A surprising consensus among diverse ways to measure democracy

Word count: 9,781

Abstract

We show that behind decades of disagreements over how to classify regime types, there is an unexpectedly strong consensus about which countries have been democratic in which years. We match many-valued classifications of democracies and autocracies to binary classifications, by finding the cutpoint that dichotomizes each many-valued measure so that it matches each binary measure as closely as possible. Despite the sharp differences between these datasets, we find nearly identical optimal cutpoints for almost every major classification. This is evidence of a strong underlying agreement between the classifications. We show that the many-valued measures, when dichotomized in the optimal way, can replicate major results in the study of democracy. We also examine what sorts of countries are counted as democracies in this consensus.

1 Introduction

Surveying the literature on democracy in 1942, Schumpeter lamented that the classical doctrine of democracy could describe "an infinite wealth of possible forms", evading consensus even after "the legal mind ransacked the lumber room of its constructs in search for tools by which to reconcile" theoretical definitions of democracy with the observed reality of politics (Schumpeter 1942, 247). In the nearly eighty years since then, political scientists have largely settled on a theory of democracy that relies on certain observable features. But when it comes to deciding exactly which features are necessary parts of democracy, our lumber rooms have produced such an unruly range of constructs that surveys of the democratization literature report broad disagreement about how to classify a particular country as a democracy or an autocracy in a particular year.

Discussing the ongoing boom in methods for classifying democracies and autocracies, Cheibub et al. (2010, 67) argue that "differences across regime measures must be taken seriously". Boix et al. (2012, 1525) reinforce that differences between the "dozens of distinct measures of democracy" are "not merely academic", since "empirical results can depend on the specific measure of democracy used". And after surveying a "proliferation of binary indices that identify different defining conditions of democracy" and measurement decisions that are both arbitrary and informative "in strikingly different ways", Skaaning et al. (2015, 1492-4) note that such endemic disagreements might actually be desirable, since "the discipline is well served by a variety of measures". Indeed, it is not surprising that scholars of democracy have developed many classifications that are each well-developed and theoretically defensible, but are substantively distinct. As Geddes (1999) wrote about the difficulty of coding democratic or autocratic transitions, "the basic problem faced by analysts is that the process of democratization varies enormously from case to case and region to region."

Certainly these authors are all correct that substantively different codings can be defended on their own terms, and should not be carelessly used interchangeably. In fact, just a few decades ago Casper and Tufis (2003) pushed back against the practice of thoughtlessly swapping out one measure of democracy for another in analyses, arguing that scholars treated the available measures of democracy "as virtually interchangeable". Because the existing classifications of democracies and autocracies were highly correlated, scholars were lulled into believing that they could freely use whichever analysis was more empirically convenient, but Casper and Tufis (2003) showed that an analysis might produce opposite results when conducted on different classifications of democracy. In the intervening years the field seems to have largely adopted Casper and Tufis's view. Now the main consensus in the study of how to classify democracies is that there is no consensus on how to classify democracies.

As efforts to classify democracies have rapidly mushroomed in the past few years, are they still characterized by the high correlations that Casper and Tufis (2003) worried were misleading, or is the situation closer to the dissonant landscape that more recent analyses have described? In Figure 1, we compare the level of agreement among the major reasonably independent dichotomous (or easily dichotomized) measures of democracy. The figure shows how often each pair of datasets agrees about whether or not a given country-year is a democracy, among those country-years that appear in both datasets.

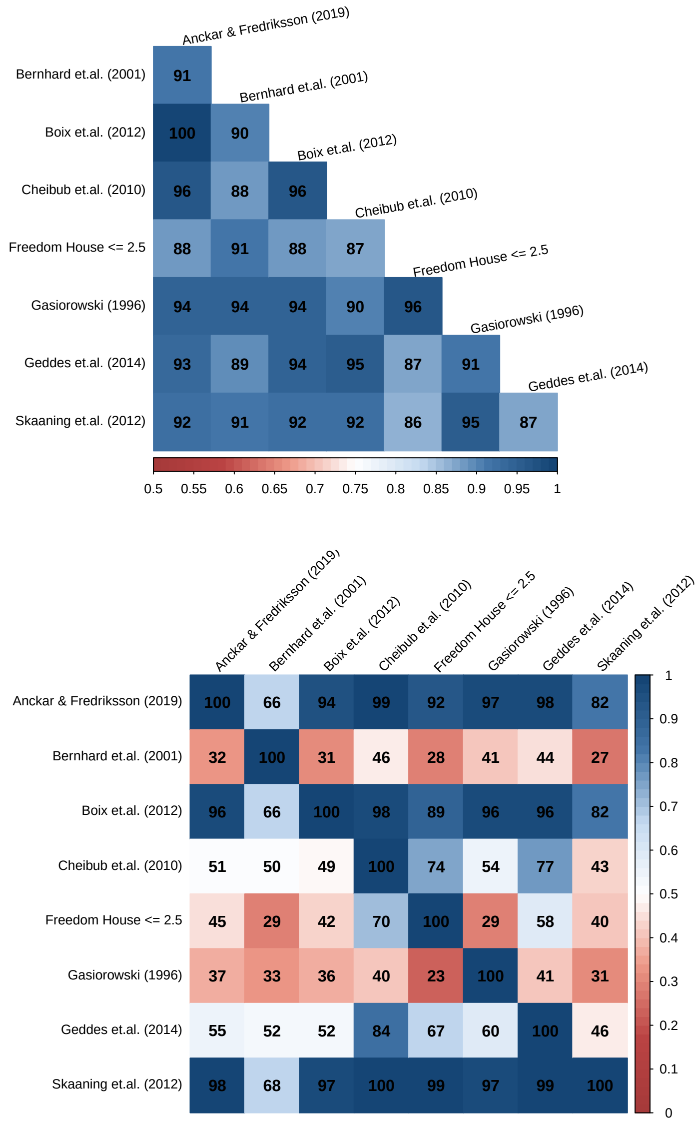

Figure 1: Top: Each cell shows the pairwise agreement between the datasets in that row and column. Bottom: Each cell shows the percentage of the dataset in that column which also appears in the dataset in that row. For example, Geddes et al. (2014) has 94% agreement with Boix et al. (2012), 96% of the cases in Geddes et al. (2014) are in Boix et al. (2012), and 52% of the cases that appear in Boix et al. (2012) also appear in Geddes et al. (2014).

The top of Figure 1 shows the pairwise agreement between each measure of democracy, while the bottom of Figure 1 shows how many country-years are included in both datasets. A typical level of agreement between these datasets is upwards of 90%. This puts numbers to the common refrain that there is little consensus in this area of study, though it is open to discussion whether that is a substantively high or low rate for classifying country-years as democracies. But pairwise comparisons can only be so informative when fewer than a quarter of the cases in one dataset are included in the other, and even if every dataset were almost perfectly correlated, it would still matter that these datasets make systematically different substantive assumptions (Casper and Tufis 2003).

And yet, in the Analysis section of this paper, we will show that lurking behind the apparent discord there is a natural way to achieve a remarkably strong consensus regarding which countries have been democracies in which years. We pick up on one of the biggest disagreements in this topic: whether to use dichotomous or many-valued measures of democracy. An idea that has recently gained traction is to dichotomize a many-valued measure of democracy so that it matches a binary classification as closely as possible, by picking a value (called a "cutpoint") such that every case above that value is a democracy and every case below that value is an autocracy.

We find the best cutpoint for matching the major many-valued measures of democracy to eight dichotomous classifications. For many-valued measures that have as many as 1001 possible cutpoints, we find that *multiple different binary classifications of democracy share exactly the same best cutpoint*, and all the cutpoints are close to each other. Even datasets with extremely different classification rules have extremely similar optimal cutpoints. This represents a deep form of agreement between classifications of democracy, which has not been captured in previous substantive discussions or pairwise comparisons. In the Replications section, we use the modal best cutpoint to produce a consensus binary classification, and we replicate a central substantive result from two major binary measures.

In this paper we identify a classification that is surprisingly similar to the codings of all of the major democracy datasets. However, our primary aim is not to produce a new dataset, and we are not arguing that this classification should be the new standard in empirical analyses that require a country-year coding of democracies. Our main point is theoretical, and our secondary point is empirical. Theoretically, we argue that there is a much stronger consensus on how to classify democracies than the current literature suggests. Empirically, we are not constructing a measure of democracy *per se*; really we are measuring *the consensus about democracy*. In this sense, our goal is closer to meta-analysis than to classification.

There are occasions when it is useful to have a measure of what other researchers think about a topic (say, in checking how novel a new classification rule is), and times when researchers may simply wish to rely on a scholarly consensus of how cases should be classified. For example, when choosing a measurement for robustness checking, it could be useful to resort to the one that most closely approximates existing measures. Also, instead of choosing a democracy measure to work with based on their own theoretical preferences, scholars may prefer that their results are agnostic to their theoretical choices, in which case a consensus measure that best approximates other main measures in the field may be of interest. Dichotomizing existing multi-valued yearly measures that span centuries also produces dichotomous classifications of democracy over earlier historical periods than the existing dichotomous measures. So it can be useful to have access to a consensus classification like the one we identify. But the central point goes beyond practicality: users of any measure of democracy should be aware that, when we compare these datasets using a fuller method than pairwise comparisons, we uncover an exceptionally strong consensus.

1.1 Democracy: A matter of degree or of type?

Since the beginnings of political science, identifying which political entities were democratic has been a major concern (Popper 2020, Schumpeter 1942). Even Dahl's 1973 seminal theoretical book on polyarchy was based on a classification of countries' democratic status circa 1969. As a consequence, over the decades, a great deal of the debate on the processes of democratization — and recently, of de-democratization — has centered around the details of how to classify political regimes.

Among the many methodological disagreements, perhaps the most critical has been whether democracies are a matter of type or of degree (Boix et al. 2012, Collier and Adcock

1999, Teorell and Lindberg 2019). Early work approached regime type as a largely all-or-nothing proposition, classifying some regimes as democracies and others as non-democracies at a given point in time (Dahl 1973, Lipset 1960). As efforts to chart countries' experiences with democracy over time became more common, researchers began to introduce numerical proxies for democratization, like the percentage of the population voting in national elections (Lerner 1958).

