Probabilistic Democracy

Muhammet A. Bas
Division of Social Science
New York University Abu Dhabi

Randall Stone
Department of Political Science
University of Rochester

Abstract

Democracy is an important concept that is difficult to measure, and existing measures have well-known weaknesses. We define democracy narrowly, as electoral accountability, and estimate the probability of democracy using a structural model. We allow for strategic voting and find evidence that voters are intimidated into supporting authoritarian leaders. Ratification of the Convention Against Torture by the country in question emboldens voters, while ratification by third parties, close relations with the United States and the incumbent's military experience increase voter intimidation. Our estimated democracy scores are highly correlated with other measures frequently used in political science, and come with important advantages, including conceptual clarity, replicability, out-of-sample score prediction, flexibility with respect to the variables and specifications used in the estimation model, estimates of uncertainty, and avoiding potential expert bias.

A minimalist definition of democracy requires that competitive elections be held, and that the incumbent step down if she loses (Przeworski, 1991; Przeworski et al., 2000). This definition does not capture everything that is commonly thought to be important about democratic institutions, but it does capture a necessary condition. Electoral accountability is essential for the operation of democratic institutions and is an important dimension along which contemporary electoral systems vary. This definition has an advantage over the multidimensional conceptions of democracy behind the commonly employed quantitative measures, because it is objective and does not require aggregation of diverse indicators. However, it defines electoral accountability as a conditional probability that needs to be estimated in some cases. Electoral accountability can be directly observed when the incumbent steps down, and its absence can be inferred when no elections are held, but when the incumbent appears to win, measuring electoral accountability is a matter of estimating the probability that the incumbent would have stepped down, conditional on losing.

To illustrate the point, consider a recent event. Presidential elections were held in 2018 in Russia, and Vladimir Putin was declared the winner. Was electoral accountability in place? Critics charged that the election was not "free and fair," because the incumbent had a number of extra-constitutional advantages. He controlled the mass media, intimidated his opponents, controlled an impressive system of patronage, and employed various forms of electoral misconduct. Supporters pointed out that Putin remains highly popular in Russia despite all of the things his Western detractors say about him, and he almost certainly won a majority of the votes. The problem is that we do not observe an outcome that would allow us to judge with certainty whether electoral accountability was in operation: it may be the case that Putin would have won a fair election, but refused to conduct one; and it may be the case that he would have refused to step down had he lost. Since we only observe the decision whether to accept an unfavorable electoral outcome when such an outcome occurs, we are left to draw uncertain inferences.

This inference problem has become more acute in recent decades, as it has become more common for authoritarian regimes to hold elections. Before the end of the Cold War, most authoritarian regimes did not hold elections, and today, most authoritarian regimes do. It has become increas-

ingly valuable to authoritarian leaders to masquerade as popularly elected representatives, both because their citizens' expectations have risen and because a range of international benefits are available if they can pass as democratically legitimate (Hyde, 2011). Elections, meanwhile, have come to play an important role in authoritarian governance; but authoritarian leaders sometimes make unintended democratic transitions when they lose elections.

Przeworski et al. (2000), whose minimalist definition we adopt, attempt to get around this problem of uncertain cases by using a conservative coding rule, and judging a political system to be democratic only after a peaceful transfer of power to the opposition has taken place. This leads to a low frequency of type I errors (false democracies) at the cost of a higher frequency of type II errors (false non-democracies). In addition, it does not allow for the existence of borderline cases, for uncertainty, or for the possibility that the probability of compliance varies between elections. To allow for these possibilities, in this paper we offer an approach that estimates democracy, understood as electoral accountability, as an alternative to coding it under uncertainty.

In what follows, we introduce a method to estimate the probability of stepping down, conditional on losing an election. The model allows for partial observability — we may only know for certain that the election was lost if we observe that the incumbent steps down — but also allows us to incorporate more detailed information that we have about particular cases to improve the efficiency of our estimates. In a few cases, our estimates of democracy differ in interesting ways from the codings by Przeworski et al. (2000) and by Polity, which is consistent with the interpretation that uncertainty about the efficacy of electoral accountability is generally greater than it appears to the analyst in retrospect. In the case of Russia, we find that it is unlikely that Putin would comply if he were voted down.[1] We can say more than this, however; we have a point estimate that changes over time and a confidence interval, and we think both of these tell us something important about Russia's political system.

