.589	12.767	13.149
v2x_freexp	Δ AIC	0.00	10002.75	8694.70	3518.86	7713.26	7134.38
	Vuong	0	6.215	7.105	3.987	6.500	7.205

In order to summarize the relative explanatory power of the institutional measures across the large number of dependent variables, we produce a simple score card based upon the Vuong statistic. The rows of Table 4 assess how W performed relative to the other institutional measures across the 30 dependent variables (shown in tables 3 A4, A5, A6 and A7). For instance, the table reports that in pairwise Vuong tests in 16 out of 30 comparisons W was preferred to Support at the $p < .01$ level (entry in the top left of table). Twice W outperformed support at lower level of statistical significance. 11 times the Vuong test failed to distinguish between W and Support and in one case Support out performed W ($.1 < p < .01$). None of the other models in the table outperforms W on any of the dependent variables, while W provides

17

Table 4: Scorecard: Common Sample for All Tests (1800-2008)

Score	Support	W (old)	Polity2	Dem6	Boix
Favors W ($p < .01$)	16	18	13	15	14
Favors W ($0.1 \geq p > .01$)	2	4	5	4	5
Indeterminate	11	8	12	11	11
Opposes W ($0.1 \geq p > .01$)	1	0	0	0	0
Opposes W ($p < .01$)	0	0	0	0	0

a better fit at least at p < 0.10 and generally at p < 0.01 between 18 and 22 times out of 30 opportunities. The score card makes evident that overall W outperforms each of the other regime-type indicators in this set of tests. Looked at from the AIC perspective, W ranks significantly ahead of all of the other models in 20 out of 30 tests.

4.1   Tests on an Expanded Set of Regime-Type Measures

The Przeworski measure of democracy was excluded from Table 4 because it is available over a smaller number of year and so its inclusion would have significantly reduced the common sample size. To include this measure in the comparison we adopt an alternative pairwise approach. For instance, the analyses reported in the Przeworski column of Table 5 are based on all observations for which both W and Przeworski measures are available. This score card also includes GWF and TL measures, each of which contain 5 indicators of regime type [4].

Table 5: Scorecard: Pairwise Samples

Score	Support	W (old)	Polity2	Dem6	Boix	Przeworski	GWF	TL
Favors W ($p < .01$)	15	19	12	15	14	14	9	13
Favors W ($0.1 \geq p > .01$)	1	2	6	4	4	7	3	3
Indeterminate	11	9	12	11	12	9	13	12
Opposes W ($0.1 \geq p > .01$)	1	0	0	0	0	0	5	2
Opposes W ($p < .01$)	2	0	0	0	0	0	0	0

Table 5, in which samples vary by pairs of regime-type indicators, tells much the same story as the previous score card. The Vuong-test results demonstrate that of the 8 alternative indicators to W, *Support* outperforms W on three dependent variables, GWF outperforms W on five dependent variable and TL outperforms W in two dependent variables. W, in contrast, fits the dependent variables significantly better than other measures between 12 and 21 times out of 30 dependent variables. W performs least well when looking at the Health and Education variables and when looking at the Private or Public Goods dependent

---

[4]Tables A8, A9, A10, A11 and A12 summarize the results for each individual dependent variable

variable in the Key Public Goods section. W performed particularly well compared to the other measures in analyses looking at Fundamental Freedom and Abuse. Because these tests, unlike the earlier tests, do not include the same sample across models the AIC test cannot be utilized. Hence, we replicated the tests reported earlier, but now including all of these models for the set of observations across which they are all observed. Doing so restricts the sample sizes to between 1,441 and 5,602. AIC for this more limited sample shows W ranks clearly first (ΔAIC > 10 and usually > 100) for 57% of the dependent variables.

4.2 Tests Within Regime Types

One of the critiques made by Gallagher and Hanson (2015) and Kennedy (2009) is that while the selectorate model was good at differentiating between democracies and non-democracies, it had limited power to discriminate between institutions within these nominal categories. To address this concern, we replicate the pairwise tests looking separately at scores across democracies (as defined by Polity2$\geq$ 6, Table 6) and at scores among non-democracies (as defined by Polity2< 6, Table 7). Within the non-democratic comparisons, the column Anoc. differentiates between anocracies ($6 > Polity2 > -6$) and autocracies ($Polity2 \leq -6$)[5] .

Table 6: Scorecard: Democracies (Polity2 $\geq$ 6)

Score	Support	W (old)	Polity2	Boix	Przeworski	GWF	TL
Favors W ($p < .01$)	4	9	10	10	4	0	5
Favors W ($0.1 \geq p > .01$)	13	13	12	13	16	16	8
Indeterminate	13	8	8	7	10	14	17
Opposes W ($0.1 \geq p > .01$)	0	0	0	0	0	0	0
Opposes W ($p < .01$)	0	0	0	0	0	0	0

Table 6 , focused on governments categorized as democratic by Polity, tells the same story as the previous two score cards. No alternative indicator statistically outperforms W whereas W frequently does a better job of sorting out values on the dependent variables than do the alternative indicators. In the pairwise comparisons, the Vuong tests significantly favor W over the alternative measures by a margin ranging from 13 of 30 dependent variables for five dimensions of TL to 23 of 30 dependent variables for the Boix measure.

The final score card, Table 7 , summarizes tests restricted to the sample of non-democratic nations as determined by Polity. In a nutshell, the score card shows us two critical pieces of information. W does better than each of the other regime-type indicators although it does not perform as strongly in this subset of cases as it did in the full set. Based on the pairwise samples, *Support* outperforms W for three dependent variables. However on the other side of the ledger, W is statistically the superior measure to Support 18

---

[5]Tables A13, A14, A15, A16, A17, A18, A19, A20, A21, A22 summarize the results for each set of dependent variables.

times. In comparison to the other measures, W is the preferred model between 10 and 19 times out of the 30 dependent variables. GWF and TL are significangtly better than W in 1 out of 30 dependent variables each. The findings are similar when we use the AIC criteria on a (much smaller) common sample. W is the preferred measure in accounting for 14 of the 30 dependent variables. No other regime indicator does as well.

Contrary to Gallagher and Hanson (2015) and Kennedy (2009) critiques that W is only useful for distinguishing between democracies and autocracies, our results suggest that, over a broad range of public and private goods, W is the superior measure both within these categories as well as across these categories. This result holds both in terms of Vuong tests across pairwise samples (that maximize the number of observations) and AIC criteria that simultaneous compare all measures on a (smaller) common sample.

Table 7: Scorecard: Autocracies (Polity2 < 6, Pairwise Samples)

Score	Support	W (old)	Polity2	Anoc.	Boix	Przeworski	GWF	TL
Favors W ($p < .01$)	12	15	6	13	14	7	6	9
Favors W ($0.1 \geq p > .01$)	6	4	5	3	5	8	4	2
Indeterminate	9	11	19	14	11	15	19	18
Opposes W ($0.1 \geq p > .01$)	3	0	0	0	0	0	1	1
Opposes W ($p < .01$)	0	0	0	0	0	0	0	0

4.3   Caveats

The proposed new measure of winning coalition size does a good job at accounting for the provision of public and private goods, a highly important aspect of governance evaluated by selectorate theory. However, we note several important caveats. First, there is no claim that W is superior to other measures in general. The W measure accounts for institutional features that determine the size of the group to whom leaders are beholden. The alternative measures reflect different theoretical perspectives and we anticipate that, in the context for which they were envisioned, they perform well.

Second, the dependent variables we examine, as well as the components of W, are all from V-Dem and so biases in coding W might correlate with biases in the coding of other variables. While this concern suggests caution, there are important factors that mitigate this risk. Much of V-Dem's data are generated by expert assessments. All of the analyses here include country fixed effects. If coding across variables is biased by the use of common panels of country experts to make estimates within each country then any within-country bias is likely to be absorbed by the fixed effects. Additionally, it is important to note that the way in which we have combined V-Dem indicators to construct W is unique and does not reflect any one existing indicator in the V-Dem dataset. In fact, the bivariate correlations between the several components of W and the

indicator of W do not look much different than the correlations between W and Polity or Przeworski or Boix or Support or other of the alternative regime-type measures.