Cutright (1963) was among the first to propose a many-valued scale of democracy, which he calculated by combining his subjective scores for a series of institutional characteristics within a given country. Cutright's measure was a standard empirical variable in cross-country comparisons and econometric analyses for nearly two decades. However, Bollen (1980) argued that Cutright's subjective measure was endogenous to the very political and economic processes it was often used to investigate. Bollen proposed, instead, a score based on country characteristics more exogenous to institutional idiosyncrasies, like the presence of free press and free opposition, the fairness of elections, and whether the legislative and executive branches (when existent) were elected. While similar in spirit to measures of democracies used today, Bollen's measure covered only a few countries for a limited number of years.

This early work on many-valued measures of democracy was partly a reaction to the perceived limits of dichotomous classification schemes. In later work with Jackman (1989) Bollen summarized the common criticism of dichotomous measures: "Dichotomizing democracy lumps together countries with very different degrees of democracy and blurs distinctions between borderline cases" (Bollen and Jackman 1989, 612).

And yet, dichotomous classifications of democracy have not only re-emerged, but become more and more common in empirical analyses involving democracy. Alvarez et al. (1996), who offered their own dichotomous classification scheme, responded to Bollen and Jackman (1989) that "it is one thing to argue that some democracies are more democratic than others and another to argue that democracy is a continuous feature over all regimes" (21).

Even if there are different degrees of democracy and autocracy, these authors argue that there is still a real discontinuity between these types, and that a country year cannot be "half-democratic: there is a natural zero point" (*ibidem*). Alvarez et al. (1996) drew on minimalistic notions of democracy, as established by Schumpeter (1942) and Dahl (1973), defining democracy as political regimes where governing offices are filled via competitive elections. For each country-year they identify whether elections were used in choosing current executive and legislative branches, whether those elections had more than one party, and whether there had ever before been an alternation in power (the turnover rule). Relying only on observable historical information about countries, they were able to classify many more countries and years than any previous study, covering a total of 141 countries through 41 years. Their classification became one of the most widely used, it has been updated several times, (Bormann and Golder 2013, Cheibub et al. 2010, Przeworski et al. 2000) and their classification scheme has been used as the basis for multiple future classifications (Boix et al. 2012, Geddes et al. 2014, Svolik 2008).

Alongside these new dichotomous measures, two ordinal indices have also been frequently used by researchers: Puddington et al. (2018) (Freedom House) and Marshall et al. (2017) (the Polity Project). These two large-scale projects attempt to combine some of the most appealing characteristics of the previous efforts to classify regime type. First, they are both ordinal measures of country-year democracy levels, following in the tradition inaugurated by Bollen (1980). Second, like Alvarez et al. (1996) these measures seek comprehensive country coverage and extensive over-time coverage, with continual updates each year. Third, despite their ordinal nature, they are nearly always described as having a specific cutpoint in their ordinal scales that would define when a country-year could be considered democratic, akin to the natural zero point mentioned by Alvarez et al. (1996).

More recently, a number of new non-dichotomous alternatives have emerged, most notably the Varieties of Democracy project (V-Dem), with its emphasis on comprehensiveness and transparency (Coppedge et al. 2019). V-Dem relies on a very large team of experts selected from around the world to classify, among other things, the degree of political freedoms, electoral fairness, and electoral competitiveness of countries in each year since the French Revolution. Using a Bayesian composition approach V-Dem produces country-year measures corresponding to different conceptions of democracy, each with estimated confidence intervals. The primary measure of levels of democracy in V-Dem, called polyarchy, offers for the first time a non-ordinal many-valued measure (it ranges from 0 to 1, with 3 decimal places), with error estimation, clear classification rules, a large international team of expert classifiers, and covering a long historical period.

Interestingly, while each of these non-dichotomous initiatives have been highly successful, often researchers choose to convert measures into dichotomous variables at some cutpoint along their scale, so that every value above the cutpoint is a democracy and every value below the cutpoint is an autocracy (Bogaards 2010). This has long been common practice with Freedom House and Polity, and the same has been done with V-Dem data (Kasuya and Mori 2019, Lührmann et al. 2018, Wahman et al. 2013). And some highly sophisticated methods have been developed to transform a large number of democratic classifications into a classification with fewer values (Lührmann et al. 2018, Pemstein et al. 2010). Many have also criticized the practice of transforming political regime degrees into types. But it is important to stipulate that this practice does not assert that it is inherently better to have either a many-valued measure of democracy or a dichotomous measure of democracy; it simply argues that these types of measures have distinct advantages and drawbacks, and for some research questions it can be useful to be able to turn one type into the other.

A particularly contentious question is which value to choose as the cutpoint, particularly since each value of the many-valued measures could result from many different combinations of their underlying components (Gleditsch and Ward 2006). Bogaards (2010) shows that altering the choice of where to cut the commonly used Freedom House and Polity ordinal measures of democracy level to create a binary classification of democracy actually changes empirical results of previously published models. And Boix et al. (2012, 1528) lament that none of the authors who use cutpoints "offer a concrete reason for the thresholds other than claiming they are intuitive or citing another study that uses the same threshold."

Nevertheless, to discipline a choice of cutpoints, there is a simple option: we could simplify a many-valued measure of democracy using whichever cutpoint produces a binary variable that agrees with previous binary measures as much as possible. Unfortunately this requires a compromise, because we might expect a many-valued measure of democracy to have a different optimal cutpoint for each binary dataset that we wish to match to (Boese 2019, Boix et al. 2012, Kasuya and Mori 2019). However, using a series of datasets that have been central to the comparative study of democracy, we will show that such a compromise might not be necessary.

2 Analysis

In this section we find the optimal cutpoint for matching the Varieties of Democracy (V-Dem) polyarchy variable (Coppedge et al. 2019) and the Polity Project's Polity IV variable (Marshall et al. 2017) onto the classifications by Anckar and Fredriksson (2019), Bernhard et al. (2001), Boix et al. (2012),[1] Cheibub et al. (2010), Gasiorowski (1996), Geddes et al. (2014), Puddington et al. (2018), and Skaaning et al. (2015). For the polyarchy variable, which ranges from 0 to 1 with three digits of precision, we show that of the 1001 possible optimal cutpoints, multiple different binary classifications of democracy share *exactly the same best value*, and all of the values are close to each other. We observe a similar pattern for polity, though it only has 21 possible values. Crucially, we argue that this represents a form of agreement between the binary classifications of democracy, and is not just some artifact of the method. This is a highly unexpected level of agreement between datasets that were produced by several groups of authors over decades, with different research questions

[1] All of the results presented here refer to the original 2012 version of the Boix et al. (2012) dataset, because our analysis will include empirical replications of portions of that paper. Note however that the dataset was updated to include country-years through 2015 (Boix et al. 2018). When our analyses are run on the 2018 version of the dataset instead of the 2012 version, none of our substantive conclusions change, and the optimal cutpoint is identical.

in mind, aiming to capture substantively distinct notions of democracy and applying those notions to different cases, and disagreeing about as many as one in every six country-years.

2.1 Finding optimal cutpoints

In Figure 1 we showed that the main existing dichotomous datasets typically have fairly strong pairwise agreement, as previous authors have found (Casper and Tufis 2003). But there are two crucial limitations to pairwise comparisons. First, the number of cases in the intersection of any two datasets shows that pairwise comparisons can rarely tell the whole story; commonly, for large datasets like Boix et al. (2012), pairwise comparisons can only be made on the basis of a quarter or a third of the dataset, leaving out thousands of country-years. It is also easy to imagine that the country-years in the intersection of any two datasets will systematically tend to be the country-years that the field is most certain about, since these will be the country-years that are more commonly studied and classified, in which case the intersection of two datasets will tend to have a misleadingly high pairwise agreement. Second, pairwise comparisons tell us nothing about how closely these dichotomous measures agree with the many high-quality many-valued measures of democracy.

This motivates the idea of using a many-valued classification of democracy that is part of a very large dataset, like V-Dem or Polity, and identifying the optimal cutpoint for matching it to a binary dataset. But first, it is necessary to describe precisely what we mean by a "cutpoint". Figure 2 illustrates the basic problem, using a fictional example of a country that became steadily more democratic during the 20th century.

