

# A Measurement Error Model of Democracy Status

Jay Ulfelder and Sean J. Taylor

February 6, 2015

## **Abstract**

We use a Bayesian measurement error model to derive a probabilistic measure of democracy from several existing dichotomous data sets. This approach accepts the premise that democracy may usefully be construed as a bivalent concept for certain theoretical and technical purposes, but it also makes more explicit our collective uncertainty about where some cases fall in that binary scheme. We believe the resulting data provide a firmer foundation than measures of countries' degree of democracy for studies that require a dichotomous measure of regime type.

***Keywords:*** democracy, measurement, Bayesian

# 1 Introduction

There is nothing worse than a  
sharp image of a fuzzy concept.

---

Ansel Adams, *Reader's Digest*

(May 1997)

Political scientists use the term “regime type” to refer to the formal and informal structure of a country’s government—more specifically, how rulers are selected and how their authority is organized and exercised. Over the past 25 years or so, scholars have devoted particular attention to measuring democracy as a regime type. The widely used Polity IV data set treats democracy as a scalar concept, but several other projects assert that democracy may usefully be construed as a bivalent concept and therefore adopt a dichotomous approach to measuring it (e.g., Boix et al. (2012)[1], Cheibub et al. (2010)[3], Goldtsone et al. (2010)[9], and Ulfelder (2012)[14]). These dichotomous measures and ones like them have been used in a number of influential studies on regime survival and other things that a country’s regime type is expected to affect.

Unfortunately, those dichotomous data sets also obscure significant uncertainty about whether or not certain cases qualify as democracies. Specific countries at specific times are identified as falling on one side or the other of that line, even when researchers knowledgeable about those cases might be unsure or disagree about where they belong. For example, all of the existing data sets concur that Norway is a democracy and Saudi Arabia is not, but they diverge on the classification of countries like Russia, Venezuela, and Pakistan, and rightly so. Some of the features we need to measure to make judgments about whether or not a regime is democratic—for example, freedoms of speech, assembly, and association—are inherently hard to observe, and it’s not exactly clear how much is enough to qualify. Other features, such as electoral fraud, are hard to observe because political actors respond to strong incentives to hide them.

These measurement challenges don't negate the validity of a bivalent concept of democracy, but they do mean that our judgments about whether or not that concept is present in specific cases will often be uncertain.

As it happens, the degree of our uncertainty about where a case belongs may itself be correlated with many of the things that researchers use data on regime type to study. As a result, findings and forecasts derived from those data are likely to be sensitive to those bivalent calls in ways that are hard to understand when that uncertainty is ignored. In principle, it should be possible to make that uncertainty explicit by reporting the probability that a case belongs in a specific set instead of making a firm yes/no decision, but that's not what the data sets we have now do.

To address this shortcoming, we use a Bayesian measurement error model to derive a probabilistic measure of democracy from several existing dichotomous data sets. This approach accepts the premise that democracy may usefully be construed as a bivalent concept for certain theoretical and technical purposes, but it makes explicit our uncertainty about where some cases fit in that either/or scheme. Instead of 0s and 1s, we end up with estimated probabilities of membership in the democracy set and confidence intervals associated with those estimates. We don't create new information about the features of countries' political regimes. Instead, we draw on previously created information to newly quantify our collective uncertainty about the presence or absence of democracy. We believe the resulting data can serve as a firmer foundation for studies that depend on a dichotomous measure of this particular regime type.

Pemstein et al. (2010)[13] use the same technique to produce the Unified Democracy Scores. Our method is very similar, but the underlying concept—and thus the results—differ in an important way. Their approach treats democracy as a matter of degree and derives a composite scalar index of democracy from several measures. By contrast, we accept a binary conceptualization of democracy and estimate the probability that a country belongs in that set.