The model is strategic, as we explain below, which in this case means that voters can take into account their expectations about whether the leader will comply when they decide how to

[1]The estimated probability was 0.36 in 2000 and 0.44 in 2004.

vote. This allows for a novel explanation for elections in authoritarian regimes: unpopular leaders can afford to take the risk of holding elections because they can rely on a portion of the population to vote for them strategically in order to avoid the conflict and disorder that would follow if the leader lost and repudiated the election. We find that voter intimidation is an important substantive explanation for electoral outcomes in semi-competitive political systems. Ratification of the Convention Against Torture by the country in question emboldens voters, while ratification by third parties, close relations with the United States, and the incumbent's military experience increase voter intimidation. We use a comparative model test to directly test the hypothesis that vote choices depend on the expected probability that the incumbent steps down when faced with electoral defeat.

Our estimated democracy scores are highly correlated with other measures frequently used in political science, including those of ACLP (Przeworski et al., 2000), Polity (Marshall and Jaggers, 2007), Freedom House, and V-Dem (Coppedge et al., 2017). In addition, our measure comes with a few desirable properties, including conceptual clarity; estimates of uncertainty; replicability; flexibility with respect to the variables and specifications used in the estimation model; ability to generate out-of-sample predictions; and avoiding potential expert bias. Our measure differs from all of the above measures except ACLP in offering a minimalist, unidimensional measure of electoral control. The high correlation with other measures of democracy that use different criteria suggests that electoral accountability is central to democratic governance. Our measure differs from all except V-Dem in offering an estimate of the uncertainty of the measure itself, and it is the only measure that is replicable using publicly available data and code, and that can be customized by researchers with particular interests by altering the sample or estimation specification. We demonstrate how the estimates can be validated by generating out-of-sample predictions, and how changing the estimation window can generate additional substantive insights. Finally, this is the first measure of democracy that does not rely on expert coding, avoiding potential subjective assessments or biases from human coding.

1 A Minimalist Conception of Democracy

The conception of democracy advanced in Przeworski (1991) and Przeworski et al. (2000) is based on the effectiveness of electoral accountability. In order for democratically elected leaders to represent the preferences of the citizenry and safeguard their liberty, the electorate must be able to replace an unsatisfactory leader. Necessary conditions for the operation of democracy are that leaders are subject to competitive elections, and that when they lose, they step down. As Przeworski (1991) puts it, in order for democracy to be a self-enforcing equilibrium, it must be the case that opposition candidates have incentives to challenge the incumbent, that the outcome of the election be uncertain, and that the incumbent prefers to concede defeat if she loses.

This way of posing the question focuses on the credibility of elections, which is a key empirical issue facing contemporary electoral systems. Elections have become ubiquitous features of even authoritarian political systems. Authoritarian leaders use plebiscites and semi-competitive elections as ways to cement their legitimacy and demonstrate their popularity to rivals and to foreign and domestic audiences. Indeed, the benefits of international recognition spur "pseudo-democrats" to invite international monitors to oversee their elections, even when they intend to cheat (Hyde, 2011). In many cases, the ranks of potential challengers are screened to prevent the emergence of real threats, either through legal maneuvers or through intimidation. The media may be closely controlled and biased in favor of the incumbent. Vote buying, ballot stuffing and electoral repression tilt the competition in the incumbent's favor. It might seem that interfering with electoral outcomes so overtly would defeat the purpose of holding elections in order to demonstrate the dictator's popularity, but recent work suggests that insecure authoritarian leaders benefit from holding unfair elections because their opponents are left uncertain about how much support they have (Rozenas, 2016). Even rigged elections can be lost, however, and this has become one of the more common routes to democratization. Meanwhile, some of the same tactics are used to advantage incumbents in a wide range of democratic states, although the more overt forms of manipulation are most common in developing countries (Stokes et al., 2013). Consequently, the dividing line be-

tween democracy and authoritarianism has become blurred, and the key feature that distinguishes between the two is the probability that the incumbent, if defeated, would in fact step down.