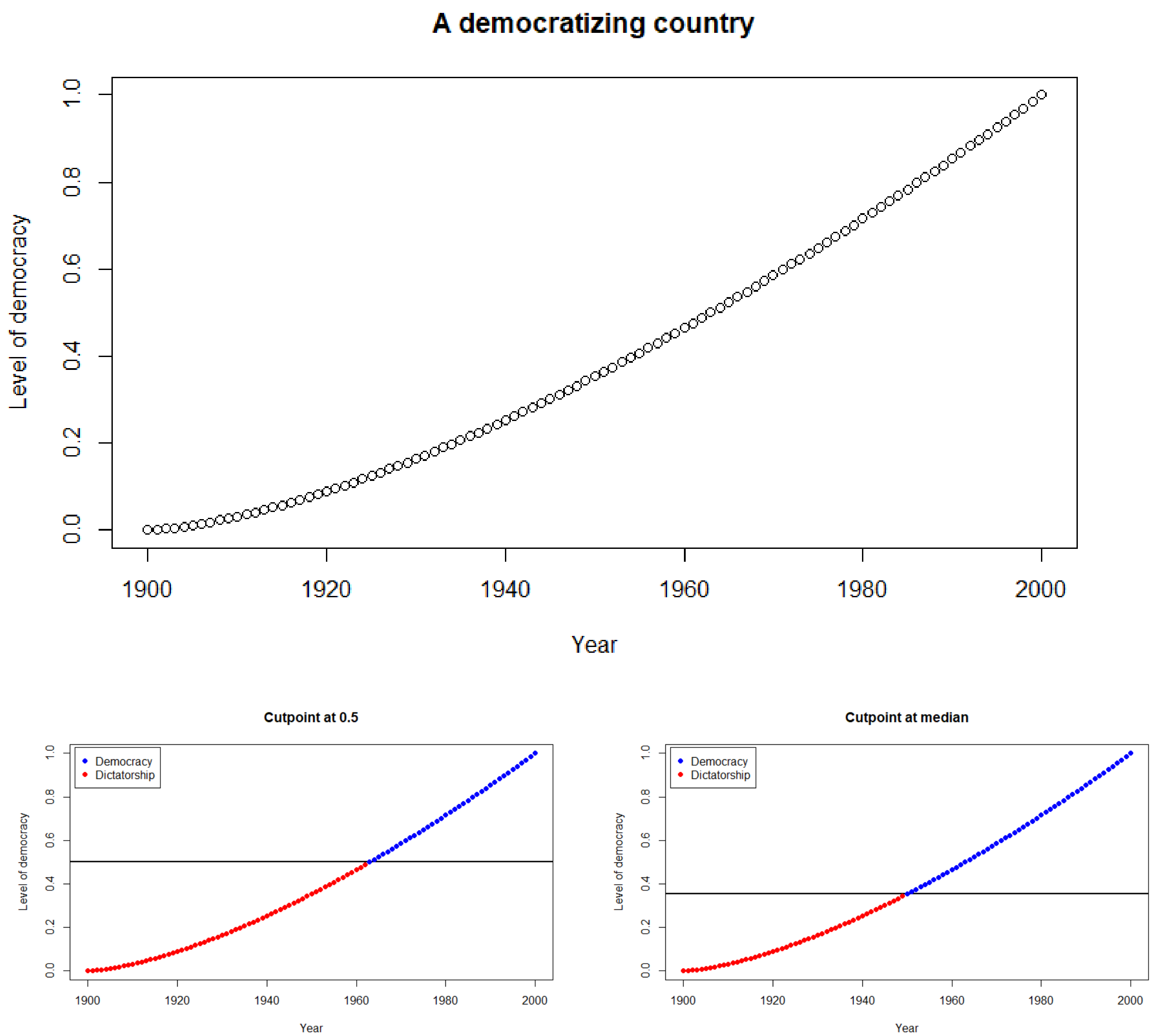

Figure 2: A fictional example to illustrate the idea of cutpoints.

The country in Figure 2 began as a full autocracy in 1900, and then became monotonically more democratic until it attained perfect democracy in the year 2000. The question is: in what year did it become a democracy? Is it when the country became "half democratic", as in the left panel in the bottom row of Figure 2? Or when the country crossed its median democracy value, as in the right panel in the bottom row of Figure 2? We could draw the line at – or even in between – any of the 101 points in Figure 2.

For a measure of democracy $\mathbf{x}$ that is bounded above some lower bound x_L and below some upper bound x_H, a "cutpoint" $\tau \in [x_L; x_H]$ is a value that translates $\mathbf{x}$ into a binary classification $\mathbf{b}$ of democracies ($b = 1$) and autocracies ($b = 0$) according to the following rule:

$$b = \begin{cases} 1 & \text{if } x \geq \tau \\ 0 & \text{if } x < \tau \end{cases}$$

where b is the element of $\mathbf{b}$ corresponding to some $x \in \mathbf{x}$.[2] This classification idea makes two claims. First, it claims that if a value $x \in \mathbf{x}$ is a democracy, then x' is also a democracy $\forall x' > x$, and conversely that if $x \in \mathbf{x}$ is an autocracy, then x' is also an autocracy $\forall x' < x$. This is easy to accept whenever x is intended to represent an amount of democracy, so that higher x-values represent a more democratic case. Second, the classification also makes the much stronger assertion that τ is the least democratic democracy. The difficult question in this method is therefore to decide where exactly to set τ.

In a vacuum, picking a cutpoint requires making a contentious (and perhaps even incoherent) substantive decision: we have to decide what proportion of the total quantity of possible democracy is a sufficient amount of democracy to be the minimum possible democracy. A common practice is to embrace that the exercise is unavoidably arbitrary, and pick, say, the middle of the range of $\mathbf{x}$ (Lührmann et al. 2018). But the fact that we have many different measures of democracy offers a method for picking a cutpoint that is grounded in the literature: as Kasuya and Mori (2019) argue, we can pick the cutpoint which best matches a many-valued measurement to a binary measurement as closely as possible. Anybody who has classified countries into either democracies or autocracies has already made a substantive judgment about what is the minimum requirement for a country to be a democracy, so this

[2] According to this rule, when the elements of $\mathbf{x}$ have a precision of $n \in \mathbb{N}$ digits, τ is actually the equivalence class of real numbers that have an identical first n digits. When we refer to "the optimal cutpoint" in this paper, what we really mean is "the optimal cutpoint of precision n".

approach translates the many-valued measure to a binary measure in whichever way matches another researcher's judgments as well as possible.

Consider the situation in Figure 3. This plot imagines a binary classification for the same burgeoning democracy in Figure 2, produced by a separate author, according to another rule for what constitutes a democracy.

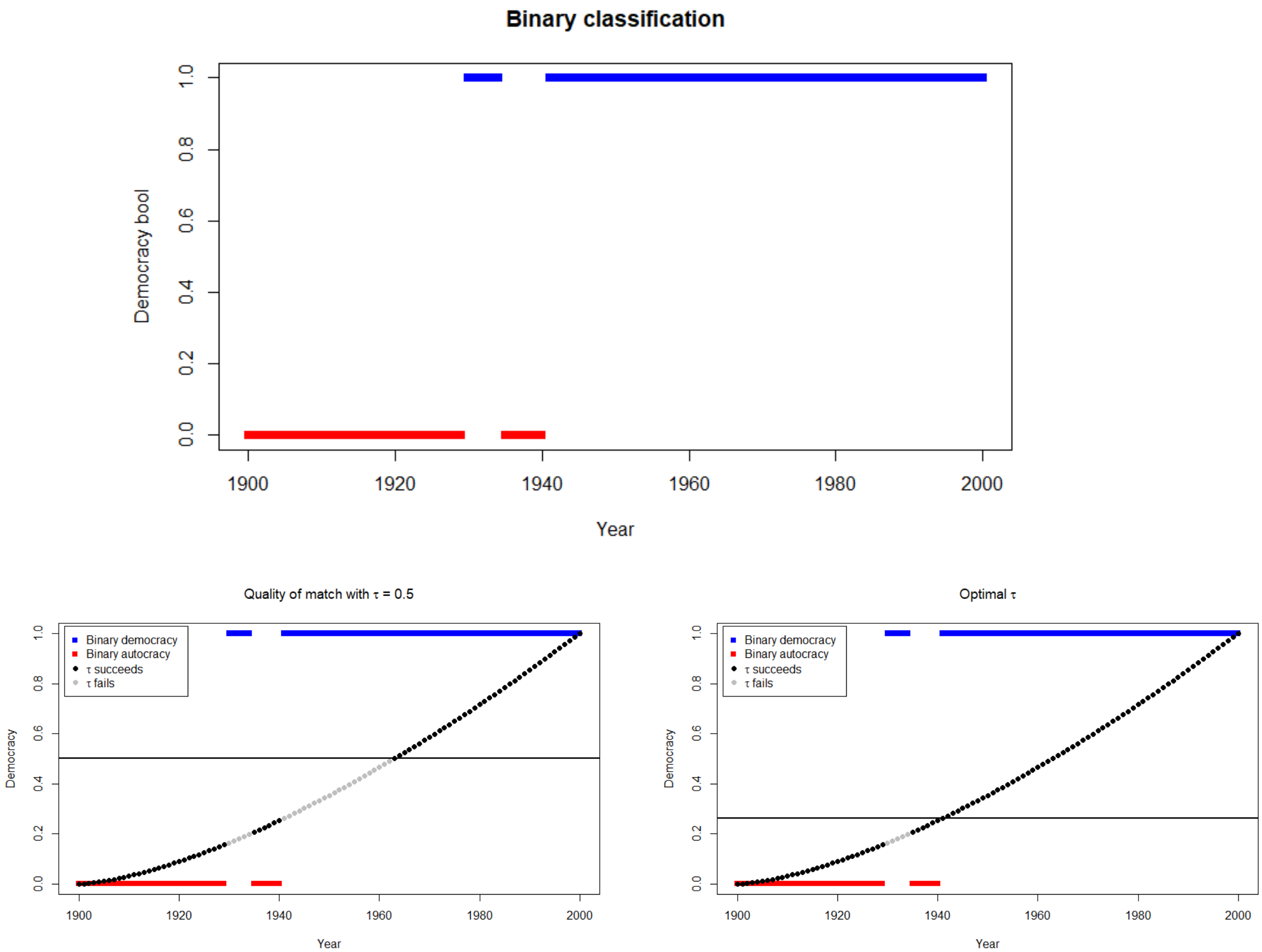

Figure 3: A fictional example of the optimal cutpoint for matching a many-valued classification to a binary classification.

This author has made their own substantive judgments about what exactly is the minimum requirement for a democracy; we can now take advantage of that existing judgment by setting a cutpoint so that our many-valued measure of democracy from Figure 2 matches the binary measure of democracy in Figure 3 as closely as possible. The bottom-left panel of Figure 3 shows that a naive cutpoint set at $\tau = 0.5$ classifies many cases differently from the binary classification. In the bottom-right panel, we identify the "optimal cutpoint", which we denote τ^*. This is the point at which increasing or decreasing the value of τ^* will not increase the number of cases that match the binary dataset.[3]

An important feature of optimal cutpoints is that they are almost never perfect. The binary classification in Figure 3 illustrates a common situation in which a country becomes democratic, then backslides into autocracy, and then returns to democracy. We can never capture this feature of the binary variable by applying a cutpoint to the many-valued classification in Figure 3, which instead argues that the level of democracy in our fictional country rose in every year. So in some situations τ^* might classify every case in the same way as the binary dataset, but in other situations τ^* might only do so for, say, 90% of cases.