We used V-Dem's variables because V-Dem provides a much longer time series that most other data sources. Examining alternative data we find similar patterns although our ability to discriminate between measures is hampered by reduced sample size. Finally, V-Dem's IRT methodology estimates the distribution of underlying latent concepts from the ordered categorical responses of expert coders. The W measure was constructed using point estimates rather than utilizing all the information about the distribution of measures. In the appendix we show replications of the analysis of Public Goods provision shown in Table 2 by repeated sampling from the distribution of the components used to create the W measure. The distribution of the AIC statistics reinforce the conclusions drawn from the measure constructed from point estimates.

5 Conclusions

We introduced a new, continuous, institutions-based indicator of relative coalition size and estimated it on a country-year basis for virtually every country each year between 1789 and 2018. Our ability to do so is due in large part to the V-Dem project's comprehensive collection of data. The breadth and scale of that project, in terms of both the number of measures and the number of years for which data are available, offer scholars vital tools for measuring and testing theory.

In head-to-head comparisons with commonly used regime-type indicators we found that W provided a better fit with the 30 dependent variables that we examined than did any of the alternatives. In terms of accounting for public and private goods provisions the W measure appears to be the best way to model the dependent variables among the set of common regime-type indicators considered here. The strength of the fit and the comparatively better performance shown by W lend added support to the theoretical account of governmental resource allocations derived from the selectorate theory. These findings go some ways toward addressing critiques (Gallagher and Hanson, 2015; Kennedy, 2009; Clarke and Stone, 2008) that selectorate theory lacks adequate empirical measures of its core concepts.

The results seem to answer the call for a better way to estimate coalition size so as to implement the insights of selectorate theory (Gallagher and Hanson, 2015; Kennedy, 2009; Clarke and Stone, 2008) and to allay concerns that the theory does not discriminate within non-democratic regimes in its ability to explain public and private goods choices. The evidence also supports the selectorate theory contention, as highlighted by Clarke and Stone, that institutions, rather than attitudes and behaviors, provide a stronger foundation for distinguishing government resource allocations within governmental categories and, indeed, across all types of governments.

While the results seem to answer the main empirical criticisms of selectorate theory, much remains to be done. The thirty dependent variables we have analyzed encompass a wide variety of governmental policy choices but there are many more to be tested with the new indicator, W, as well as with $\frac{W}{S}$. Space precludes us from doing so here but future research will expand on the tests reported here to encompass issues of governmental stability, endogenous institution change, foreign policy, and more. For now, mindful of Hempel's concerns about the shift from theory to observable quantities, we believe some useful progress has been made in further explaining and exploring differences in performance across regimes.

References

Akaike, Hirotugu. 1974. "A new look at the statistical model identification." *IEEE transactions on automatic control* 19(6):716–723.

Alvarez, Mike, José Antonio Cheibub, Fernando Limongi and Adam Przeworski. 1996. "Classifying political regimes." *Studies in Comparative International Development* 31(2):3–36.

Banks, Arthur and Kenneth A Wilson. 2019. *Cross-national time-series data archive.* Jerusalem, Israel: DATABANKS INTERNATIONAL.

Boix, Carles, Michael Miller and Sebastian Rosato. 2013. "A complete data set of political regimes, 1800–2007." *Comparative Political Studies* 46(12):1523–1554.

Bollen, Kenneth A and Robert W Jackman. 1989. "Democracy, stability, and dichotomies." *American Sociological Review* pp. 612–621.

Bueno de Mesquita, Bruce, Alastair Smith, Randolph Siverson and James Morrow. 2003. *The Logic of Political Survival.* Cambridge, MA: The MIT Press.

Burnham, Kenneth P and David R Anderson. 1998. *Model Selection and Inference: A practical information-theoretic approach.* Springer, New York.

Clarke, Kevin A and Randall W Stone. 2008. "Democracy and the logic of political survival." *American Political Science Review* pp. 387–392.

Collier, David and Robert Adcock. 1999. "Democracy and dichotomies: A pragmatic approach to choices about concepts." *Annual review of political science* 2(1):537–565.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I Lindberg, Jan Teorell, David Altman, Michael Bernhard, Agnes Cornell, M Steven Fish, Lisa Gastaldi et al. 2021. "V-Dem Codebook v11.".

Coppedge, Michael, John Gerring, David Altman, Michael Bernhard, Steven Fish, Allen Hicken, Matthew Kroenig, Staffan I Lindberg, Kelly McMann, Pamela Paxton et al. 2011. "Conceptualizing and measuring democracy: A new approach." *Perspectives on Politics* pp. 247–267.

Dahl, Robert A. 1971. *Polyarchy: Participation and opposition.* New Haven, CT: Yale university press.

Elkins, Zachary. 2000. "Gradations of democracy? Empirical tests of alternative conceptualizations." *American Journal of Political Science* pp. 293–300.

Gallagher, Mary E and Jonathan K Hanson. 2015. "Power tool or dull blade? Selectorate theory for autocracies." *Annual Review of Political Science* 18.

Geddes, Barbara, Joseph George Wright, Joseph Wright and Erica Frantz. 2018. *How dictatorships work: Power, personalization, and collapse.* Cambridge University Press.

Geddes, Barbara, Joseph Wright and Erica Frantz. 2014. "Autocratic breakdown and regime transitions: A new data set." *Perspectives on Politics* 12(2):313–331.

Gleditsch, Kristian S and Michael D Ward. 1997. "Double take: A reexamination of democracy and autocracy in modern polities." *Journal of Conflict Resolution* 41(3):361–383.

Kennedy, Ryan. 2009. "Survival and accountability: An analysis of the empirical support for electorate theory?" *International Studies Quarterly* 53(3):695–714.

Levitsky, Steven and Daniel Ziblatt. 2018. *How Democracies Die.* Crown Books.

Marquardt, Kyle L and Daniel Pemstein. 2017. "IRT models for expert-coded panel data." *V-Dem Working Paper* 41.

Marshall, Monty G, Ted Robert Gurr and Keith Jaggers. 2016. "Polity IV project: Political regime characteristics and transitions, 1800-2015." *Center for Systemic Peace* .

Morrow, James D, Bruce Bueno De Mesquita, Randolph M Siverson and Alastair Smith. 2008. "Retesting selectorate theory: Separating the effects of W from other elements of democracy." *American Political Science Review* pp. 393–400.

Pemstein, Daniel, Kyle L Marquardt, Eitan Tzelgov, Yi-ting Wang, Joshua Krusell and Farhad Miri. 2019. "The V-Dem measurement model: latent variable analysis for cross-national and cross-temporal expert-coded data?" *V-Dem Working Paper* 21.

Przeworski, Adam and Fernando Limongi. 1997. "Modernization: Theories and facts." *World politics* 49(2):155–183.

Teorell, Jan, Michael Coppedge, Staffan Lindberg and Svend-Erik Skaaning. 2019. "Measuring polyarchy across the globe, 1900–2017." *Studies in Comparative International Development* 54(1):71–95.

Teorell, Jan and Staffan I Lindberg. 2019. "Beyond democracy-dictatorship measures: a new framework capturing executive bases of power, 1789–2016." *Perspectives on Politics* 17(1):66–84.

Vreeland, James Raymond. 2008. "The Effect of Political Regime on Civil War: Unpacking Anocracy." *Journal of Conflict Resolution* 52(3):401–425.
URL: *https://doi.org/10.1177/0022002708315594*

Vuong, Quang H. 1989. "Likelihood ratio tests for model selection and non-nested hypotheses." *Econometrica: Journal of the Econometric Society* pp. 307–333.

Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data*. MIT press.