This conception is minimalist in the sense that it identifies only a necessary condition for democracy, and not a sufficient condition. It is an empirical question whether competitive elections and electoral accountability guarantee the free exercise of a wide range of rights and liberties. According to the Freedom House scale, in contrast, these rights and liberties are the defining features of democracy. Similarly, institutional features such as division of powers, constraints on the executive, and an independent judiciary may be necessary for electoral accountability to be effective, but they are not part of the minimalist definition. In contrast, these institutional features are the key defining features of democracy according to the polity project (Jaggers and Gurr, 1995).

Our definition does not incorporate political participation. While eschewing the term democracy, Dahl (1973) argued that polyarchy was defined along two dimensions, contestation and participation. Barber (2003) argued that the quality of democracy depended on the breadth and depth of participation. The notion that participation is central to democratic governance and depends on a supportive political culture goes back to De Tocqueville (2003), and finds expression in a long line of comparative behavioral studies of political culture (Verba and Almond, 1963). Moreover, our definition does not impose restrictions on the membership of the electorate, which is often held to be a key defining feature of democracy. An institutional view of democracy proposed by Bueno de Mesquita et al. (2003) argues that the defining features of political systems are the size of the selectorate that chooses the leader and the size of the necessary winning coalition.

More broadly, the minimalist definition of democracy does not make any claims about representation. This may be regarded as a theoretical advantage, because electoral accountability has more secure game-theoretic micro-foundations than representation. Representative notions of democracy run into difficulties because diverse preferences of members of society have to be aggregated by institutions. Riker (1982) argued against what he regarded as populist conceptions of democracy on the grounds that formal theory suggested that substantive representation was not really feasible in a world with two or more salient policy dimensions. His "liberal" conception of

democracy was narrower, and similar to Przeworski's conception. Similarly, the chief concern of early democratic theorists was to prevent the usurpation of power by a tyrant. For Calvin (1536) and Locke (1689), rebellion was justified to overthrow tyranny, but not in order to ensure that government policies reflected majority preferences. Similarly, the Federalists argued for the division of powers on grounds that this created the means and incentives for officials to hold each other in check. Institutional design was chiefly useful in order to ensure that democracy was a self-enforcing equilibrium (Hamilton et al., 1788).

For better or for worse, we focus on a minimalist conception of democracy as effectiveness of electoral accountability. A parsimonious definition has advantages in terms of measurement. We are not required to make subjective judgments, to aggregate indicators that represent diverse concepts, or to choose arbitrary weights to attach to the subcomponents (Treier and Jackman, 2008). There is no conceptual drift between our definition and measurement of the variable. Defined minimally, democracy is the probability of stepping down, conditional on losing an election.

1.1 A Statistical Model of Probabilistic Democracy

To estimate the probability of an incumbent leader stepping down, conditional on losing an election, we build a structural model of a strategic interaction between a representative voter (V) and an incumbent leader (I) during an election, represented in Figure 1. In this simple model, the voter decides to reelect or replace the leader. If the incumbent loses, she has the option of stepping down or manipulating the election and announcing victory. The game has three outcomes: re-election (RE), election manipulation (MP), and the leader stepping down (SD). The outcome of this interaction is partially observable: we observe whether the incumbent steps down, but not whether the voter chooses to re-elect. Consequently, if the leader remains in power, it is possible that the leader was re-elected, but also possible that the leader lost the election but manipulated the results to stay in power. The statistical model we propose below aims to separate such cases and provide estimates of the probability of each scenario.