But behind this simple picture is a far thornier problem: the extreme diversity in how authors operationalize democracy suggests that if we identify the best cutpoint for matching a many-valued measure of democracy to several different binary measures of democracy, we should expect to find a series of very different optimal cutpoints. As Kasuya and Mori (2019) explain, the researcher is then faced with a difficult decision about which cutpoint to use for their analysis. We will now demonstrate a highly convenient fact: the literature on democratic classifications, which includes many different notions of what exactly constitutes a democracy, consistently agrees about the best cutpoint for dichotomizing the main many-valued measures of democracy.

2.2 Different datasets have the same cutpoint

In this section we find the optimal cutpoint for matching V-Dem's polyarchy variable and the Polity Project's polity variable onto each binary classification of democracy, and we argue that it represents a new form of substantive agreement between the datasets. Figure 4 plots

[3] Note that multiple cutpoints could perform equally well, so τ^* is not necessarily unique.

the pairwise agreement between each dichotomous dataset and polyarchy, when polyarchy is dichotomized at every possible cutpoint. Figure 5 shows the same information for polity.

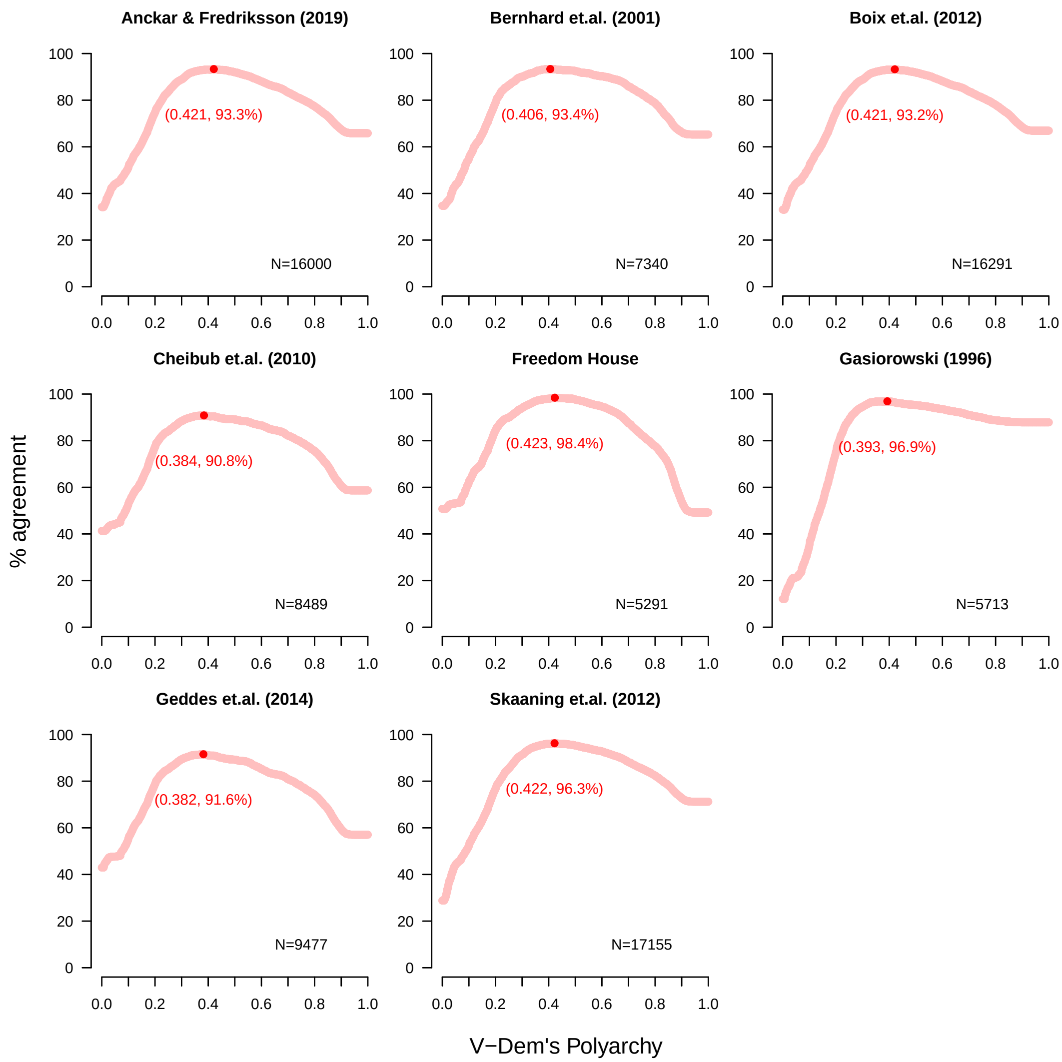

Figure 4: Each panel shows how successful every possible polyarchy cutpoint is at matching a dichotomous dataset. The dot represents the optimal cutpoint and the corresponding agreement level. In the Appendix we vary certain coding decisions.

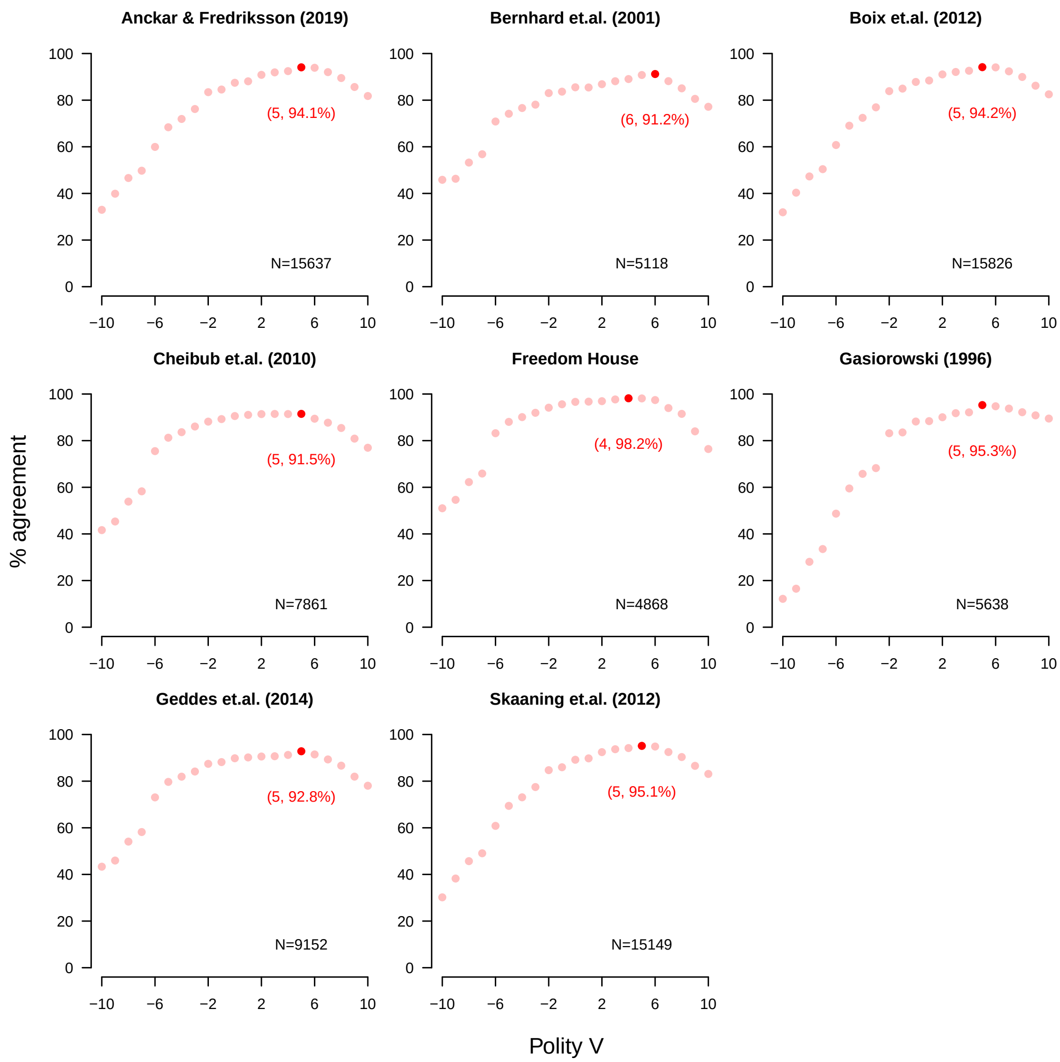

Figure 5: Optimal polity cutpoints.

Figure 4 plots the quality of the match between polyarchy and each binary dataset when τ takes on every possible value in the range of polyarchy, and Figure 5 shows the same data for polity. Figure 4 presents a remarkable regularity: multiple extremely different datasets, with radically different cases and coding methodologies, share the same optimal polyarchy cutpoint. The optimal polyarchy cutpoints for all of the dichotomous measures are in deep agreement with one another, differing by only 0.041, which is only 4% of the total range of polyarchy. We find exactly identical optimal cutpoints between Anckar and Fredriksson (2019) and Boix et al. (2012), and we are exceptionally close to finding that half of the datasets have the same optimal cutpoint: the optimal cutpoint shared by Anckar and Fredriksson (2019) and Boix et al. (2012) is just 0.001 less than the optimal cutpoint of Skaaning et al. (2015), which in turn is just 0.001 less than the optimal cutpoint for Freedom House. A difference of 0.001 is as similar as cutpoints can possibly be without being exactly identical. And the remaining three datasets are also tightly grouped, with optimal cutpoints ranging from 0.382 to 0.393. In this group, Geddes et al. (2014) and Cheibub et al. (2010) almost exactly agree with one another, with optimal cutpoints that differ only by 0.002.

Figure 5 shows a similar result. The datasets also closely agree about how to optimally dichotomize polity. Six of the eight datasets have the same optimal polity cutpoint of $\tau^* = 5$, while the other two datasets have adjacent values of $\tau^* = 4$ and $\tau^* = 6$.[4] Another identical cutpoint is mentioned by Pemstein et al. (2010, 444), who found that the original Przeworski et al. (2000) coding (from which some of these datasets are descended) has an optimal cutpoint "somewhere near 5 on the polity scale".