In fuzzy set theory (Zadeh, 1965)[17], this probability represents a case’s degree of membership in the democracy set, not how democratic it is. The distinction between a country’s degree of membership in that set and its degree of democracy is subtle but meaningful, and the former will sometimes be a better fit for an analytic task than the latter. As Collier and Adcock (1999)[4] argue, “How scholars understand and operationalize [democracy] can and should depend in part on what they are going to do with it.” In other words, the choice of measures depends, in part, on the aims and design of the ensuing research. If a researcher needs to distinguish categorically between democracies and non-democracies in order to estimate differences in some other quantity across the two sets—or to analyze life courses within one of those sets, as one does with event history analysis—it makes more sense to base that split on a probabilistic measure of set membership than on an arbitrarily chosen cut point on a scalar measure of democracy. We still have to choose a break point to distinguish between categories, but  $p \geq 0.5$  has a natural interpretation (“probably a democracy”) that suits the task in a way that an arbitrary cut point on an index does not. And, of course, we can still perform sensitivity analyses by moving the break point and seeing how much our estimates change as a result.

## 2 Methods

Imagine that you have a set of cases that may or may not belong in some category of interest—here, democracy. Now imagine that you’ve got a set of experts who vote up or down on the membership status of each of those cases and sometime disagree. We can get a simple estimate of the probability that a given case belongs in the set of interest by converting the experts’ up or down votes to 1s and 0s and then averaging the series, and that’s not necessarily a bad idea.

If, however, we suspect that some experts are more error prone than others, and that those errors follow certain patterns, then we can do better by gleaning those patterns and adjusting the averages accordingly. That’s exactly what the Bayesian measurement error model does. It compares the experts’ votes across the entirety of the data, estimates the extent to which each one is in agreement or at odds with the others, and treats that variance as an estimate of each experts’ error rate. This error rate, in turn, is used to weight the contribution of each experts’ vote when calculating the posterior estimate of each case’s status. Instead of an unweighted average of the experts’ votes, we get something like a weighted average, where the weights are inversely proportional to the size of each experts’ apparent error rate. This inverse-error-rate-weighted average of the expert signals should be more reliable than the unweighted version, if the assumption that there is systematic bias in the experts’ judgments is largely correct.

## 2.1 Model Description

Formally, let each country’s (indexed by  $i$ ) true, unobserved probability of democracy at the end of year  $t$  be denoted  $d_{it} \in (-\infty, \infty)$ . Let  $\text{Logit}^{-1}(d_{it})$  represent the unconditional probability that an expert would say that country  $i$  is democratic at period  $t$ . As  $d_{it}$  approaches  $\infty$ , all experts would agree that the country is democratic and as it approaches  $-\infty$  experts would agree that the country is not. When  $d_{it} = 0$ , it is fair coin-flip whether an expert would report that country  $i$  is democratic in period  $t$ . For interpretability, we report  $\text{Logit}^{-1}(d_{it})$  instead of  $d_{it}$  in our results, but the underlying state uses an unbounded scale for estimation.

We have access to signals from a pool of experts indexed by  $j$ . Each year, each expert may report their information regarding whether the country is

democratic using a binary label  $d_{it}^j \in \{0, 1\}$ .<sup>1</sup> We assume their report is produced by the following process:

$$d_{it}^j \sim \text{Bernoulli}(\text{Logit}^{-1}(\frac{d_{it}}{\sigma_j} + \alpha_j))$$

Here  $\alpha_j$  represents the bias in expert  $j$ 's signal. Higher values correspond to a tendency to overstate the level of democracy in a country. The precision parameter  $\sigma_j$  represents how precisely the expert's signal corresponds to reality. As it increases, the expert's signal is more likely to be a biased coin flip rather than information with respect to  $d_{it}$ . The best possible expert will have  $\alpha_j = 0$  and as small a  $\sigma_j$  (which is strictly positive) as possible.

For identification, we assume that no expert's signal is systematically wrong:  $\sigma_j > 0$ . Without this assumption, there are two possible values for  $d_{it}$  rationalized by any given set of expert labels: one in which most of the experts are right, and one in which most of them are wrong.