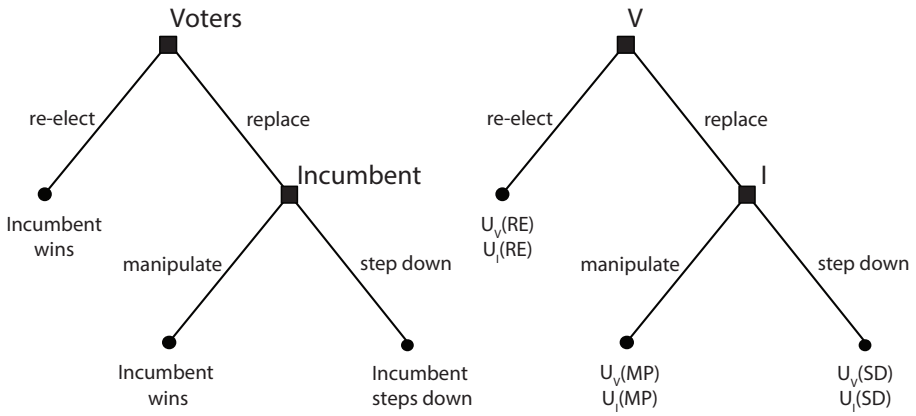

Figure 1: A Model of Probabilistic Democracy

The partial observability strategic probit maximum likelihood estimator we use is proposed by Bas and Stone (2014), and builds on a partial observability model introduced by Poirier (1980). Non-strategic applications of such models in political science include Przeworski and Vreeland (2000) and Vreeland (2003). Partial observability implies that in the statistical model, the outcome variable is binary, taking values $\{\text{SD, }\tilde{}\text{SD}\}$. Unlike standard logit or probit models for binary dependent variables, this model takes into account strategic interaction and allows non-monotonic and interactive effects of individual regressors. Actors' utilities from various outcomes in Figure 1 are modeled as linear functions of regressors. For identification reasons, some of these utilities are normalized to zero: while $\mathrm{U_I(SD)} = \mathrm{X_I^{SD}}\beta_{\mathrm{I}}^{\mathrm{SD}}$, $\mathrm{U_V(SD)} = \mathrm{X_V^{SD}}\beta_{\mathrm{V}}^{\mathrm{SD}}$, and $\mathrm{U_V(MP)} = \mathrm{X_V^{MP}}\beta_{\mathrm{V}}^{\mathrm{MP}}$ are estimated as three separate utility equations, the rest are normalized as $\mathrm{U_I(MP)} = 0$ and $\mathrm{U_V(RE)} = 0$ to enable estimation. The three estimated utilities are interpreted relative to the normalized utility for each player.

Below, we briefly describe the likelihood function for the strategic version of the estimator. We focus on the agent-error specification, where the expected utility from each choice for each actor has a random component ϵ_{ji} that follows a standard normal distribution (Signorino, 1999).[2] Hence,

[2]As a common assumption in such models, we assume that the random components ϵ_{ji} are independent from each other to facilitate estimation. This follows Signorino (2002)'s finding that modeling strategic interdependence is more important than accounting for potentially correlated disturbances. For example models that allow for correlations, see

the voter's probability of replacing the leader, p_V, and the leader's probability of stepping down after an election loss, p_I, are defined as follows:

$$\begin{aligned} p_V = \Pr_V(\text{Replace}) &= \Pr\left(EU_V(\text{Replace}) + \epsilon_{V2} \geq EU_V(\text{Reelect}) + \epsilon_{V1}\right) \\ &= \Phi\left(\frac{p_I U_V(SD) + (1-p_I)U_V(MP) - U_V(RE)}{2}\right) \\ p_I = \Pr_I(\text{StepDown}) &= \Pr(EU_I(\text{StepDown}) + \epsilon_{I2} \geq EU_I(\text{Manipulate}) + \epsilon_{I1}) \\ &= \Phi\left(\frac{U_I(SD) - U_I(MP)}{2}\right) \end{aligned} \tag{1}$$

Since the RE and MP outcomes are only partially observable, the outcome probabilities for the partial observability strategic probit model are given as follows:

$$\begin{aligned} \Pr(SD) &= p_V p_I \\ \Pr(\sim SD) = \Pr(RE) + \Pr(MP) &= 1 - p_V p_I \end{aligned}$$

We estimate this model using maximum likelihood (MLE), based on the following log-likelihood function:

$$\ln(L) = \sum_i^N I_{SD}(\ln(p_V) + \ln(p_I)) + (1 - I_{SD})\ln(1 - p_V p_I)$$

where I_{SD} is an indicator variable recording whether the leader stepped down after the election or not. Estimating the parameters of this model will allow us to calculate our measure of democracy, $\Pr_I(\text{StepDown})$, as defined above in equation 1.