It is important to remember, though, that polity has only 21 values; up to the precision available in polity, we find the same tight similarity as in polyarchy, but the finding that polyarchy values are often identical is all the more surprising since it has 1001 possible values.

It is intriguing enough that the values of τ^* are always identical or nearly identical. But the figures show an even stronger result. Notice that most of the panels in those figures have the same characteristic curve (it is easier to see in Figure 4, but it is also true of Figure 5): very low cutpoints are similarly unsuccessful at matching the many-valued measure to each dichotomous measure, the quality of the match becomes sharply better as we choose cutpoints closer to the middle of the range, and then the quality of the match decreases much more gradually as we pick cutpoint values towards the top of the range. The fact that nearly all of these curves look similar means that every cutpoint is similarly successful

[4] We also identified the optimal cutpoint for matching polyarchy and polity onto a very different dataset: Svolik (2008), which classifies consolidated versus transitional democracies rather than democracies versus autocracies. Amazingly, the division between consolidated and transitional democracies in that dataset has the same optimal polyarchy cutpoint as Skaaning et al. (2015): $\tau^* = 0.422$, and its polity cutpoint of 4 matches that of Freedom House. These are the only global classifications for which we identified optimal cutpoints; we also identified cutpoints for two regional classifications (see the Appendix).

at mimicking nearly all of the eight dichotomous datasets. So, a cutpoint that matches polyarchy (or polity) to one binary classification with about a 50% success rate will also match polyarchy (or polity) to another dataset with about a 50% success rate.

As strong as the similarity in the optimal cutpoints is, this is an even more intriguing result: not only do the best cutpoints have similar values and similar success rates, but all of the possible cutpoints are similarly successful. As we have discussed, cutpoints add information that pairwise comparisons cannot address, so this finding builds on the conventional wisdom that these datasets tend to be fairly highly correlated with one another.

We have checked the best cutpoints for matching polyarchy to each binary dataset, and we separately did the same for polity. But can the cutpoints method also give us insights into how closely polity and polyarchy relate to each other? So far we have only discussed how to use cutpoints to match a many-valued variable onto a dichotomous variable. But we could also use them to match a many-valued variable onto another many-valued variable. A natural way to match polyarchy onto polity using cutpoints would be to proceed in two steps:

1. Turn polity into a binary variable

2. Find the optimal cutpoint for matching polyarchy onto that binary variable

In Table 1, we show the optimal cutpoint for matching polyarchy onto every possible polity dichotomization; since polity ranges from -10 to 10, there are 19 possible ways to dichotomize polity (excluding the bounds), each with a corresponding τ^*.

Polity dichotomization	Polyarchy τ^*	% Agree
-9	0.04	93%
-8	0.04	88%
-7	0.04	85%
-6	0.18	80%
-5	0.20	81%
-4	0.21	82%
-3	0.24	82%
-2	0.29	85%
-1	0.32	85%
0	0.34	87%
1	0.34	87%
2	0.34	89%
3	0.40	91%
4	0.40	91%
5	0.42	93%
6	0.45	93%
7	0.47	92%
8	0.69	91%
9	0.79	92%

Table 1: Optimal V-Dem cutpoints for matching onto the binary datasets that are created when we take every possible polity dichotomization (excluding the bounds, which would turn it into all ones or all zeroes).

One strategy for dichotomizing polity might be to pick the modal result of polity's optimal cutpoint for matching the binary datasets, since that will produce the binary classification that is the most similar to the literature. In Figure 5, polity's optimal cutpoint was found to be in the range from around 4 to 6. Conveniently, 4 is also the value that has been suggested by Marshall et al. (2017) as a potential democratic cutpoint. And once again, we are greeted with the same recurring optimal cutpoint value: when polity is dichotomized in the way that best matches the literature, the optimal polyarchy cutpoint for matching to that binary dataset is almost identical to the recurring optimal cutpoints in Figure 4. And what is more, when we dichotomize polity in the way that most closely agrees with other parts of the literature, we are able to match polyarchy to it better than if we dichotomize it in any other way.

This result ties the literature together even more strongly than our previous findings: the very best polyarchy cutpoint for matching to any possible dichotomization of polity happens to a) correspond to the optimal polity threshold for matching to the binary datasets, b) correspond to the recommended polity cutpoint, and c) also be the best polyarchy cutpoint for matching to the binary datasets. For all of these things to be true, there must be an exceptional level of agreement between the binary datasets, polyarchy, and polity.

2.3 Similar cutpoints represent substantive agreement

We have shown that different datasets have similar optimal cutpoints, even in cases where they could have more than a thousand different values. When we dichotomized polyarchy or polity in the way that most closely matched eight different dichotomous datasets, we found identical or nearly identical cutpoints every time. We emphasize that what is surprising is *not* the specific value of the cutpoints that we found; it would have been exactly as surprising to repeatedly find optimal cutpoints of, say, 0.3 or 0.6. What is surprising is that we found the same cutpoint values *over and over again* in very different places.

Out of 1001 possible cutpoint values (and 21 in the case of polity), we found nearly or exactly identical cutpoints between eight dichotomous datasets that have:

- ↪ **Variation in authors**: The datasets were written by different groups of authors, sometimes decades apart
- ↪ **Variation in cases**: The datasets have such radically different temporal and spatial coverage that as many as four out of every five country-years in one dataset are not even included in another dataset
- ↪ **Empirical disagreements**: The datasets disagree about as many as one in every six cases
- ↪ **Theoretical disagreements**: Every one of the datasets introduced innovations in substance or measurement specifically designed to make its conception of democracy different from previous classifications
- ↪ **Different definitions of democracy**: The datasets vary on such fundamental issues as whether or not suffrage is a necessary feature of democracies

↪ **Different purposes**: Some of the datasets were constructed with specific research questions in mind while others were constructed for more general use

↪ **Variation in coding methods**: Some of the datasets are coded by hand according to explicit rules, others are coded by aggregating scores assigned by experts or summarizing discussions among experts, and still others are the outcome of statistical models

It is therefore anything but obvious that the cutpoint which maximizes agreement between polyarchy and one of these datasets should also maximize agreement between polyarchy and another dataset. And yet, that is exactly what we found.

But is it possible that the similarity of the cutpoints, while unexpected, does not tell us anything substantive? We consider three possible objections to our claim that the similarity of the cutpoints represents an unexpectedly strong consensus in the field.

One natural concern is that perhaps there is some methodological reason that this particular cutpoint will keep appearing over and over again. But we already saw strong evidence against this, when we found nearly identical polity cutpoints in addition to the nearly identical polyarchy cutpoints. That suggests that the pattern is not just a special feature of one of those variables, or an unlikely random event. In the Appendix we go further, by running a series of simulations which show that the similarity of the cutpoints is not caused by the tool itself, or by the distributions of polyarchy or polity, or by some meaningless feature of the dichotomous variables. This establishes that their sameness really must represent some level of substantive agreement about democracy.

Another possible objection is that we have placed emphasis on the cutpoints coinciding up to 1, 2, or even 3 decimal places, but it is far from obvious that the third decimal place of a measure of democracy has an interpretable meaning. It is quite reasonable to argue that we should not focus too much on the third digit of polyarchy, and we invite the reader to completely ignore the third digits of our results. Then, instead of having two pairs of matching cutpoints, we find that fully *half* of the eight datasets have matching cutpoints: when rounded to the nearest two digits, Anckar and Fredriksson (2019), Boix et al. (2012), Skaaning et al. (2015), and Freedom House all have an optimal cutpoint of 0.42. So the cutpoints are just as interesting for a reader who only considers two decimal places as they are to a reader who considers all three. Indeed, in the Replications section we will take the consensus cutpoint to be 0.42, rather than splitting hairs about whether it is 0.421 or 0.422 or 0.423.

One last objection might be that we have taken no account of the fact that we are identifying optimal cutpoints for datasets that develop very different definitions and operationalizations of democracy, and we do not focus on any differences in coding rules. This is certainly true: the differences in how these authors conceptualize democracy *is exactly what makes the similarity of the cutpoints so remarkable*. The only commonality we insist on is that they are all classifying a level of democracy.

An essential caveat is that nothing about this exercise implies that one classification is more correct than another. We should not be understood as claiming that many-valued classifications are superior to dichotomous ones or *vice versa*, or that one dataset is closer than another to some objective ground truth of what democracy is, or that any differences between datasets are examples of one author being right and another author being wrong. There are major differences between different classifications, and many insightful papers have been written about the pros and cons of their various approaches. But as real as those differences surely are, we have shown that the agreement between datasets runs far deeper than conventional wisdom suggests.

3 Replications

We have argued that tremendously similar cutpoints represent surprisingly strong agreement about what constitutes a democracy. They do so by allowing us to boil either polyarchy or polity down to a dichotomous classification of democracies that is exceptionally similar to the major existing dichotomous classifications. But is this agreement, in the words of Boix et al. (2012), "merely academic"?

In the same spirit of meta-analysis in which we motivated this investigation, we now pick central empirical results from two papers that introduced very different and important dichotomous measures, Boix et al. (2012) and Cheibub et al. (2010), and we show that these core results can be fully replicated with polyarchy dichotomized using the optimal cutpoint of $\tau^* = 0.42$.

3.1 Replicating Boix et al. (2012)

The primary purpose of Boix et al. (2012) is to present a dataset, but together with that dataset the authors include descriptive plots of the association between their democratic measure and three core variables: GDP per capita, land equality, and latitude. They find that "economic modernization variables have steadily declined in their correlation with democracy over time".

In Figure 6 we replicate each of those main plots. In the interest of an independent replication, we use the GDP per capita value from Coppedge et al. (2019) rather than the one used in Boix et al. (2012), and rather than the land equality value used in Boix et al. (2012) we use the (distinct but closely related) urbanisation value from Coppedge et al. (2019), but there is no similar reason to use a different latitude variable.[5] In Figure 6, red squares show the values when we re-implement the analysis in Boix et al. (2012) and run it on their democracy classification; blue circles show the values when we run that same analysis on polyarchy dichotomized at $\tau^* = 0.42$.