We seek to infer parameters  $(d_{it}, \alpha_j, \sigma_j)$  from observed labels  $d_{it}^j$ . In order to close the model, we need to make assumptions about the prior distribution of these parameters. For the expert characteristics, we assume uninformative priors with some reasonable bounds, corresponding to the assumptions that no expert is too biased or noisy.

$$\alpha_j \sim \text{Uniform}(-2, 2)$$

$$\sigma_j \sim \text{Uniform}(0, 10)$$

For the latent state  $d_{it}$ , we explore a few different sets of assumptions. Each one imposes different structure on how the democracy state variable evolves over time, yielding more or less precision in the estimates at the cost of imposing potentially incorrect assumptions.

---

<sup>1</sup>Note we do not need to assume that all experts report a label for each country/year.

### 2.1.1 No model for democracy

$$d_{it} \sim \text{Uniform}(-10, 10)$$

Each country-year is a random draw from a uniform distribution with large absolute bounds. This assumption puts very little structure on the inference problem and lets the data drive most of the results, with the cost that we may less precisely estimate democracy.

### 2.1.2 Country random walks

$$d_{i1} \sim \text{Cauchy}(0, \lambda)$$

$$\epsilon_{it} \sim \text{Cauchy}(0, \nu)$$

$$d_{it} = d_{i,t-1} + \epsilon_{it}$$

In this model, each country starts period 1 with a level of democracy that is determined by a draw from the Cauchy distribution with median 0 and scale parameter  $\lambda$ . In subsequent periods, their levels of democracy receive a shock distributed with a median 0 Cauchy distribution with a different scale parameter  $\nu$  representing the relative per-period stability of democracy. Since the Cauchy distribution is fat-tailed, it can account for sudden shifts in democratic status while still imposing a mild smoothness assumption on the democracy measure.

## 2.2 Estimation

Due to complexity of the model and the use of priors to stabilize estimates, we used an efficient Hamiltonian Monte Carlo sampler (Hoffman and Gelman, 2012)[8] and computed our results using a state-of-the-art Bayesian model compiler (Stan Development Team, 2013)[12]. We simulated four Markov chains for 500 iterations, discarding the first 250 iterations for burn-in. We then used the

last 250 draws for estimation. We evaluated convergence by computing a potential scale reduction factor for each estimated parameter in the model (Gelman and Rubin, 1992)[7].

## 2.3 Data Sources

For inputs to our measurement error model, we compiled binary measures of democracy from the following five country-year data sets, all of which define the underlying concept in similar fashion.

- Cheibub, Gandhi, and Vreeland’s Democracy and Dictatorship (**DD**) Dataset (Cheibub et al., 2010)[3], which covers the period 1946–2008 and defines democracies as “regimes in which governmental offices are filled as a consequence of contested elections.”
- The “complete dataset of political regimes” (**BMR**) developed by Boix, Miller, and Rosato (2012) [1], which covers the period 1800–2007. They code as democracies cases in which “decisions to govern the state are taken through voting procedures that are free and fair” and a majority of the male population may vote.
- A binary indicator of democracy derived from Polity IV using the Political Instability Task Force’s coding rules (**PITF**) as described in Goldstone et al. (2010)[9]. Here, a country is considered a democracy if its chief executive is chosen in competitive elections (EXREC equal to 7 or 8) and political competition is not suppressed (PARCOMP of 0 or greater than 2). Polity IV covers the period 1800–2013, so this series does, too.
- The Democracy/Autocracy data set (**DA**) created by Ulfelder (2012)[14], which covers the period 1955–2010. In this data set, democracy is defined as “a form of government in which a free citizenry fairly chooses and routinely holds accountable its rulers.” That condition, in turn, is said to



exist when elected officials rule, elections are fair and competitive, politics is inclusive, and civil liberties are protected.

- The lists of electoral democracies (**FH**) in Freedom House’s annual Freedom in the World reports[5]. Freedom House only began producing this list in 1989, but it has been updated annually since then and currently runs through 2013. Its coding rules focus on regular, competitive elections, the absence of “massive” voter fraud, and universal adult suffrage.