1.2 Estimating Democracy

We use the National Elections Across Democracy and Autocracy (NELDA) Data Set to identify competitive elections for the office of the national leader (president or prime minister) during the period from 1945 to 2008 (Hyde, 2011), defined as elections in which multiple parties were legal, at least one opposition party competed, and multiple candidates appeared on the ballot. Our dependent variable, *Stepdown*, answers the question, "Did the incumbent leader step down after the election?" In terms of our theoretical model, the zeroes reflect partial observability: they include

Lewis and Schultz (2003) and Bas, Signorino and Whang (2014). The existence of partial observability makes such extensions infeasible in our application.

cases in which the incumbent won fair elections, and also cases in which the incumbent would have lost a fair election, but was able to manipulate the results to avoid stepping down.

In order to assist with the identification of the partial observability model, we incorporate information about a few of the zeroes.[3] This practice can be viewed as a middle ground between expert coding of all partially observable cases as in Przeworski et al (2000) and a purely estimation based measurement approach. We use a Monte Carlo simulation (described in the appendix) to study the effect of introducing varying numbers of anchoring observations on the Root Mean Squared Error (RMSE) of the estimator. Figure 2 plots the average RMSE of the democracy measure as a function of the proportion of the partially observable cases that are revealed through Monte Carlo iterations. As the figure illustrates, adding a small number of anchoring observations significantly reduces the RMSE from the partial observability baseline. After this initial gain, however, additional anchoring observations have diminishing returns, and the estimator approaches the full observability threshold. The intuition is that partial observability models place high demands on their identification assumptions, so a small number of fully-identified observations can substantially improve the model.

In order to identify the model, we normalize some utilities to zero and make assumptions about which covariates are relevant to which choices. Figure 1 shows that the leader moves only after the voter has chosen to replace her. Consequently, we can estimate only one set of coefficients for the leader, which represents the difference in utilities between manipulating the election (MP) and stepping down (SD), conditional on not having been reelected. Without loss of generality, we normalize the leader's utility of manipulation (MP) to zero, so positive coefficients indicate covariates that make stepping down more attractive than manipulation. Things are more complicated for the voter. Without loss of generality, we normalize the voter's utility of reelecting the leader to zero.

[3]Elections coded as competitive wins are the elections in the United States, Canada, UK, France (after 1947), Germany, Sweden, Finland, Norway, Denmark, Netherlands, and Belgium; those coded as manipulated are Iran in 2009, Zimbabwe in 2008, Ukraine 2004 (first election), Ethiopia in 2005, Guyana in 1980, Philippines in 1986, Zambia in 2001, Haiti in 1995 and 2000, and Togo in 2003 and 2005 (identified by Hyde 2011).