[5] These changes in variables mean that the precise numbers we report sometimes differ in substantively unimportant ways from the numbers that the original authors report. We also wrote our own code based on the authors' descriptions of their analyses.

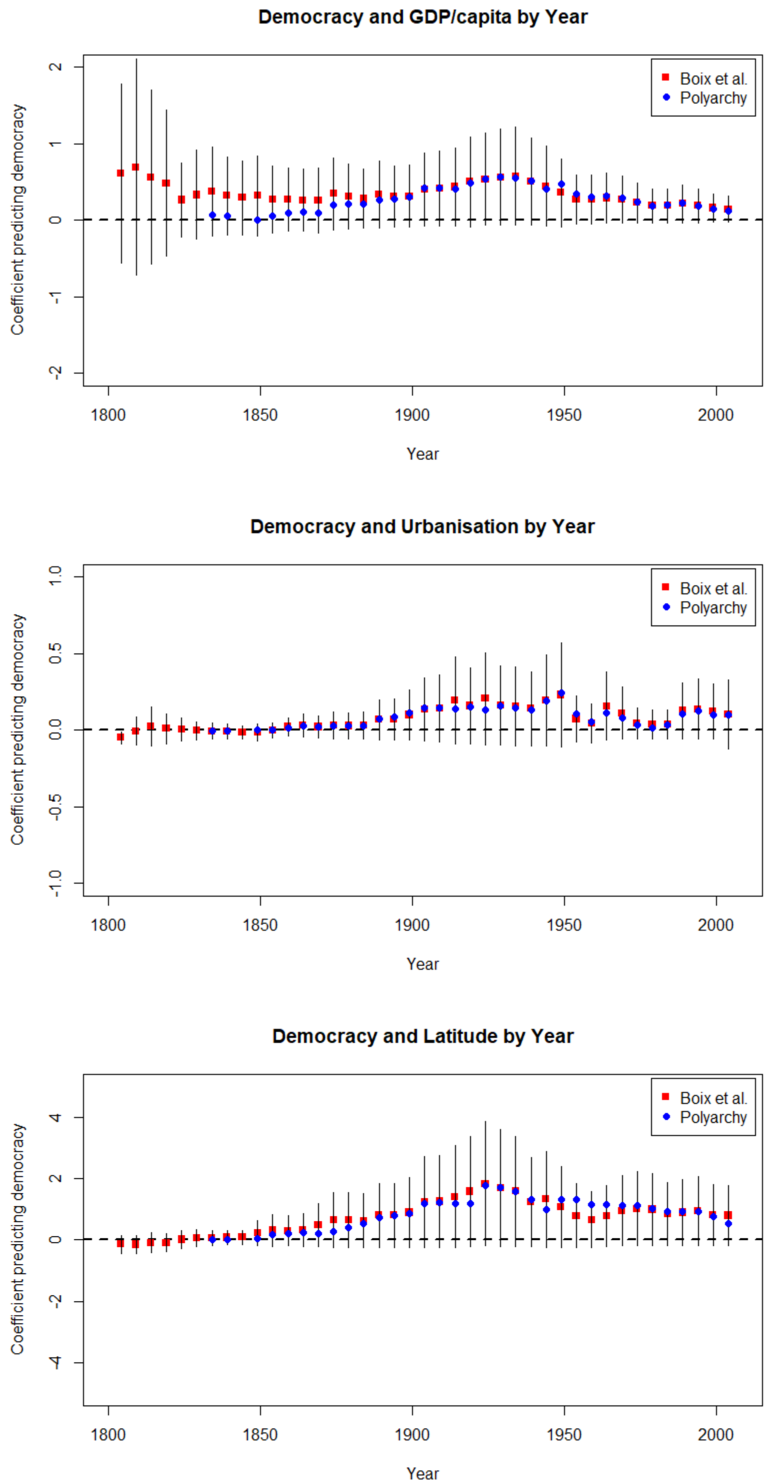

Figure 6: Replication of (respectively, from top to bottom) Figure 3, Figure 4, and Figure 5 in Boix et al. (2012), using their democracy coding and polyarchy dichotomized at $\tau^* = 0.42$. The years that do not appear have too few observations.

We very closely replicate the pattern in Boix et al. (2012) that economic variables have declined in their capacity to predict level of democracy; the polyarchy dataset would give the same conclusion. Urbanization behaves slightly differently from land equality, but in ways which do not affect the overall pattern of the results, and here polyarchy and the democracy variable of Boix et al. (2012) behave almost identically. Latitude also very closely resembles the figure in Boix et al. (2012), up to units, and the coefficients of the dichotomized polyarchy variable depart only slightly from Boix's classification. In every case, the coefficient estimate for the dichotomized polyarchy variable is within the confidence interval of the coefficient estimate for the variable from Boix et al. (2012), and would yield the same substantive conclusions.

3.2 Replicating Cheibub et al. (2010)

The presentation of the dataset in Cheibub et al. (2010) is organized around a series of replications of past work, to illustrate theoretical subtleties of their measure of democracy with respect to previous measures. The first of these is a replication of a main result in Fearon and Laitin (2003). In Table 2 we show that the dichotomized polyarchy dataset behaves exactly as the dataset in Cheibub et al. (2010) does: the same variables are significant, with the same signs and nearly identical magnitudes. The consensus classification is exactly as successful as the Cheibub et al. (2010) dataset in replicating the results of Fearon and Laitin (2003).

	Fearon and Laitin (2003)	Cheibub et al. (2010)	Polyarchy replication
Anocracy	**0.54**		
	(0.24)		
Democracy	0.11		
	(0.31)		
Dictatorship with legislature		-0.36	-0.40
		(0.21)	(0.21)
Instability	**0.53**		
	(0.24)		
Prior war	**-0.84**	**-0.85**	**-0.87**
	(0.31)	(0.32)	(0.32)
GDP/capita	**-0.31**	**-0.34**	**-0.35**
	(0.07)	(0.07)	(0.07)
log(population)	**0.27**	**0.25**	**0.26**
	(0.07)	(0.07)	(0.07)
log(% mountainous)	**0.20**	**0.24**	**0.25**
	(0.09)	(0.09)	(0.09)
Noncontinuous state	0.33	0.37	0.38
	(0.28)	(0.28)	(0.28)
Oil exporter	**0.79**	**0.93**	**0.92**
	(0.28)	(0.27)	(0.27)
New state	**1.63**	**1.52**	**1.59**
	(0.35)	(0.34)	(0.34)
Ethnic fractionalization	0.15	0.10	0.11
	(0.37)	(0.37)	(0.37)
Religious fractionalization	0.43	0.52	0.48
	(0.51)	(0.51)	(0.51)
Constant	**-7.09**	**-6.48**	**-6.48**
	(0.76)	(0.73)	(0.73)
N	6,217	6,255	6,255

Table 2: Cheibub (2010) Table 2 replication. Bolded numbers are significant at $p < 0.05$.

Since we saw in section 2.2 that the polyarchy dichotomization is highly correlated with the Boix et al. (2012) and Cheibub et al. (2010) datasets, one might have anticipated that the authors' findings can be replicated with the dichotomized polyarchy variable. But if this is expected after reading the cutpoint analysis in section 2.2, it was certainly not obvious at the outset that we could take a variable from one dataset, dichotomize it in just one way, and replicate findings from two different datasets (Högström 2013).

We emphatically are not arguing that all binary codings of democracy are redundant, or that a dichotomized version of polyarchy can replace any binary coding of democracy in any analysis and produce the same results. To the contrary, in the Appendix we find mixed success when we deliberately push the limits of dichotomized polyarchy as a replication tool, by attempting to replicate analyses from a very different coding of democracies: Svolik (2008). Our point was to establish just how deep the consensus in the field actually runs, in direct contradiction of how it is usually portrayed.

With these replications, we have shown that lying latent in the field is a consensus classification that not only matches existing datasets exceptionally well, but actually reproduces the core findings that motivated some of the major previous datasets. We will now dig into that consensus: what exactly is this understanding of democracy that the field has converged on?

4 What is the consensus?

Knowing that there is a consensus is half the story: we now probe what exactly that consensus is. We proceed in three steps. First, we examine how much the dichotomous datasets agree with one another, and what drives their disagreements. Second, we dichotomize polyarchy at the optimal cutpoint, and we explore how the resulting dataset compares to the majority positions of the dichotomous datasets. Third, we briefly discuss the nature of polyarchy dichotomized at the optimal cutpoint, to better understand the characteristics of the consensus that we have identified in this paper.

4.1 Consensus among binary classifications

For the sake of clarity, in this section we select a smaller number of datasets to focus on. We will focus on polyarchy rather than repeating the following analysis for polity (but the high correlation in Table 1 suggests an analysis with polity would get similar results), and on just four binary classifications. As shown in Table 3, these represent the range of coding rules and practices among democratic classifications.

Dataset	Requirements for full democracy
Boix et al. (2012)	All of: ↪Executive elected in popular elections ↪Executive responsible to voters or legislature ↪Legislature chosen in free and fair elections ↪Most adult men have voting rights
Cheibub et al. (2010)	All of: ↪Free and fair elections ↪Minimum level of suffrage
Geddes et al. (2014)	None of: ↪Executive in power without direct, fair, competitive election ↪Rule change limiting electoral competition ↪Major parties blocked from elections by military
Puddington et al. (2018) (Freedom House)	Expert survey with discussion and review

Table 3: The varying rules for coding a country as a democracy.