In estimating our measurement model, we decided to make no prior assumptions about the five sources’ relative reliability. Different users might prefer to make stronger assumptions about the sources’ relative error rates, and those assumptions would have some effect on the resulting estimates. Absent a compelling prior reason to trust some sources’ coding decisions more than others, however, we chose to treat observed divergence from the central tendency as our best estimate of each source’s error rate.

### 3 Results

Over the next few pages, we discuss the estimated parameters of our models and the probabilities of democracy produced by one of them, the country random walks model described in Section 2.1.2. we chose to focus on the estimates from that model because they were most consistent with our intuition about levels and change over time in specific cases. Data produced by both models are available in our replication materials, however.

#### 3.1 Estimates of Bias and Variance

Figure 1 summarizes the estimates we obtained of our five sources’ bias and variance from the modeling strategies. Among other things, these estimates tell us that PITF’s Polity-based indicator and, to a lesser extent, Freedom House

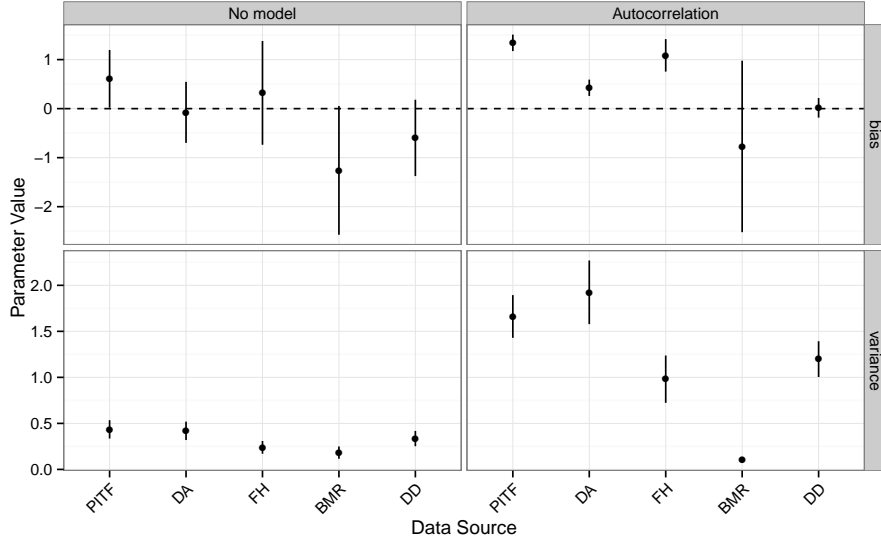


Figure 1: Posterior distributions for expert bias ( $\alpha_j$ ) and variance ( $\sigma_j$ ) parameters for our two models. Points represent posterior means while lines represent the 95% credible intervals.

list of electoral democracies are more inclined to see democracy where the other sources do not. By contrast, the MBR data set is clearly the most conservative.

The autocorrelation model requires inference on two additional parameters:  $\lambda$  which has credible interval  $[14.0, 20.0]$  and  $\nu$  which has credible interval  $[0.260, 0.36]$ .

### 3.2 Country-Year Estimates

Figure 2 uses a histogram to show the distribution of the estimated probabilities of democracy from the autocorrelation model for the years 1955 to 2008, the period for which we have inputs from at least four of the five source data sets. Reassuringly, the distribution has the expected shape. Most countries most of the time are confidently identified as democracies ( $p$  approximately 1) or non-democracies ( $p$  approximately 0), but the membership status of a non-