Appendix:

A New Indicator of Coalition Size: Tests Against Standard Regime-Type Indicators

Bruce Bueno de Mesquita[1] and Alastair Smith[2]

Wilf Family Department of Politics

New York University

19 West 4th St, New York NY 10012

[1]Bruce.buenodemesquita@nyu.edu

[2]Alastair.smith@nyu.edu

January 2022

Contents

1 Definitions and Coding Rules

The winning coalition size is constructed using variables from V-Dem data set, version 10 (2020). The definitions of the relevant variables below are taken directly from the V-Dem codebook.

1.1 EMB autonomy (v2elembaut)

Question: Does the Election Management Body (EMB) have autonomy from government to apply election laws and administrative rules impartially in national elections? Clarification: The EMB refers to whatever body (or bodies) is charged with administering national elections (Pemstein et al., 2019).

Table A1: Coding of v2elembaut

Response	
0	No. The EMB is controlled by the incumbent government, the military, or other de facto ruling body.
1	Somewhat. The EMB has some autonomy on some issues but on critical issues that influence the outcome of elections, the EMB is partial to the de facto ruling body.
2	Ambiguous. The EMB has some autonomy but is also partial, and it is unclear to what extent this influences the outcome of the election.
3	Almost. The EMB has autonomy and acts impartially almost all the time. It may be influenced by the de facto ruling body in some minor ways that do not influence the outcome of elections.
4	Yes. The EMB is autonomous and impartially applies elections laws and administrative rules.

1.2 Opposition parties autonomy (v2psoppaut)

Question: Are opposition parties independent and autonomous of the ruling regime? Clarification: An opposition party is any party that is not part of the government, i.e., that has no control over the executive (Pemstein et al., 2019).

Table A2: Coding of v2psoppaut

Response	
0	Opposition parties are not allowed.
1	There are no autonomous, independent opposition parties. Opposition parties are either selected or co-opted by the ruling regime.
2	At least some opposition parties are autonomous and independent of the ruling regime.
3	Most significant opposition parties are autonomous and independent of the ruling regime.
4	All opposition parties are autonomous and independent of the ruling regime.

1.3 Barriers to parties (v2psbars)

Project Manager(s): Allen Hicken Question: How restrictive are the barriers to forming a party? Clarification: Barriers include legal requirements such as requirements for membership or financial deposits, as well as harassment (Pemstein et al.,

2019).

Table A3: Coding of v2psbars

Response	
0	Parties are not allowed.
1	It is impossible, or virtually impossible, for parties not affiliated with the government to form (legally).
2	There are significant obstacles (e.g. party leaders face high levels of regular political harassment by authorities).
3	There are modest barriers (e.g. party leaders face occasional political harassment by authorities).
4	There are no substantial barriers.

1.4 Hereditary dimension index (D) (v2x_ex_hereditary)

Question: To what extent is the power base of the chief executive determined by hereditary succession? Clarification: Representing one of five regime dimensions, each of which may be more or less present in any given case, this index taps into the extent to which the appointment and dismissal of the chief executive is based on hereditary rule (Teorell and Lindberg, 2019).

Aggregation: The index is based on whether the "chief executive" was (a) appointed through hereditary sucession or by a royal council, and (b) can be dismissed by a royal council. Both condition (a) and (b) are coded as present (1) or not (0); we then average across the two. In nominally dual systems, where the head of state (HOS) and the head of government (HOG) are not the same individual, we determine who is the "chief executive" by comparing HOS and HOG powers over the appointment and dismissal of cabinet ministers. We aggregate across the two executives by taking the average weighted by their relative powers over cabinet formation and dismissal.

1.5 Military dimension index (D) (v2x_ex_military)

Question: To what extent is the power base of the chief executive determined by the military? Clarification: Representing one of five regime dimensions, each of which may be more or less present in any given case, this index taps into the extent

to which the appointment and dismissal of the chief executive is based on the threat or actual use of military force (Teorell and Lindberg, 2019).

Aggregation: The index is based on whether the "chief executive" was (a) appointed through a coup, rebellion or by the military, and (b) can be dismissed by a the military. Both condition (a) and (b) are coded as present (1) or not (0); we then average across the two. In nominally dual systems, where the head of state (HOS) and the head of government (HOG) are not the same individual, we determine who is the "chief executive" by comparing HOS and HOG powers over the appointment and dismissal of cabinet ministers. We aggregate across the two executives by taking the average weighted by their relative powers over cabinet formation and dismissal.

1.6 Ruling party dimension index (D) (v2x_ex_party)

Question: To what extent is the power base of the chief executive determined by a ruling party? Clarification: Representing one of five regime dimensions, each of which may be more or less present in any given case, this index taps into the extent to which a ruling party appoints and dismisses the chief executive (Teorell and Lindberg, 2019).

Aggregation: The index is based on whether the "chief executive" was (a) appointed by the ruling party, and (b) can be dismissed by the ruling party. Both condition (a) and (b) are coded as present (1) or not (0); we then average across the two. In nominally dual systems, where the head of state (HOS) and the head of government (HOG) are not the same individual, we determine who is the "chief executive" by comparing HOS and HOG powers over the appointment and dismissal of cabinet ministers. We aggregate across the two executives by taking the average weighted by their relative powers over cabinet formation and dismissal.

2 Assessment of Government Policy Provision

We compare the explanatory power of different institutional indicators across 30 measures of policy outcomes, government performance, freedoms and welfare. For presentational convenience we organize and report on these dependent variables clustered into five groups: Key Pubic Goods, Key Private Goods, Fundamental Freedoms, Abuses and Health and Education.

2.1 Dependent Variables

The 30 measures of government performance are drawn from V-Dem data. We find similar results looking at output measures from other data sources such as the World Bank's World Development Indicators but focus on V-Dem's variables because of their much greater temporal coverage. Here we provide a brief description of each variables and its source.

2.1.1 Key Public Goods

1. Public Goods (v2peapspol), meaning "access to basic public services, such as order and security, primary education, clean water, and healthcare, distributed equally across political groups" (Coppedge et al., 2021, p. 204);

2. Goods Ratio, a constructed variable that normalizes a general measure of public goods provision (v2peapspol) to fall between 0 and 1 and divides that quantity by itself plus a normalized version of a general measure of public sector corruption (v2excrptps) with the direction of the corruption variable reversed so that larger values indicate more public sector corruption and lower values indicate less public sector corruption. This constructed variable is intended to evaluate the selectorate theory's expectation that as W increases, the proportion of revenue spent by government on public goods increases relative to the amount committed to private goods;

3. Public Goods or Clientelism (v2dlencmps); that is, the extent to which spending is on particularistic goods (defined by V-Dem as pork, clientelistic spending and private goods) or on public goods "intended to benefit all communities within a society" (Coppedge et al., 2021, p. 151);

4. Transparent Laws (v2cltrnslw); that is, transparent and predictable laws;

5. Elite Want Common Good (v2dlcommon), meaning the extent to which the elite act in pursuit of the common good;

6. Public Want Common Good (v2dlengage), the extent to which the public is engaged in advancing the common good; and

7. Free Expression (v2x_freexp).

The average bivariate correlation among the variables in the first table is 0.66 so they are clearly measuring different aspects of public goods provision. The second set of tables assess key private goods, with the V-Dem variables reversed in direction, as appropriate, so that higher values reflect greater provision of the private good and smaller values reflect a lesser provision – or none at all – of the private good. The mean correlation among the variables in this table is 0.55:

2.1.2 Key Private Goods

1. State Owns Economy (v2clstown), meaning the extent to which the state has centralized control over production and other economic activity;

2. Corruption (v2excrptps);

3. ExecutiveCorrupt (v2x_execorr), meaning that the executive allows corruption;

4. Public Sector Theft (v2exthftps);

5. Clientelism (v2xnp_client) meaning that public monies are use for political gain; and

6. Judge Corrupt (v2jucorrdc) indicating the extent to which judges are corrupt.

2.1.3 Fundamental Freedoms

The next table in each analysis focuses attention on the prevalence (or absence) of fundamental freedoms and has (an unsigned) average correlation of 0.66:

1. Respect for Constitution (v2exrescon), executive respect for the constitution;

2. Rule of Law (v2x_rule), the extent to which there is rule of law;

3. Free Elections (v2elfrfair), how free and fair elections are;

4. Free Movement (v2xcl_dmove), the extent to which people enjoy freedom of movement;

5. Civil Liberties (v2x_clpol);

6. Political Rights (reversed order for e_fh_pr so that higher values mean more rights);

7. Religious Freedom (v2clrelig);

8. Judicial Independence (v2juhcind,);

9. Property Rights (v2xcl_prpty).

2.1.4 Abuses

The fourth table summarizing the regression analysis results focuses our attention on abuses of the populace designed to protect private goods allocations. The average correlation among these dependent variables is 0.54:

1. Media Censored (v2mecenefm), that is,the frequency of government efforts to censor the media;

2. Packed Courts (v2jupack), meaning the executive packs or influences the courts;

3. Torture (v2cltort);

4. Slave Labor (v2xcl_slave), so that larger values indicate greater use of forced or enslaved labor.

2.1.5 Health and Education

The final set of regression assessments are for variables measuring Health and Education. The bivariate (unsigned) correlations among these dependent variables average 0.53:

1. Infant Mortality (e_peinfmor) for which, of course, higher values are worse;

2. Campus Freedom (v2casurv), meaning that campuses are free from government surveillance;

3. Academic Freedom (v2xca_academ); and

4. Health Care (v2pehealth), assessed as the extent to which there is equal access to quality health care across the population.

2.2 Tests to Compare Model Fit for Different Institutional Measures

For each dependent variable we run a fixed effects regression for each institutional measure. These analyses also include a measure of wealth (Ln(GDP per capita), V-Dem's e_migdppcln), population size (Ln(population), V-Dem's e_mipopula) and fixed effects for nations and years.[1]

To test whether W or an alternative institutional measure better accounts for policies and outcomes we rely on two non-nested tests: Akaike information criteria (AIC) and the Vuong test (Akaike, 1974; Vuong, 1989). Both tests are grounded in information theory. We discuss each in turn.

The AIC test provides a means of ranking alternative non-nested model specifications in terms of goodness of fit. To avoid over-fitting a model, the log-likelihood is discounted according to the number of model parameters included. The preferred model is the one with the smallest AIC statistic. Given our focus on W, we report ΔAIC as the difference between the AIC for W and the AIC for each of the alternative measures. A positive ΔAIC indicates selecting in favor of W, while a negative ΔAIC suggests the alternative measure is preferred.[2]

We apply extremely stringent criteria to evaluate whether the difference-scores reported by AIC are statistically meaningful. As Burnham and Anderson (1998, p. 128) report, "If a model has $4 < \Delta < 7$ there is definite evidence that the model is not the K-L best model, and if $7 < \Delta < 10$, there is strong evidence that the model is not the K-L best model. Finally, if $\Delta > 10$, there is very strong evidence that the model is not the K-L best model."[3] Burnham and Anderson (1998) assess a difference of 4 to 7 in AIC scores as equivalent to the probability of a significant difference in models equal to $p < 0.05$. We assess all of the models and identify which one – if any – ranks first in goodness-of-fit if all of the other models have AIC scores that are all at least 10 points larger than the model in question. Hence, if a model other than W fits best on a given dependent variable, it will have an ΔAIC score at least 10 points lower than any other model, including, of course, W. Conversely, if W fits best then all other models will have ΔAIC scores at least 10 points larger than W's. No significant difference is judged to have been found if no model's score is more than 10 points lower than the score for each other model.

---

[1] We report standard errors corrected for the nation level clustering, although the model comparison statistics reported are not affected by the form of standard error corrections.

[2] Alternative statistics, such as Bayesian Information Criteria and small sample corrections for AIC, provide different metrics to discount the number of parameters in a model. However, since all the specification contain the same number of parameters the alternative statistics have identical results (Burnham and Anderson, 1998).

[3] The Kullback-Leibler (K-L) information criterion assesses the extent to which one probability distribution differs from another, providing a mechanism for evaluating the comparative fit of alternative models.

The Vuong test, like the AIC score, relies on the logic of the difference in probability distributions across models based on the Kullback-Leibler (K-L) information criterion. The Vuong test, however, is intended to compare pairs of models to each other rather than larger sets of models as in the case of AIC (Vuong, 1989). The null hypothesis in the Vuong test is that each model produces the same likelihood. Under the null hypothesis the test statistics is t-distributed. We parameterize the test such that a positive Vuong statistic suggests W is preferred to the alternative institutional measure. If the magnitude of the t-statistic is greater than 1.96, then the null hypothesis that both models are equally good is rejected at the 5% level. In comparing models we report both the AIC test and the cluster robust Vuong test that Woolridge (2010, section 13.11.2) recommends for panel data.

3 Analyses of Government Policies for a Common Sample

In the main text, Table 3 reports tests comparing the performance of W to equivalently normalized versions of $W(old)$, Polity2, Dem6 and Boix on a common sample for seven key public goods. Each cell in table 3 reports four statistics: 1) estimates of the coefficient associated with the institutional measure, $\hat{\beta}$, 2) t-statistic as to whether the coefficient estimate is significantly different from zero, $\hat{\beta}/\hat{se}$, 3) Akaike information criteria (AIC) test, and 4) Vuong test. Tables 3, A4, A5, A6 and A7 contain fixed effects regressions examining Key Pubic Goods, Key Private Goods, Fundamental Freedoms, Abuses and Health and Education, respective. The analyses use a common sample for each dependent variable.

Table A4: Model Comparisons: Key Private Goods

Variable	Stats	W	Support	W (old)	Polity2	Dem6	Boix
State Owns Economy	$\hat{\beta}$	-2.407	0.024	-0.514	-1.104	-0.454	-0.576
obs = 11448	$t = \hat{\beta}/\hat{se}$	-4.968	0.064	-2.443	-4.619	-3.847	-4.179
v2clstown	Δ AIC	0.00	1920.82	1724.36	834.28	1529.29	1277.08
	Vuong	0	2.326	2.312	1.978	2.228	2.202
Corruption	$\hat{\beta}$	-0.063	-0.228	-0.042	-0.052	-0.052	-0.044
obs = 11446	$t = \hat{\beta}/\hat{se}$	-1.572	-3.761	-2.612	-2.156	-3.762	-2.928
v2excrptps	Δ AIC	0.00	-679.49	-7.92	-90.65	-326.28	-201.20
	Vuong	0	-1.421	-0.069	-0.633	-1.565	-1.205
Executive Corrupt	$\hat{\beta}$	-0.209	-0.258	-0.095	-0.103	-0.082	-0.073
obs = 11446	$t = \hat{\beta}/\hat{se}$	-3.197	-3.628	-3.414	-2.920	-4.445	-3.563
v2x_execorr	Δ AIC	0.00	59.02	261.49	166.87	31.97	124.33
	Vuong	0	0.173	1.051	0.947	0.138	0.652
Public Sector Theft	$\hat{\beta}$	-0.927	-1.421	-0.355	-0.427	-0.381	-0.327
obs = 11390	$t = \hat{\beta}/\hat{se}$	-3.814	-5.768	-2.738	-2.969	-4.413	-3.631
v2exthftps	Δ AIC	0.00	-174.78	298.36	189.91	-14.22	109.34
	Vuong	0	-0.615	1.363	1.212	-0.071	0.646
Clientelism	$\hat{\beta}$	-0.126	-0.237	-0.084	-0.068	-0.070	-0.061
obs = 11419	$t = \hat{\beta}/\hat{se}$	-1.816	-2.945	-3.479	-2.050	-3.326	-2.570
v2xnp_client	Δ AIC	0.00	-196.45	-13.98	34.30	-168.25	-84.79
	Vuong	0	-0.778	-0.080	0.289	-0.898	-0.539
Judge Corrupt	$\hat{\beta}$	-0.239	-0.661	-0.175	-0.083	-0.179	-0.141
obs = 11413	$t = \hat{\beta}/\hat{se}$	-0.816	-2.671	-1.813	-0.655	-2.527	-1.841
v2jucorrdc	Δ AIC	0.00	-166.35	-14.62	31.33	-117.17	-56.11
	Vuong	0	-1.122	-0.157	0.420	-1.268	-0.815