We saw, in Figure 1, that these cases typically have fairly strong pairwise agreement. So in the rare cases where they disagree, what drives those disagreements? Of the 199 distinct countries that appear in at least one of these four binary datasets, 108 of them (a clear majority) have no year in which one author considers them democratic while another author considers them autocratic. Among the countries that have a disagreement, there are a total of 721 individual disagreements, so among these countries the average number of years in dispute is less than 7. Only 24 countries (12%) have more than 10 years coded as a democracy in at least one dataset while being coded as an autocracy in another. Table 4 records these countries and the years in question.[6]

Country	Number of disagreements	Years
Botswana	45	1966 – 2010
Guatemala	32	1954, 1958 – 1962, 1966 – 1981, 1986 – 1995
Lebanon	32	1944 – 1970, 1975, 1976, 2006 – 2008
Costa Rica	30	1920 – 1949
The Netherlands	26	1871 – 1896
Fiji	25	1970 – 1986, 1992 – 1999
The Gambia	22	1972 – 1993
Namibia	21	1990 – 2010
Lesotho	20	1966 – 1970, 1994 – 2008
Argentina	19	Many individual years
Armenia	18	1991 – 2008
Guyana	18	1972, 1994 – 2008
Ghana	16	Many individual years
El Salvador	15	1972 – 1975, 1984 – 1994
Pakistan	15	1947 – 1957, 1976, 1988, 1999, 2008
South Africa	15	1994 – 2008
Thailand	15	1975, 1976, 1979 – 1988, 1991, 1992, 2006
Paraguay	14	1989 – 2002
Haiti	12	1947 – 1950, 1991, 1995 – 1999, 2007, 2008
Nigeria	12	1966, 1979, 1983, 1999 – 2007
Peru	12	Many individual years
Sri Lanka	12	1977 – 1982, 1989 – 1994
Guinea-Bissau	11	1994 – 1997, 2000 – 2007
Panama	11	Many individual years

Table 4: Countries that have more than 10 years which were coded as a democracy in at least one dataset while being coded as an autocracy in at least one other dataset.

[6]Disagreements between countries can sometimes arise from coding decisions that might at first appear arbitrary, such as whether a country-year observation is coded based on its status at the start or end of a year. But every coding decision arises from substantive considerations. For example, one author may believe that the existence of a transition is sufficient to switch between democratic and autocratic, while another author is interested in spells that last a majority of the year. We leave these types of disagreements in the datasets for the same reason that we do not modify the datasets to match any of our other substantive preferences: we are interested in capturing consensus in the field as it exists.

4.2 Consensus between binary classifications and polyarchy

The binary classifications continue to show strong consensus with each other, but how well do they align with polyarchy? In Figure 7, for each polyarchy value, we plot how frequently the authors of the dichotomous datasets consider a country-year with that amount of polyarchy to be a democracy. We should expect an increasing trend, with most cases at low polyarchy values being classified as autocracies, and most cases at high polyarchy values being classified as democracies. The figure confirms this expectation.

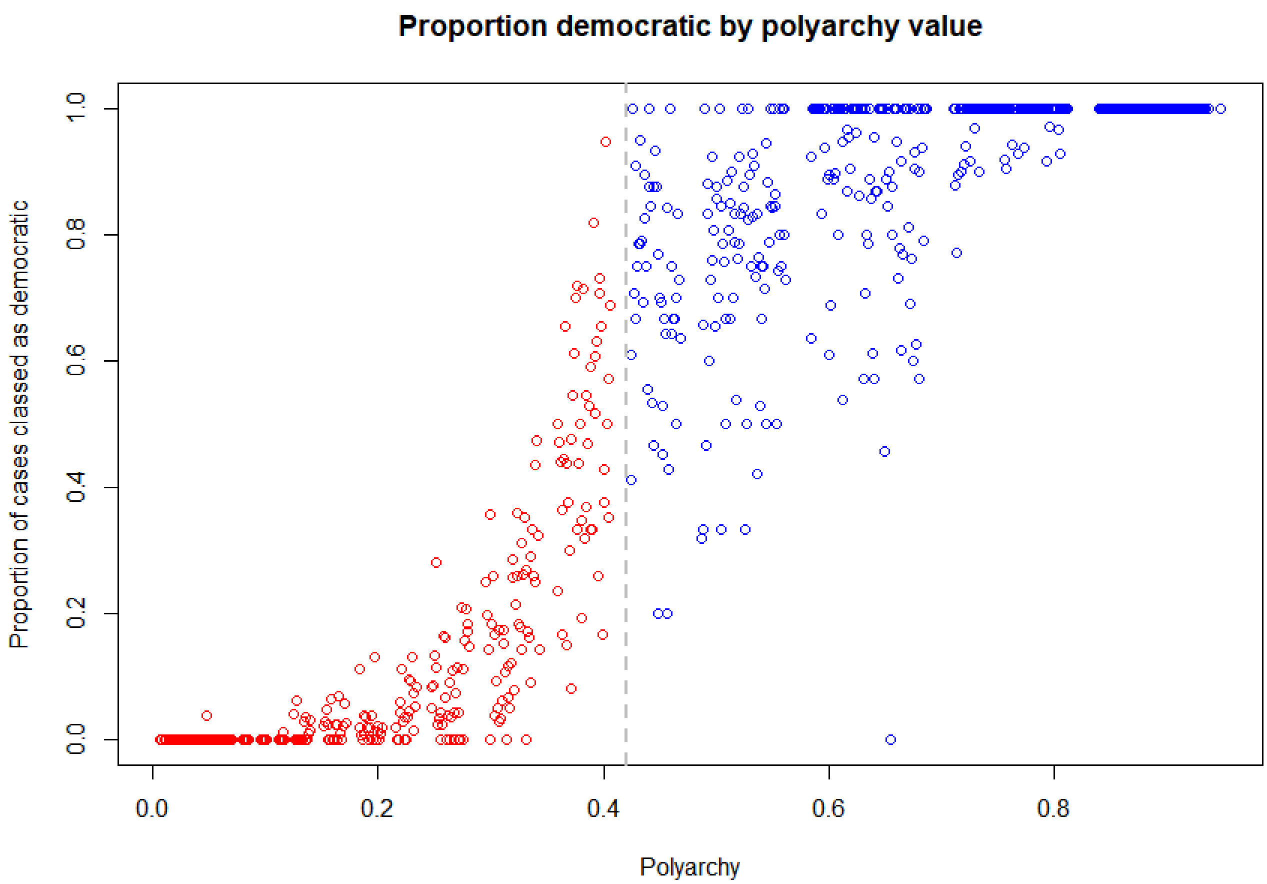

Figure 7: Proportion of cases in all datasets classified as democratic by polyarchy. The cases are coloured according to the cutpoint $\tau = 0.42$, which is marked with a vertical line.

To obtain the proportion for each polyarchy value in Figure 7, we first check the cases that appear in V-Dem with that polyarchy value. Then, we count the number of times that those cases were classified as a democracy in the dichotomous datasets. Finally, we divide that number by the total number of times that any observation at that polyarchy value was included in a binary dataset. The strong right-skew of the points in Figure 7 gives us a simple explanation for why the optimal cutpoint is around 0.42: binary coders are more willing to classify ambiguous cases as democracies than autocracies. When they encounter a case that is right on the line, they are more likely to bump it up than to bump it down.

Where does the cutpoints-based consensus differ from the majority of binary classifications?[7] $16,141$ country-years appear in V-Dem and at least one binary dataset, after discounting the 185 exactly tied country-years. Of these, 301 country-years are classified as autocracies in a majority of the binary datasets, but are considered by the cutpoint method to be democracies. 713 country-years are considered democracies by the binary datasets, but the cutpoint method classifies them as autocracies. So the cutpoint is more likely to classify as autocratic cases that most authors consider democratic, and it is less likely to classify as democratic cases that most authors consider autocratic. Similar to the pattern in Figure 7, when most binary datasets consider a case to be autocratic, it is consistently considered autocratic, and most potentially ambiguous cases are classified as democracies.[8]

When the cutpoint method considers a case to be a democracy but most binary datasets call it an autocracy, we will call it a "cutpoint-only democracy", and we call the reverse case a "cutpoint-only autocracy". Table 5 shows the ten countries with the most cutpoint-only democracies, and the ten countries with the most cutpoint-only autocracies.

[7] In the Appendix we break these patterns down to the level of country-years, and we discuss the advantages of the cutpoints method over just checking the majority judgment of the binary classifiers.

[8] Many of these are considered by the Regimes of the World dataset to be electoral autocracies (Lührmann et al. 2018).

Country	Cutpoint-only democracies	Cutpoint-only autocracies
Senegal	**20** years: 1979, 1982 – 1999	
Burkina Faso	**19** years: 1979, 1992 – 2010	
Namibia	**19** years: 1990 – 2008	
Zambia	**18** years: 1993 – 2010	
Mozambique	**17** years: 1994 – 2010	
Tanzania	**17** years: 1994 – 2010	
Cyprus	**16** years: 1961 – 1976	
Bosnia and Herzegovina	**13** years: 1997 – 2009	
Iraq	**13** years: 2006 – 2018	
Côte d'Ivoire	**12** years: 1997, 2001 – 2011	
United States of America		**96** years: 1800 – 1833, 1836 – 1897
Greece		**77** years between 1864 and 1974
Guatemala		**35** years between 1945 and 1992
Chile		**33** years between 1909 and 1953
Colombia		**29** years between 1937 and 1982
Lebanon		**25** years between 1946 and 1975
Pakistan		**20** years between 1950 and 1996
Panama		**19** years between 1949 and 1990
Norway		**17** years: 1886 – 1897, 1941 – 1945
Ecuador		**16** years: 1948 – 1962, 1979

Table 5: The 10 countries with the most cutpoint-only democracies or autocracies.

Cutpoint-only autocracies are countries that meet the minimalistic criteria used for most binary classifications of democracy, but have a polyarchy value under 0.42. The most frequent cutpoint-only autocracy is the United States from its founding through the late 1800s. There is a straightforward explanation for this disagreement: the early United States satisfies many of the simple requirements, like an elected executive and a minimum level of male suffrage, that were used to create most of the binary classifications (See Table 3), but it was also rife with egregiously non-democratic institutions that might lower its polyarchy value (Mickey 2015).

Cutpoint-only democracies are countries that do not meet the minimalistic criteria used for binary classifications of democracy, while having polyarchy of at least 0.42. This is a much rarer situation than cutpoint-only autocracies. Two particularly instructive cutpoint-only democracies are Namibia and Senegal.