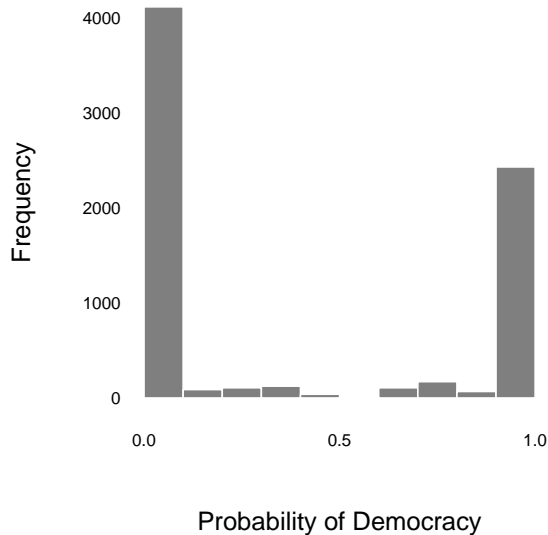


Figure 2: Histogram of estimated probabilities of democracy from autocorrelation model, 1955–2008

trivial subset of cases is more uncertain. Also, because this data summarized in this histogram span most of the second half of the twentieth century, we should expect to see —and do —that country-years of non-democracy outnumber country-years of democracy. If we were to restrict the period of observation to the post –Cold War period, the balance would shift in favor of democracies, albeit only marginally.

Figure 3 shows the mean probability of democracy for all regions of the world for each year during the period 1960–2008. The most striking feature of this plot is the upward trend over time that begins in the early 1970s —the start of what Huntington (1991)[10] dubbed the “third wave” of democratization —in all but two regions, the exceptions being South and Central Asia and the Middle East and North Africa. By the late 2000s, nearly all countries in the Americas and Europe and Eurasia are confidently identified as democracies. By contrast, the

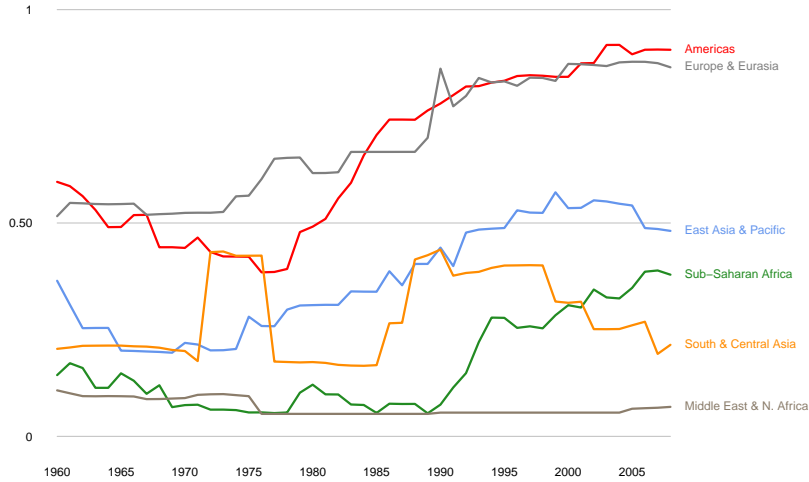


Figure 3: Mean Probability of Democracy by Region and Year, 1960–2008

mean condition in sub-Saharan Africa, East Asia and the Pacific, and South and Central Asia is ambiguity, with regional averages of approximately 50 percent. The Middle East and North Africa is the odd region out, with a regional mean close to 0 and very little variation over the 50-year period of observation.

Figure 4 drills down even further, comparing scores for some of the countries that (re-)gained their independence when the Soviet Union disintegrated in 1991. For purposes of this visualization, we limited the plot to the seven states in that region that have exhibited the most variation in their democracy status over the past 25 years. According to estimates from our model, the nine other post-Soviet states are either confidently identified since 1991 as democracies (Estonia, Latvia, and Lithuania) or non-democracies (Azerbaijan, Kazakhstan, Uzbekistan, Tajikistan, and Turkmenistan), so we left them off the chart to avoid clutter.