Table A5: Model Comparisons: Fundamental Freedoms

Variable	Stats	W	Support	W (old)	Polity2	Dem6	Boix
Respect for Constitution	$\hat{\beta}$	3.580	2.533	1.444	1.772	0.870	0.986
obs = 11396	$t = \hat{\beta}/\hat{se}$	10.819	6.538	8.740	10.277	8.396	9.348
v2exrescon	Δ AIC	0.00	3670.63	3190.31	1792.69	3344.22	2830.77
	Vuong	0	4.438	4.201	3.363	4.872	4.533
Rule of Law	$\hat{\beta}$	0.537	0.453	0.201	0.260	0.161	0.168
obs = 11448	$t = \hat{\beta}/\hat{se}$	12.952	6.033	7.522	8.584	8.583	9.705
v2x_rule	Δ AIC	0.00	2885.76	3209.09	1802.80	2268.21	2005.11
	Vuong	0	3.933	4.653	3.165	3.736	3.622
Free Elections	$\hat{\beta}$	4.936	2.717	1.357	2.403	1.168	1.259
obs = 3200	$t = \hat{\beta}/\hat{se}$	22.704	6.549	6.811	12.708	10.122	11.831
v2elfrfair	Δ AIC	0.00	1470.74	1460.17	450.81	1031.12	904.17
	Vuong	0	5.088	6.053	2.203	4.577	4.456
Free Movement	$\hat{\beta}$	0.562	0.413	0.174	0.264	0.148	0.160
obs = 11448	$t = \hat{\beta}/\hat{se}$	9.365	5.167	5.539	7.035	6.050	6.437
v2xcl_dmove	Δ AIC	0.00	2404.74	2657.28	1401.99	2040.35	1794.37
	Vuong	0	2.758	3.850	2.931	3.402	3.391
Civil Liberties	$\hat{\beta}$	1.232	0.717	0.448	0.642	0.318	0.340
obs = 11447	$t = \hat{\beta}/\hat{se}$	35.239	7.825	11.936	18.310	13.170	14.536
v2x_clpol	Δ AIC	0.00	13030.29	11761.11	5675.11	10790.72	9989.22
	Vuong	0	7.132	8.363	5.797	7.922	8.273
Political Rights	$\hat{\beta}$	6.983	5.929	4.181	4.559	2.366	2.742
obs = 5844	$t = \hat{\beta}/\hat{se}$	21.811	9.335	16.971	22.217	14.012	16.894
e_fh_pr	Δ AIC	0.00	3277.89	1445.71	-453.81	1582.84	486.56
	Vuong	0	6.138	3.218	-1.353	3.857	1.339
Religious Freedom	$\hat{\beta}$	3.126	1.518	0.924	1.540	0.805	0.846
obs = 11447	$t = \hat{\beta}/\hat{se}$	8.193	4.363	4.642	7.443	6.859	6.683
v2clrelig	Δ AIC	0.00	3619.57	3348.62	1495.92	2613.53	2420.66
	Vuong	0	3.169	3.519	2.622	3.040	3.232
Judicial Independence	$\hat{\beta}$	2.841	1.788	0.839	1.357	0.775	0.770
obs = 11293	$t = \hat{\beta}/\hat{se}$	11.156	5.271	5.410	8.522	7.204	7.119
v2juhcind	Δ AIC	0.00	3047.90	3099.07	1515.13	2225.93	2220.95
	Vuong	0	4.336	4.387	3.066	3.958	4.049
Property Rights	$\hat{\beta}$	0.456	0.271	0.115	0.202	0.121	0.122
obs = 11448	$t = \hat{\beta}/\hat{se}$	6.646	3.830	3.669	5.424	5.287	5.035
v2xcl_prpty	Δ AIC	0.00	2293.38	2422.82	1361.03	1686.85	1662.13
	Vuong	0	2.336	3.043	2.550	2.171	2.445

Table A6: Model Comparisons: Abuse

Variable	Stats	W	Support	W (old)	Polity2	Dem6	Boix
Media Censored	$\hat{\beta}$	5.401	3.384	2.158	2.996	1.536	1.674
obs = 9952	$t = \hat{\beta}/\hat{se}$	20.971	6.854	9.338	16.461	11.999	12.847
v2mecenefm	Δ AIC	0.00	6308.56	5299.15	1875.46	4553.81	3862.55
	Vuong	0	5.511	6.729	3.035	5.441	5.681
Packed Courts	$\hat{\beta}$	-1.959	-0.990	-0.692	-1.049	-0.611	-0.631
obs = 9942	$t = \hat{\beta}/\hat{se}$	-6.389	-2.530	-4.763	-6.713	-6.401	-6.789
v2jupack	Δ AIC	0.00	1369.39	1140.95	400.20	707.93	633.38
	Vuong	0	3.129	2.757	1.446	1.934	1.840
Torture	$\hat{\beta}$	-4.175	-2.760	-1.718	-2.235	-1.158	-1.210
obs = 11448	$t = \hat{\beta}/\hat{se}$	-16.377	-6.606	-12.018	-13.109	-9.862	-9.771
v2cltort	Δ AIC	0.00	4514.13	3705.40	1409.38	3380.01	3104.50
	Vuong	0	6.208	6.268	2.764	5.757	5.970
Slave Labor	$\hat{\beta}$	-0.306	-0.282	-0.093	-0.143	-0.083	-0.092
obs = 11448	$t = \hat{\beta}/\hat{se}$	-6.201	-3.386	-4.241	-4.859	-4.907	-4.781
	v2xclslave Δ AIC	0.00	824.68	1237.38	639.00	895.76	722.62
	Vuong	0	1.563	2.606	1.882	2.178	1.955

Table A7: Model Comparisons: Health and Education

Variable	Stats	W	Support	W (old)	Polity2	Dem6	Boix
Infant Mortality	$\hat{\beta}$	-25.757	-28.798	-24.633	-19.961	-10.380	-11.173
obs = 9412	$t = \hat{\beta}/\hat{se}$	-3.349	-2.629	-4.367	-3.993	-3.055	-3.518
e_peinfmor	Δ AIC	0.00	81.89	-246.42	-132.82	-3.36	-28.64
	Vuong	0	0.551	-1.392	-1.045	-0.025	-0.220
Campus Free	$\hat{\beta}$	4.687	3.830	1.915	2.623	1.385	1.453
obs = 9153	$t = \hat{\beta}/\hat{se}$	18.671	7.902	8.157	13.593	10.954	11.384
v2casurv	Δ AIC	0.00	5162.44	4871.13	1704.02	3996.12	3606.77
	Vuong	0	4.356	5.783	3.235	4.593	4.858
Academic Freedom	$\hat{\beta}$	1.073	0.781	0.439	0.589	0.309	0.330
obs = 9280	$t = \hat{\beta}/\hat{se}$	20.090	7.520	8.886	14.430	11.557	12.148
v2xca_acad	Δ AIC	0.00	8022.88	7210.89	3117.89	6243.95	5566.47
	Vuong	0	4.557	6.214	4.083	4.929	5.306
HealthCare	$\hat{\beta}$	0.319	1.420	0.311	0.143	0.219	0.216
obs = 9976	$t = \hat{\beta}/\hat{se}$	0.891	3.780	1.972	0.793	2.334	2.004
v2pehealth	Δ AIC	0.00	-480.73	-61.39	25.93	-94.66	-93.84
	Vuong	0	-1.645	-0.936	0.412	-0.977	-1.312

4 Analyses of Government Policies using Pairwise Comparisons

Tables A8, A9, A10, A11 and A12 contain fixed effects regressions examining Key Pubic Goods, Key Private Goods, Fundamental Freedoms, Abuses and Health and Education, respective. The analysis are conducted on a pairwise common sample (an observation is included if the data is available for W and the alternative institutional measure). For each dependent variable we report the number of observations and the Vuong t-statistics.

