A Probabilistic Measure of Democracy  
Derived from Dichotomous Data Sets

Jay Ulfelder and Sean J. Taylor

**Introduction**

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 (Cheibub et al 2010, Boix et al 2012, Ulfelder 2012). 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.

**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 follows 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.

Formal representation of model

For inputs to our measurement error model, we compiled binary measures of democracy from the following five country-year data sets, all of which use similar definitions.

* Cheibub, Gandhi, and Vreeland’s (2010) Democracy and Dictatorship (DD) data set, which covers the period 1946–2008 and defines democracies as “regimes in which governmental offices are filled as a consequence of contested elections.”
* Boix, Miller, and Rosato’s (2012) “complete dataset of political regimes,” 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](http://www.systemicpeace.org/polityproject.html) using the Political Instability Task Force’s[coding rules](http://onlinelibrary.wiley.com/doi/10.1111/j.1540-5907.2009.00426.x/abstract) (Goldstone et al. 2010). 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 lists of electoral democracies in Freedom House’s annual [*Freedom in the World*](http://www.freedomhouse.org/report-types/freedom-world#.U_zo2vk7u-M)reports. Freedom House only began producing this list in 1989, but it has been updated annually ever since, right through 2013. Its coding rules focus on regular, competitive elections, the absence of “massive” voter fraud, and universal adult suffrage.
* Ulfelder’s (2012) Democracy/Autocracy data set, which covers the period 1955–2010. Ulfelder defines a democracy 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.

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.

**Results**

Discuss estimates of bias and variance for the five sources.

Figure 1 below shows a histogram of the estimates for the period 1989–2007, the only years for which we have inputs from all five of the source data sets. Reassuringly, the distribution has the expected shape. Most countries most of the time are readily identified as democracies or non-democracies, but the membership status of a sizable subset of cases is more uncertain. Also, because this histogram looks only at the post–Cold War period, we are not surprised to see that cases confidently identified as democracies slightly outnumber cases confidently identified as non-democracies.

Of course, we can and should also look at the estimates for specific cases. I know a little more about countries that emerged from the collapse of the Soviet Union than I do about the rest of the world, so I like to start there when eyeballing regime data. The chart below compares scores for several of those countries that have exhibited more variation over the past 20+ years. Most of the rest of the post-Soviet states are slammed up against 1 (Estonia, Latvia, and Lithuania) or 0 (e.g., Uzbekistan, Turkmenistan, Tajikistan), so we left them off the chart. We also limited the range of years to the ones for which data are available from all five sources. By drawing strength from other years and countries, the model can produce estimates for cases with fewer or even no inputs. Still, the estimates will often be less reliable for those cases, so we thought we would focus on the estimates based on a common set of “votes.”

Those estimates look about right to me. For example, Georgia’s status is ambiguous and trending less likely until the [Rose Revolution](http://en.wikipedia.org/wiki/Rose_Revolution) of 2003, after which point it’s probably but not certainly a democracy, and the trend bends down again soon thereafter. Meanwhile, Russia is fairly confidently identified as a democracy after the [constitutional crisis of 1993](http://en.wikipedia.org/wiki/1993_Russian_constitutional_crisis), but its status becomes uncertain around the passage of power from Yeltsin to Putin and then solidifies as most likely authoritarian by the mid-2000s. Finally, Armenia was one of the cases I found most difficult to code when building the Democracy/Autocracy data set for the Political Instability Task Force, so I’m gratified to see its probability of democracy oscillating around 0.5 throughout.

One nice feature of a Bayesian measurement error model is that, in addition to estimating the scores, we can also estimate confidence intervals to help quantify our uncertainty about those scores. The plot below shows Armenia’s trend line with the upper and lower bounds of a 90-percent confidence interval. Here, it’s even easier to see just how unclear this country’s democracy status has been since it regained independence. From 1991 until at least 2007, its 90-percent confidence interval straddled the toss-up line. How’s that for uncertain?

**Conclusion**

The estimates from our measurement error model look great to me. As someone who has spent a lot of hours wringing my hands over the need to make binary calls on many ambiguous regimes (Russia in the late 1990s? Venezuela under Hugo Chavez? Bangladesh between coups?), I think these estimates are accurately distinguishing the hazy cases from the rest and even doing a good job estimating the extent of that uncertainty.

Pemstein, Melton, and Meserve (2010) use the same technique to produce the [Unified Democracy Scores](http://www.unified-democracy-scores.org/). Our method is virtually identical, but the concept and thus the results are a little different. Their approach treats democracy as a matter of degree and derives a composite scalar index of democracy-ness from several measures. By contrast, we accept a binary conceptualization of democracy and estimate the probability that a country qualifies for that label.

In fuzzy set theory (Zadeh 1965), 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 potentially meaningful, and the former will sometimes be a better fit for an analytic task than the latter. As Collier and Adcock (1999) 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 on the aims and design of the research. For example, if a researcher wants to distinguish categorically between democracies and non-democracies in order to estimate differences in some other quantity across the two sets, it will often make more sense to base that split on a probabilistic measure of set membership than an arbitrarily chosen cut point on a scalar measure of democracy. We still have to choose a break point to distinguish between categories, but “greater than 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. For example, a more cautious approach might only identify as democracies those cases for which the lower bound of the 90-percent confidence interval is above 0.5, which we might interpret as “almost certainly a democracy.”

Finally, if they were routinely updated, these estimates could also be useful to applied statistical forecasters whose models depend on binary classification of political regimes as democracies and non-democracies. 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.

**Works Cited**

Boix, Carles, Michael Miller, and Sebastian Rosato. 2012. “A Complete Dataset of Political Regimes, 1800-2007.” *Comparative Political Studies* Vol. 46, No. 12, pp. 1523-1554.

Casper, Gretchen and Claudiu Tufis. 2003. “Correlation Versus Interchangeability: The Limited Robustness of Empirical Findings on Democracy Using Highly Correlated Data Sets.” *Political Analysis* Vol. 11, No. 2: 196-203.

Cheibub, José Antonio, Jennifer Gandhi, and James Raymond Vreeland. 2010. “Democracy and Dictatorship Revisited.” *Public Choice*, Vol. 143, No. 2-1, pp. 67–101.

Pemstein, Daniel, Steven A. Meserve, and James Melton. 2010. “Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type.” *Political Analysis*, pp. 1-24.

Ulfelder, Jay. 2012. “Democracy/Autocracy Data Set,” http://hdl.handle.net/1902.1/18836 V1 [Version]

Zadeh, Lofti. 1965. “Fuzzy sets.” *Information Control* Vol. 8, pp. 338–353.