As they should, the sharp bends in estimated scores correspond to recognizable events. For example, Georgia’s status is ambiguous and trending less likely

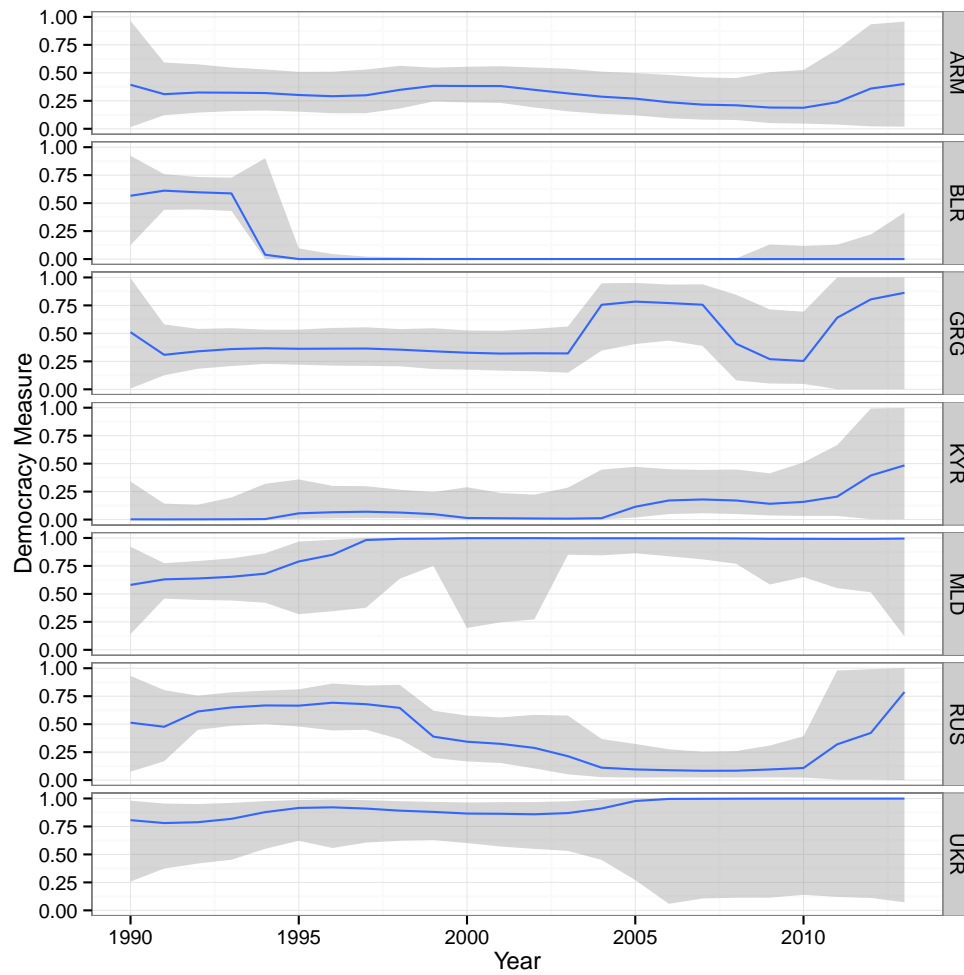


Figure 4: Estimated annual probabilities of democracy and their 95-percent confidence intervals for selected Soviet successor states, 1991–2013

until the Rose Revolution of 2003, after which point it's probably but not certainly a democracy, and the trend bends down again soon thereafter as Mikhail Saakashvili's United National Movement quickly consolidated power. Meanwhile, Russia is identified as a likely democracy after gaining independence, but its status becomes uncertain as power passes from Yeltsin to Putin and then solidifies as a likely non-democracy around the flawed legislative elections of 2003. For Moldova, the model's point estimate of the probability increases through the early 1990s and then sticks at approximately 1, but the uncertainty about that estimate widens in the early 2000s, a period of constitutional change and political instability. Finally, Kyrgyzstan is the lone case in this chart that began the period as a near-certain non-democracy but saw a significant increase in the ensuing two decades. It ticks up slightly after general elections in 1995, only to fall back near 0 after incumbent Askar Akayev won a second term in flawed balloting in 2000. It nudges back into more uncertain territory after the Tulip Revolution of 2005, however, and hovers right around the 0.5 threshold at the end of the period of observation.