Table A8: Model Comparisons (Pairwise): Key Public Goods

Variable	Stats	Support	W (old)	Polity2	Dem6	Boix	Przeworski	GWF	TL
Public Good	Obs.	11838	11955	10386	10386	10393	7510	7236	11954
	Vuong	-0.299	3.029	2.047	1.530	1.451	2.070	1.459	2.954
Goods Ratio	Obs.	11838	11955	10386	10386	10393	7510	7236	11954
	Vuong	-1.175	1.381	1.096	0.047	-0.133	0.401	-1.153	1.852
Public Goods or Clientelism	Obs.	13779	13961	12063	12063	12053	7632	7321	13877
	Vuong	-3.409	-0.511	0.398	-0.837	-0.612	-0.709	-2.039	-1.438
Transparent Laws	Obs.	13776	13958	12060	12060	12050	7632	7321	13875
	Vuong	2.622	4.035	3.615	4.965	4.003	2.831	1.456	3.766
Elite Want Common Good	Obs.	11988	12105	10470	10470	10515	7632	7321	12104
	Vuong	-3.062	1.794	0.956	1.300	1.221	2.224	0.106	0.982
Public Want Common Good	Obs.	13783	13965	12065	12065	12055	7632	7321	13881
	Vuong	4.383	5.485	3.789	6.794	6.349	5.261	5.172	4.318
Free Expression	Obs.	13757	13939	12041	12041	12030	7632	7321	13856
	Vuong	5.930	6.778	4.249	6.557	7.227	7.064	5.954	6.271

Table A9: Model Comparisons (Pairwise): Key Private Goods

Variable	Stats	Support	W (old)	Polity2	Dem6	Boix	Przeworski	GWF	TL
State Owns Economy	Obs.	13783	13966	12065	12065	12055	7632	7321	13881
	Vuong	2.298	2.328	2.131	2.306	2.355	2.470	1.360	-1.140
Corruption	Obs.	13779	13961	12063	12063	12053	7632	7321	13877
	Vuong	-0.480	-0.258	-0.385	-1.422	-1.230	-1.143	-1.812	-1.156
Executive Corrupt	Obs.	13779	13923	12025	12025	12015	7632	7321	13877
	Vuong	1.553	1.316	1.014	0.122	0.513	-0.815	-1.873	-1.244
Public Sector Theft	Obs.	13723	13905	12007	12007	11997	7632	7321	13821
	Vuong	1.135	1.235	1.317	0.132	0.700	-0.722	-1.651	-0.236
Clientelism	Obs.	13705	13872	12034	12034	12010	7632	7321	13791
	Vuong	-0.002	-0.115	0.430	-0.760	-0.898	-0.385	-1.555	-1.444
Judge Corrupt	Obs.	13727	13910	12030	12030	12020	7598	7321	13825
	Vuong	0.625	0.179	0.278	-1.501	-1.077	-1.273	-1.767	-1.697

Table A10: Model Comparisons (Pairwise): Fundamental Freedoms

Variable	Stats	Support	W (old)	Polity2	Dem6	Boix	Przeworski	GWF	TL
Respect for Constitution	Obs.	13725	13855	11974	11974	11950	7632	7321	13810
	Vuong	4.589	3.233	3.547	4.854	4.508	1.939	1.326	1.573
Rule of Law	Obs.	13783	13966	12065	12065	12055	7632	7321	13881
	Vuong	4.193	4.619	3.370	3.703	3.304	1.918	0.342	4.462
Free Elections	Obs.	3661	3712	3378	3378	3349	1979	1912	3692
	Vuong	6.814	7.227	2.361	4.718	4.744	5.871	4.667	6.822
Free Movement	Obs.	13783	13966	12065	12065	12055	7632	7321	13881
	Vuong	1.214	3.272	2.515	3.084	3.218	3.173	1.563	2.384
Civil Liberties	Obs.	13781	13963	12064	12064	12054	7632	7321	13880
	Vuong	7.278	8.413	5.859	7.985	8.592	8.868	8.071	8.353
Political Rights	Obs.	6625	6706	6315	6315	6236	5144	4961	6706
	Vuong	6.415	3.659	-1.125	3.955	0.697	2.361	3.715	5.094
Religious Freedom	Obs.	13781	13964	12064	12064	12054	7632	7321	13879
	Vuong	3.098	3.657	2.777	3.130	3.437	3.868	3.329	2.529
Judicial Independence	Obs.	13527	13695	11908	11908	11884	7601	7321	13615
	Vuong	3.849	4.201	3.131	3.825	4.024	3.635	2.403	3.692
Property Rights	Obs.	13782	13965	12064	12064	12055	7632	7321	13880
	Vuong	1.229	2.944	2.535	2.223	2.598	3.743	2.454	0.292

Table A11: Model Comparisons (Pairwise): Abuse

Variable	Stats	Support	W (old)	Polity2	Dem6	Boix	Przeworski	GWF	TL
Media Censored	Obs.	11963	12080	10446	10446	10490	7632	7321	12079
	Vuong	5.675	6.500	3.029	5.400	5.791	5.444	5.357	5.128
Packed Courts	Obs.	11933	12050	10436	10436	10481	7598	7321	12049
	Vuong	3.304	2.613	1.406	1.992	1.826	2.011	-0.232	1.621
Torture	Obs.	13783	13966	12065	12065	12055	7632	7321	13881
	Vuong	4.648	4.628	2.809	5.722	5.439	3.043	1.342	4.632
Slave Labor	Obs.	13782	13965	12065	12065	12055	7632	7321	13880
	Vuong	-1.453	2.800	2.048	2.203	2.289	2.866	2.302	0.657

Table A12: Model Comparisons (Pairwise): Health and Education

Variable	Stats	Support	W (old)	Polity2	Dem6	Boix	Przeworski	GWF	TL
Infant Mortality	Obs.	10513	10652	9556	9556	9969	7227	6926	10615
	Vuong	1.296	-0.497	-0.402	0.384	0.213	0.861	-1.402	-1.294
Campus Freedom	Obs.	10478	10538	9631	9631	9558	6928	6762	10537
	Vuong	4.799	5.630	3.284	4.669	4.847	3.391	3.185	4.891
Academic Freedom	Obs.	10605	10665	9758	9758	9685	7008	6842	10664
	Vuong	4.926	5.917	4.129	4.952	5.218	5.056	4.124	5.233
Health Care	Obs.	11988	12105	10470	10470	10515	7632	7321	12104
	Vuong	-2.401	-1.521	0.377	-1.102	-1.385	-1.178	-2.081	-1.724

5 Democracies

Tables A13, A14, A15, A16 and A17 contain fixed effects regressions examining Key Pubic Goods, Key Private Goods, Fundamental Freedoms, Abuses and Health and Education, respective. The analysis is restricted to nation-year observation coded as democracies (Polity2 $\geq$ 6). The analysis are conducted on a pairwise common sample (an observation is included if the data is available for W and the alternative institutional measure). For each dependent variable we report the number of observations and the Vuong t-statistics.