From 1990 until 2008, Namibia is considered a democracy by the cutpoints method, but is considered an autocracy in most binary datasets. However, during these decades its polyarchy value ranges between 0.6 and 0.7, far exceeding the cutpoint of 0.42. Taken together with the prominence of Botswana in Table 4, one feature stands out as the likely reason that the binary classifications are divided on these cases: the turnover rule. Can a country be a democracy if it has never experienced a transition from one party to another? Authors who follow Przeworski et al. (2000) in answering "no" will classify Namibia and Botswana as autocratic, since they have been governed respectively by the SWAPO Party and the Botswana Democratic Party since independence; authors who answer "yes" may or may not code these cases as democratic. The appearance of Senegal supports this interpretation.

There is another noteworthy feature of the cutpoint-only democracies compared to the cutpoint-only autocracies: strong regional groupings. Of the ten cutpoint-only democracies listed in Table 5, seven are in Africa. V-Dem is substantially more likely to assign high polyarchy to African countries than the binary datasets are to classify them as democratic. In contrast, of the ten cutpoint-only autocracies, five are in Latin America, while three represent periods in the distant history of European or North American democracies.

We have seen that there is a surprisingly strong consensus on which countries are democracies and which are autocracies, and we have a sense of the scale and types of disagreements within this consensus. But what is it about democracies and autocracies that makes the 0.42 cutpoint substantively special? What sorts of countries have polyarchy above that cutpoint, and how are they different from the countries below that cutpoint? In the Appendix we discuss the dangers of hunting for substantive drivers of the optimal cutpoint after having already identified it; here we limit ourselves to descriptive statements.

The classification that represents a solid consensus of democratization scholars emphasizes the following traits. It firmly insists that democracies always have elections, but countries with elections can be autocracies. Elections in democracies might actually not be fully free or fair, while elections in autocracies are almost never free or fair. Almost all democracies have freedom of association and freedom of expression, and while many autocracies have no free association or expression, a small number of them do. Finally, democracies nearly always have full suffrage, and while historically many autocracies did not.

5 Conclusion

It would be easy to read the literature on classifying democracies and come away with the impression that political scientists have almost no agreement on one of the oldest and most important topics in our field. For decades we have known that measures of democracy are all highly correlated, which led to a completely justifiable and empirically important focus on why those datasets are not interchangeable. But that has given rise to the false impression that these datasets are rife with disputes and fundamentally at odds.

In this paper we have shown that there is actually very deep agreement on what constitutes a democracy. We addressed one of the major controversies in the literature — whether democracy is a matter of degree or of kind — by dichotomizing many-valued measures of democracy so that they match dichotomous measures of democracy as closely as possible. At the outset it seemed perfectly obvious that there would be no single best way to dichotomize a many-valued measure of democracy so that it matches datasets written by different authors, decades apart, about different institutions in different countries during different centuries, with sometimes conflicting definitions of democracy, coded through formal rulesets or discussions or predictive modeling, and which actively disagree about as many as one in every six cases. And yet, we found that there is one consistent best way to dichotomize multiple many-valued measures of democracy to match these wildly diverse datasets. And that matching is so similar to the existing datasets that it can actually replicate a series of core results in the study of democracies.

While showing that classifications of democracies agree superbly with each other, we have made no attempt to judge how close each of those classifications is to some objective definition of democracy. Nor have we passed judgment on whether some bright line separates democracies and autocracies. These two questions are the subjects of venerable and sophisticated debates, and those debates should and will continue. But they should continue with the awareness that, so far, we agree about much more than we realize.

Different people will have different definitions of democracy; some will be more similar to the consensus of the field, some more at odds with it. And that raises two disparate interpretations of our findings. On the one hand, perhaps our theoretical understanding of democracy is a great deal more harmonious than the conventional wisdom would suggest. So while differing classification rules have represented genuine theoretical disagreements, those disagreements only affect a small number of cases. On the other hand, there might be many tenable definitions of democracy that have yet to be used in a large cross-national classification. Perhaps there is some unseen reason that it is simply too difficult to construct a completely novel and radically different classification of democracy.

Skaaning et al. (2015) are far from alone in arguing that "the discipline is well served by a variety of measures". While the identical recurring cutpoints are a marker of how much classifiers of democracies actually agree, they are also a warning flag to anyone who sets out to construct a new measure of democracy. Even if a researcher develops a coding rule that they believe makes very important substantive departures from previous classifications, the result is nevertheless very likely to closely resemble the consensus classification.

References

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

Anckar, Carsten and Cecilia Fredriksson. 2019. "Classifying political regimes 1800–2016: a typology and a new dataset." *European Political Science* 18(1): 84–96.

Bernhard, Michael, Timothy Nordstrom, and Christopher Reenock. 2001. "Economic performance, institutional intermediation, and democratic survival." *Journal of Politics* 63(3): 775–803.

Boese, Vanessa A. 2019. "How (not) to measure democracy." *International Area Studies Review* 22(2).

Bogaards, Matthijs. 2010. "Measures of democratization: From degree to type to war." *Political Research Quarterly* 63(2): 475–488.

Boix, Carles, Michael Miller, and Sebastian Rosato. 2012. "A complete data set of political regimes, 1800-2007." *Comparative Political Studies* XX(X): 1–32.

Boix, Carles, Michael Miller, and Sebastian Rosato. 2018. "Boix-Miller-Rosato Dichotomous Coding of Democracy, 1800-2015."

Bollen, Kenneth and Robert Jackman. 1989. "Democracy, stability and dichotomies." *American Sociological Review* 54(4).

Bollen, Kenneth A. 1980. "Issues in the comparative measurement of political democracy." *American Sociological Review*: 370–390.

Bormann, Nils-Christian and Matt Golder. 2013. "Democratic electoral systems around the world, 1946–2011." *Electoral Studies* 32(2): 360–369.

Casper, Gretchen and Claudiu Tufis. 2003. “Correlation versus interchangeability: The limited robustness of empirical findings on democracy using highly correlated data sets.” *Political Analysis* 11(2): 196–203.

Cheibub, José Antonio, Jennifer Gandhi, and James Raymond Vreeland. 2010. “Democracy and dictatorship revisited.” *Public Choice* 143: 67–101.

Collier, David and Robert Adcock. 1999. “Democracy and dichotomies: A pragmatic approach to choices about concepts.” *Annual Review of Political Science* 2: 537–565.

Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Michael Bernhard, M. Steven Fish, Adam Glynn, Allen Hicken, Anna Lhrmann, Kyle L. Marquardt, Kelly McMann, Pamela Paxton, Daniel Pemstein, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Agnes Cornell, Lisa Gastaldi, Haakon Gjerlw, Valeriya Mechkova, Johannes von Rmer, Aksel Sundström, Eitan Tzelgov, Luca Uberti, Yi ting Wang, Tore Wig, and Daniel Ziblatt. 2019. “V-dem dataset.”

Cutright, Phillips. 1963. “National political development: Measurement and analysis.” *American Sociological Review*: 253–264.

Dahl, Robert Alan. 1973. *Polyarchy: Participation and opposition.* Yale University Press.

Fearon, James D. and David D. Laitin. 2003. “Ethnicity, insurgency, and civil war.” *American Political Science Review* 97(1): 75–90.

Gasiorowski, Mark J. 1996. “An overview of the political regime change dataset.” *Comparative Political Studies* 29(4): 469–483.

Geddes, Barbara. 1999. “What do we know about democratization after twenty years?” *Annual Review of Political Science* 2: 115–144.

Geddes, Barbara, Joseph Wright, and Erica Frantz. 2014. “Autocratic breakdown and regime transitions.” *Perspectives on Politics* 12(2).

Gleditsch, Kristian Skrede and Michael D. Ward. 2006. "Diffusion and the international context of democratization." *International organization*: 911–933.

Högström, John. 2013. "Does the choice of democracy measure matter? comparisons between the two leading democracy indices, freedom house and polity iv." *Government and Opposition* 48(2): 201–221.

Kasuya, Yuko and Kota Mori. 2019. "Better regime cutoffs for continuous democracy measures." *V-Dem working paper*.

Lerner, Daniel. 1958. *Democracy and development: political institutions and well-being in the world, 1950-1990.* Free Press.

Lipset, Seymour Martin. 1960. *Political Man: The Social Bases of Politics.* Anchor.

Lührmann, Anna, Marcus Tannenberg, and Staffan I. Lindberg. 2018. "Regimes of the world (row): Opening new avenues for the comparative study of political regimes." *Politics and Governance* 6: 60–77.

Marshall, Monty G., Ted Robert Gurr, and Keith Jaggers. 2017. "Polity iv project."

Mickey, Robert. 2015. *Paths Out of Dixie: The Democratization of Authoritarian Enclaves in America's Deep South, 1944-1972.* Princeton University Press.

Pemstein, Daniel, Stephen A. Meserve, and James Melton. 2010. "Democratic compromise: A latent variable analysis of ten measures of regime type." *Political Analysis* 18.

Popper, Karl R. 2020. *The open society and its enemies.* Princeton University Press.

Przeworski, Adam, Michael Alvarez, José Antonio Cheibub, and Fernando Limongi. 2000. *Democracy and development: political institutions and well-being in the world, 1950-1990.* Cambridge University Press.

Puddington, Arch, Jennifer Dunham, Elen Aghekyan, Christopher Brandt, Rukmani Bhatia, Isabel Linzer, Shannon O'Toole, Sarah Repucci, and Tyler Roylance. 2018. "Freedom in the world."

Schumpeter, Joseph A. 1942. "Capitalism, socialism, and democracy." *New York*.

Skaaning, Svend-Erik, John Gerring, and Henrikas Bartusevičius. 2015. "A lexical index of electoral democracy." *Comparative Political Studies* 48(12): 1491–1525.

Svolik, Milan. 2008. "Authoritarian reversals and democratic consolidation." *American Political Science Review* 102(2): 153–168.

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.

Wahman, Michael, Jan Teorell, and Axel Hadenius. 2013. "Authoritarian regime types revisited: updated data in comparative perspective." *Contemporary Politics* 19.