Observers who know the region will probably be surprised to see the suggestion in those plots that Russia and Armenia became more recognizably democratic after 2010. In fact, those upturns are artifacts of attrition in the mix of data sets contributing "votes" to our estimates. In our analysis file, the D&D data set ends in 2007, MBR in 2008, and Ulfelder and Freedom House in 2010, leaving only PITF's Polity-based indicator for the last few years of those series. From the estimates of bias presented in Figure 1, we know that PITF/Polity is more inclined to see democracy when the other sources do not, and Armenia and Russia happen to be two cases where that divergence occurs. On Russia, for example, PITF/Polity sees an unbroken spell of democracy from 1991 onward, while DDR never sees democracy; MBR sees it ending in 1999, Ulfelder in 2003, and Freedom House in 2004.

Instead of truncating those plots to the period for which votes are available

from most or all of the source data sets, we chose to present them as is because this example helps illustrate some important features of the results. One of the strengths of the Bayesian measurement model is its ability to generate estimates from series that do not overlap completely, allowing us to continue updating our data even as some of the sources lag in their reporting or wink out entirely. At the same time, those updates will inevitably be shaped by the specific mix of contributing sources, and our certainty in the resulting estimates will decline (represented here by confidence intervals that widen as the pool of voters shrinks). Researchers working with the resulting data can decide whether to adopt a more conservative approach and only use estimates from periods when most or all sources cast votes—here, 1955 to 2010—or they might prefer to use as much available information as possible and work with the full series. In any case, we assume that it is better to have that choice than not.

We can also compare our model’s estimates of the *probability* of democracy with the estimates of countries’ *degree* of democracy that Pemstein et al. (2010) [13] get from their measurement model (i.e., the Unified Democracy Scores). Figure 5 is a scatter plot of observations from the two data sets for the period 1955–2008. Unsurprisingly, the two series are highly correlated. At the same time, we can also see that there are some real differences between them, especially at the edges. The smears of points along the 0 and 1 lines show that there is substantial variation in the degree of democracy among cases that our model confidently identifies as democracies or non-democracies in a categorical sense. This variation at the edges is relevant to analyses concerned with observing or estimating the effects of differences in countries’ degree of democracy, but it is not germane to analyses concerned with democracy as distinct regime type. What’s more, the fact that some countries with  $\text{Pr}(\text{democracy})$  approximately 1 have  $\text{UDS} < 0$  and some with  $\text{Pr}(\text{democracy})$  approximately 0 have  $\text{UDS} > 0$  suggests that the two series are not interchangeable. A categorical measure of democracy derived from  $\text{UDS} > 0$  will not produce the same strata as one

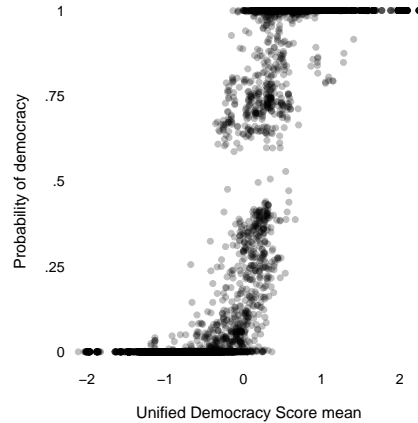


Figure 5: Comparison of probabilities of democracy estimated with measurement model and mean Unified Democracy Scores, 1955–2008

derived from  $\Pr(\text{democracy}) > 0.5$ . In other words, the choice of measures matters (cf. Casper and Tufis (2003)[2], Wilson (2013)[16]) and we believe that, for conceptual or practical reasons, ours is more appropriate for research in which democracy is treated as a matter of kind rather than degree.

## 4 Conclusion

We have described the motivation behind, design of, and output from a Bayesian measurement error model of countries’ categorical status as democracies or non-democracies. As we have noted throughout the paper, this approach accepts the premise that it is sometimes useful to construe democracy as a binary concept, but our model also makes more explicit our uncertainty about where many cases fall in the resulting two-part typology. Instead of another batch of 0s and 1s, we report probabilities of membership in the democracy set and the upper and lower bounds of confidence intervals associated with those estimates. We have not collected new information about the features of countries’ political regimes.