Table A13: Model Comparisons for Democracies (Pairwise): Key Public Goods

Variable	Stats	Support	W (old)	Polity2	Boix	Przeworski	GWF	TL
Public Goods	Obs.	4387	4394	4394	4119	2865	2806	4394
	Vuong	1.410	2.444	2.237	2.585	2.084	1.648	1.712
Goods Ratio	Obs.	4387	4394	4394	4119	2865	2806	4394
	Vuong	0.924	1.972	2.001	1.913	1.635	1.409	1.552
Public Goods or Clientelism	Obs.	4687	4694	4694	4417	2874	2815	4694
	Vuong	1.657	1.534	1.664	1.747	1.258	0.998	0.632
Transparent Laws	Obs.	4687	4694	4694	4417	2874	2815	4694
	Vuong	2.107	2.580	2.701	2.825	2.692	2.125	1.996
Elite Want Common Good	Obs.	4396	4403	4403	4128	2874	2815	4403
	Vuong	1.686	2.484	2.535	2.562	2.195	2.100	2.290
Public Want Common Good	Obs.	4687	4694	4694	4417	2874	2815	4694
	Vuong	1.542	3.573	3.585	3.435	2.743	1.993	2.199
Free Expression	Obs.	4687	4694	4694	4417	2874	2815	4694
	Vuong	2.610	3.145	3.109	2.987	2.545	1.883	2.873

Table A14: Model Comparisons for Democracies (Pairwise): Key Private Goods

Variable	Stats	Support	W (old)	Polity2	Boix	Przeworski	GWF	TL
State Owns Economy	Obs.	4687	4694	4694	4417	2874	2815	4694
	Vuong	0.854	0.684	0.609	0.720	1.729	1.447	-0.125
Corruption	Obs.	4687	4694	4694	4417	2874	2815	4694
	Vuong	0.591	1.260	1.041	1.364	1.260	1.420	1.042
Executive Corrupt	Obs.	4687	4694	4694	4417	2874	2815	4694
	Vuong	2.609	2.532	2.474	2.653	1.938	2.061	2.050
Public Sector Theft	Obs.	4659	4666	4666	4389	2874	2815	4666
	Vuong	0.827	1.091	1.448	1.504	1.841	1.725	0.765
Clientelism	Obs.	4687	4694	4694	4417	2874	2815	4694
	Vuong	1.926	1.695	1.758	1.971	1.425	1.870	1.391
Judge Corrupt	Obs.	4687	4694	4694	4417	2874	2815	4694
	Vuong	0.128	0.787	0.162	0.860	0.113	0.663	0.727

Table A15: Model Comparisons for Democracies (Pairwise): Fundamental Freedoms

Variable	Stats	Support	W (old)	Polity2	Boix	Przeworski	GWF	TL
Respect for Constitution	Obs.	4687	4694	4694	4417	2874	2815	4694
	Vuong	1.818	2.218	2.059	2.037	1.245	0.956	0.972
Rule of Law	Obs.	4687	4694	4694	4417	2874	2815	4694
	Vuong	3.502	3.338	3.176	3.041	2.784	2.607	2.499
Free Elections	Obs.	1467	1470	1470	1392	885	885	1470
	Vuong	2.914	3.260	2.722	3.013	1.704	2.040	2.685
Free Movement	Obs.	4687	4694	4694	4417	2874	2815	4694
	Vuong	1.905	2.249	2.082	2.069	1.709	1.247	1.594
Civil Liberties	Obs.	4687	4694	4694	4417	2874	2815	4694
	Vuong	2.929	3.512	3.445	3.179	2.771	2.610	3.323
Political Rights	Obs.	2933	2937	2937	2672	2074	2096	2937
	Vuong	1.861	0.376	0.179	-0.336	0.639	0.962	0.235
Religious Freedom	Obs.	4687	4694	4694	4417	2874	2815	4694
	Vuong	1.515	1.981	1.994	2.022	2.024	1.807	1.599
Judicial Independence	Obs.	4675	4682	4682	4405	2874	2815	4682
	Vuong	2.526	2.276	2.358	2.203	1.745	1.115	1.825
Property Rights	Obs.	4687	4694	4694	4417	2874	2815	4694
	Vuong	1.556	2.203	1.886	2.336	1.963	1.883	0.841

Table A16: Model Comparisons for Democracies (Pairwise): Abuse

Variable	Stats	Support	W (old)	Polity2	Boix	Przeworski	GWF	TL
Media Censored	Obs.	4396	4403	4403	4128	2874	2815	4403
	Vuong	2.475	3.046	3.063	2.183	2.588	1.951	1.288
Packed Courts	Obs.	4396	4403	4403	4128	2874	2815	4403
	Vuong	3.147	3.501	3.488	2.767	1.772	1.718	3.819
Torture	Obs.	4687	4694	4694	4417	2874	2815	4694
	Vuong	1.873	2.214	2.191	2.231	2.096	1.006	0.949
Slave Labor	Obs.	4687	4694	4694	4417	2874	2815	4694
	Vuong	1.062	1.288	1.147	1.226	1.300	0.942	-0.738

Table A17: Model Comparisons for Democracies (Pairwise): Health and Education

Variable	Stats	Support	W (old)	Polity2	Boix	Przeworski	GWF	TL
Infant Mortality	Obs.	4081	4087	4087	4078	2833	2775	4087
	Vuong	1.247	2.038	1.546	2.247	1.255	1.045	0.216
Campus Freedom	Obs.	4174	4178	4178	3906	2749	2690	4178
	Vuong	1.706	3.547	2.794	3.177	1.903	2.077	2.465
Academic Freedom	Obs.	4280	4284	4284	4012	2808	2749	4284
	Vuong	2.245	3.029	2.756	2.884	2.224	2.042	2.795
HealthCare	Obs.	4396	4403	4403	4128	2874	2815	4403
	Vuong	1.167	1.257	1.177	1.316	1.125	0.858	0.930

6 Autocracies

Tables A18, A19, A20, A21 and A22 contain fixed effects regressions examining Key Pubic Goods, Key Private Goods, Fundamental Freedoms, Abuses and Health and Education, respective. The analysis is restricted to nation-year observation coded as non-democracies (Polity2 < 6). The analysis are conducted on a pairwise common sample (an observation is included if the data is available for W and the alternative institutional measure). For each dependent variable we report the number of observations and the Vuong t-statistics.

Table A18: Model Comparisons for Autocracies (Pairwise): Key Public Goods

Variable	Stats	Support	W (old)	Polity2	Anoc.	Boix	Przeworski	GWF	TL
Public Goods	Obs.	5952	5992	5992	5992	5812	4274	4171	5992
	Vuong	-1.648	1.505	1.294	1.588	0.928	0.216	-0.003	0.594
Goods Ratio	Obs.	5952	5992	5992	5992	5812	4274	4171	5992
	Vuong	-2.277	0.834	0.675	0.828	0.482	-0.714	-1.319	-0.280
Public Goods or Clientelism	Obs.	7237	7329	7329	7329	7080	4348	4245	7286
	Vuong	-2.500	-0.360	0.347	0.327	0.309	-0.305	-1.485	-1.183
Transparent Laws	Obs.	7234	7326	7326	7326	7077	4348	4245	7283
	Vuong	3.399	4.090	3.143	4.694	4.233	1.524	1.228	2.969
Elite Want Common Good	Obs.	6027	6067	6067	6067	5887	4348	4245	6067
	Vuong	-2.322	0.419	-0.265	1.199	-0.183	1.194	-0.796	-0.432
Public Want Common Good	Obs.	7239	7331	7331	7331	7082	4348	4245	7288
	Vuong	4.010	5.335	2.401	3.543	4.086	3.522	4.147	1.877
Free Expression	Obs.	7215	7307	7307	7307	7058	4348	4245	7265
	Vuong	5.142	6.232	3.137	4.806	5.284	4.709	5.525	5.270

Table A19: Model Comparisons for Autocracies (Pairwise): Key Private Goods

Variable	Stats	Support	W (old)	Polity2	Anoc.	Boix	Przeworski	GWF	TL
State Owns Economy	Obs.	7239	7331	7331	7331	7082	4348	4245	7288
	Vuong	2.114	2.263	1.485	1.751	2.175	1.875	1.553	-1.615
Corruption	Obs.	7237	7329	7329	7329	7080	4348	4245	7286
	Vuong	-1.575	-0.025	0.058	0.049	-0.032	-0.103	-1.061	-1.308
Executive Corrupt	Obs.	7237	7291	7291	7291	7042	4348	4245	7286
	Vuong	-1.169	0.201	0.427	0.745	0.588	-0.864	-1.648	-1.265
Public Sector Theft	Obs.	7207	7299	7299	7299	7050	4348	4245	7256
	Vuong	-1.411	0.447	0.668	0.641	0.576	-0.031	-1.362	-0.710
Clientelism	Obs.	7210	7300	7300	7300	7051	4348	4245	7259
	Vuong	-0.876	-0.370	-0.093	0.129	0.023	0.135	-1.232	-2.119
Judge Corrupt	Obs.	7204	7296	7296	7296	7047	4314	4245	7253
	Vuong	-1.068	0.240	0.171	0.119	0.246	0.035	-1.858	-1.398

Table A20: Model Comparisons for Autocracies (Pairwise): Fundamental Freedoms

Variable	Stats	Support	W (old)	Polity2	Anoc.	Boix	Przeworski	GWF	TL
Respect for Constitution	Obs.	7187	7240	7240	7240	6991	4348	4245	7235
	Vuong	2.996	3.065	3.128	4.118	3.379	0.869	0.960	0.356
Rule of Law	Obs.	7239	7331	7331	7331	7082	4348	4245	7288
	Vuong	1.694	2.841	2.236	3.666	2.932	0.945	0.213	2.166
Free Elections	Obs.	1844	1864	1864	1864	1783	979	958	1848
	Vuong	5.160	6.099	1.028	4.412	4.810	3.905	1.771	4.526
Free Movement	Obs.	7239	7331	7331	7331	7082	4348	4245	7288
	Vuong	1.714	2.700	1.257	1.468	2.765	1.964	1.683	1.609
Civil Liberties	Obs.	7238	7330	7330	7330	7081	4348	4245	7288
	Vuong	6.253	7.578	5.230	5.591	6.571	6.379	7.350	6.829
Political Rights	Obs.	3345	3378	3378	3378	3201	2769	2689	3378
	Vuong	6.290	2.899	-0.055	3.702	2.813	2.421	4.064	5.123
Religious Freedom	Obs.	7238	7330	7330	7330	7081	4348	4245	7287
	Vuong	3.336	3.669	2.685	2.965	3.486	2.787	1.644	0.255
Judicial Independence	Obs.	7096	7186	7186	7186	6937	4317	4245	7145
	Vuong	2.272	2.869	1.312	1.541	2.488	2.503	0.967	1.033
Property Rights	Obs.	7238	7330	7330	7330	7082	4348	4245	7287
	Vuong	1.847	2.164	1.956	1.975	2.232	1.977	1.141	-1.361

Table A21: Model Comparisons for Autocracies (Pairwise): Abuse

Variable	Stats	Support	W (old)	Polity2	Anoc.	Boix	Przeworski	GWF	TL
Media Censored	Obs.	6003	6043	6043	6043	5863	4348	4245	6043
	Vuong	4.795	5.692	2.298	3.520	4.236	3.857	4.269	4.213
Packed Courts	Obs.	5993	6033	6033	6033	5853	4314	4245	6033
	Vuong	2.464	2.285	0.945	2.045	2.024	1.433	-0.314	-0.801
Torture	Obs.	7239	7331	7331	7331	7082	4348	4245	7288
	Vuong	3.545	4.286	1.094	2.857	4.078	2.214	1.937	3.236
Slave Labor	Obs.	7239	7331	7331	7331	7082	4348	4245	7288
	Vuong	0.225	1.854	0.956	1.251	1.884	1.721	1.516	0.378

Table A22: Model Comparisons for Autocracies (Pairwise): Health and Education

Variable	Stats	Support	W (old)	Polity2	Anoc.	Boix	Przeworski	GWF	TL
Infant Mortality	Obs.	5381	5447	5447	5447	5384	3989	3893	5411
	Vuong	0.499	-1.327	-0.671	0.894	0.641	0.478	-1.252	-1.040
Campus Freedom	Obs.	5417	5453	5453	5453	5279	3890	3878	5453
	Vuong	3.255	4.225	2.587	3.576	3.428	1.826	2.486	3.659
Academic Freedom	Obs.	5438	5474	5474	5474	5300	3911	3899	5474
	Vuong	3.482	4.561	3.357	3.981	3.810	3.068	3.692	4.023
HealthCare	Obs.	6027	6067	6067	6067	5887	4348	4245	6067
	Vuong	-0.006	1.064	0.757	-0.211	1.018	0.598	-1.212	-1.547

7   Repeated Sampling

V-dem utilizes an item response theory (IRT) methodology to convert codings from country experts into a latent score on each dimension using Bayesian MCMC algorithms (Pemstein et al., 2019; Marquardt and Pemstein, 2017). The method generates random samples from the posterior distribution of the latent variable (and numerous other threshold parameters). The measure of W in the main text used the mean value from these distributions. A legitimate concern is that the inferences drawn from using these point estimates might differ significantly from those drawn considering the distribution of W. Here we address this concern.

For release of version 7 of their data, V-dem provides a sample of 900 random draws from the posterior distribution. These samples are archived at https://curate.nd.edu/show/c534fn1308k (accessed 8/15/2021). We use these samples for the v2elembaut, v2psbars and v2psoppaut variables to create 900 versions of the W measure and repeat the analysis in Table 2. Version 7 did not contain samples for the fourth component of our index and so we used the point estimates in version 11. Likewise we used the point estimates of PublicGoods in Version 11. For each of the 900 samples we recreate W and calculate the AIC criteria for the fixed effect regression of PublicGoods. Figure 1 shows the density of the distribution of the AIC statistic (for the common sample of 9,478 observations for which we have the W components in version 7 as well as the other measures). The vertical lines correspond to the AIC statistic using the alternative measures. The figure shows that for all 900 draws of W from the posterior distribution, the AIC for the W measure is substantially smaller than the AIC for the other measures. This figure provides reassurance that the inference drawn from the measure created using point estimates reflects the inferences taking the full distribution into account.
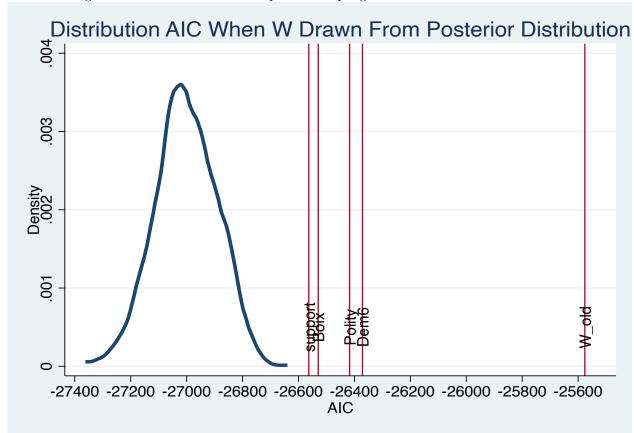
Figure 1: Distribution of AIC: Repeated Sampling of W from Posterior Distribution

References

Akaike, Hirotugu. 1974. "A new look at the statistical model identification." *IEEE transactions on automatic control* 19(6):716–723.

Burnham, Kenneth P and David R Anderson. 1998. *Model Selection and Inference: A practical information-theoretic approach.* Springer, New York.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I Lindberg, Jan Teorell, David Altman, Michael Bernhard, Agnes Cornell, M Steven Fish, Lisa Gastaldi et al. 2021. "V-Dem Codebook v11.".

Marquardt, Kyle L and Daniel Pemstein. 2017. "IRT models for expert-coded panel data." *V-Dem Working Paper* 41.

Pemstein, Daniel, Kyle L Marquardt, Eitan Tzelgov, Yi-ting Wang, Joshua Krusell and Farhad Miri. 2019. "The V-Dem measurement model: latent variable analysis for cross-national and cross-temporal expert-coded data?" *V-Dem Working Paper* 21.

Teorell, Jan and Staffan I Lindberg. 2019. "Beyond democracy-dictatorship measures: a new framework capturing executive bases of power, 1789–2016." *Perspectives on Politics* 17(1):66–84.

Vuong, Quang H. 1989. "Likelihood ratio tests for model selection and non-nested hypotheses." *Econometrica: Journal of the Econometric Society* pp. 307–333.

Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data.* MIT press.