Instead, we have drawn on information previously and painstakingly created by other researchers to quantify our collective uncertainty about the presence or absence of democracy in countries over time.

We believe the resulting data provide a firmer foundation than measures of countries’ degree of democracy for the many studies that require a dichotomous distinction between democracies and non-democracies. The approach we have used is flexible, so it can easily be expanded to include other dichotomous indicators of democracy as inputs. This technique could also be applied to other regime types, such as military or personal rule, for which categorical data already exist but don’t always agree (cf. Geddes et al. (2014)[6] and Wahman et al. (2013)[15]).

If they were routinely updated, these estimates could also be useful to applied statistical forecasters whose models depend on binary classification of political regimes. Instead of generating a single estimate based on a crisp categorization of each case, forecasters could generate two estimates for each case—one as a democracy and one as a non-democracy—and then calculate a weighted average of the two, where the weight is provided by the estimated probability of democracy. This ensemble approach is standard practice in meteorology and other fields because it makes sense and generally produces more accurate results, and we expect the same would hold in forecasts of political events such as coups d’état and onsets of civil war.

## References

- [1] Boix C, Miller M and Rosato S (2012) A complete dataset of political regimes, 1800-2007. *Comparative Political Studies* 46 (12): 1523–1554.
- [2] Casper G and Tufis C (2003) Correlation versus interchangeability: the limited robustness of empirical findings on democracy using highly correlated data sets. *Political Analysis* 11 (2): 196–203.

- [3] Cheibub JA, Gandhi J and Vreeland JR (2010) Democracy and dictatorship revisited. *Public Choice* 143 (2-1): 67–101.
- [4] Collier D and Adcock R (1999) Democracy and dichotomies: a pragmatic approach to choices about concepts. *Annual Review of Political Science* 2: 537–565.
- [5] Freedom House (2014) *Freedom in the World*. Freedom House: Washington, DC.
- [6] Geddes B, Wright J and Frantz E (2014) Autocratic breakdown and regime transitions: a new data set. *Perspectives on Politics* 12 (2): 313–331.
- [7] Gelman A and Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Statistical science* 7 (4): 457–472.
- [8] Hoffman MD and Gelman A (2014) The no-U-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *The Journal of Machine Learning Research* 15 (1): 1593–1623.
- [9] Goldstone JA, Bates RH, Epstein DL, Gurr TR, Lustik MB, Marshall MG, Ulfelder J and Woodward ML (2010) A global model for forecasting political instability. *American Journal of Political Science* 54 (1): 190–208.
- [10] Huntington S (1991) *The Third Wave: Democratization in the Late Twentieth Century*. Norman, OK: University of Oklahoma Press.
- [11] Marshall MG and Jaggers K (2002) *Polity IV Project: political regime characteristics and transitions, 1800–2002, dataset user’s manual*. Center for International Development and Conflict Management, University of Maryland, College Park.
- [12] Stan Development Team (2014) *Stan Modeling Language Users Guide and Reference Manual, Version 2.5.0*.

- [13] Pemstein D, Meserve SA and Melton J (2010) Democratic compromise: a latent variable analysis of ten measures of regime type. *Political Analysis* 18 (4): 426–449.
- [14] Ulfelder J (2012) Democracy/Autocracy Data Set. <http://hdl.handle.net/1902.1/18836> V1 [Version]
- [15] Wahman M, Teorell J and Hadenius A (2013) Authoritarian regime types revisited: updated data in comparative perspective. *Contemporary Politics* 19 (1): 19–34.
- [16] Wilson MC (2013) A discreet critique of discrete regime type data. *Comparative Political Studies* 47 (5): 689–714.
- [17] Zadeh L (1965) Fuzzy sets. *Information Control* 8: 338–353.