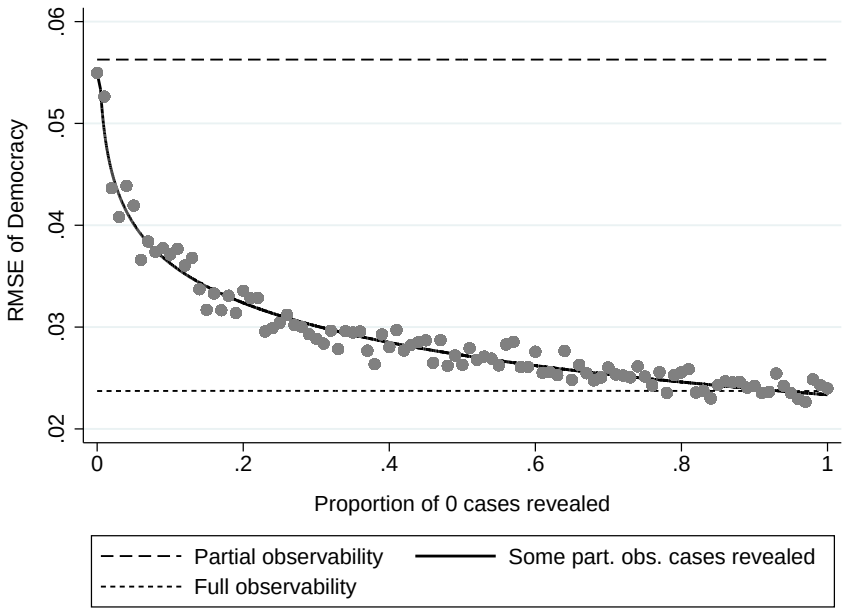

Figure 2: Effect of Including Anchoring Observations

This means that any covariates that we assign to the MP or SD outcomes in the voter's utility are used to estimate the difference in utilities between that outcome and reelecting the leader. The difference between reelection and stepping down is that the leader changes, so we assign covariates to SD that measure the voter's valuation of the leader: economic performance (GDP, growth and financial crises) and security performance (interstate and intrastate conflict and hostility level). The leader does not step down if she manipulates the election, so leader performance is irrelevant to the comparison between that outcome and reelection, but in that case there is potential for civil unrest. Consequently, we do not include leader performance variables in the voter's utility for MP, but do include variables relevant to civil unrest: the leader's military background, ideal point difference with the United States, ratification of the Convention against Torture (CAT), mountainous terrain, ethno-linguistic fractionalization, and the length of the leader's tenure. Other variables assigned to the leader's utility but not to the voter's utilities are assumed to affect the voter's utility only indirectly, through their effects on the leader's choice. This set includes lagged election outcome (precedent for stepping down); percentage of other countries that have ratified the CAT (which

affects the leader's retirement options); and variables that affect the risk of a coup (urbanization, military personnel and polarization (Vogt et al., 2015)).

The results of our analysis are presented in Table 1, where the coefficients in each column represent the effects of covariates on a particular actor's utility for a particular outcome. The marginal effects of each variable on each actor's binary choice and their confidence intervals are presented in Table 2. Because the model is strategic, any variable that appears in both the voters' and the leader's utility function has a compound effect on the vote choice: it influences the voters' valuation of outcomes and the voters' assessment of the probability that the leader will step down if defeated. Consequently, the statistical significance of coefficients may not coincide with statistical significance of marginal effects.[4]

The estimated coefficients for SD (second column) represent the effects of the covariates on the difference between the utility of removing the leader peacefully and reelecting her. We assume that factors related to the evaluation of leader performance influence this choice: economic well-being (GDP per capita, economic growth, and the occurrence of an economic crisis) and security (inter-state war, intrastate or civil war, and severity of conflict). The economic variables have the results that are expected from the economic voting literature. GDP per capita has the strongest effect: a one-standard deviation increase in GDP per capita, or $10,000, decreases the probability that the electorate chooses to replace the leader by 33.7 percentage points. The level of economic growth has a weaker effect in the same direction—a one-standard-deviation increase in growth decreases the probability by 7.6 percentage points—and economic crises have a marginally significant effect of encouraging replacement of the leader. Economic crises, as defined by Reinhart and Rogoff (2009), include financial crises, banking crises, exchange rate crises, sovereign debt crises or repudiation of domestic debt. Their effect is only marginally significant in the main specification,

[4]The model effectively interacts all of the variables in the leader's utility with all of the variables in the voters' utility when calculating the voters' choice, so the usual caveats about the statistical significance of interactive estimates apply. We discuss this when it plays an important role in the interpretation. Marginal effects assume all variables at their means (modes for categorical variables).

Table 1: A Statistical Model of Probabilistic Democracy

	Voters' MP Utility $\beta_{\mathbf{V}}^{\mathbf{MP}}$	**Voters' SD Utility** $\beta_{\mathbf{V}}^{\mathbf{SD}}$	**Leader's SD Utility** $\beta_{\mathbf{I}}^{\mathbf{SD}}$
Lagged Election Outcome	-	-	1.353***
			(0.220)
Military background	2.372	-	-0.829***
	(1.543)		(0.245)
Ideal Point distance	2.318***	-	0.462***
	(0.703)		(0.147)
GDP per cap.	-	-0.026*	0.108***
		(0.014)	(0.020)
GDP per cap. Growth	-	-7.681**	-
		(3.387)	
CAT ratifier	3.250**	-	-
	(1.503)		
% CAT ratifiers	-	-	-2.035***
			(0.413)
Mountainous Terrain	-0.010	-	-
	(0.030)		
ELF	-1.130	-	-
	(2.304)		
ln(Tenure)	-7.330***	-	-
	(1.376)		
Interstate Conflict	-	-1.558***	-
		(0.482)	
Intrastate Conflict	-	0.055	-
		(0.385)	
Hostility Level (avg.)	-	0.462***	-
		(0.146)	
Financial crises (#)	-	0.205*	-
		(0.119)	
Urban Population	-	-	0.938
			(0.830)
Military Personnel pc	-	-	0.143
			(0.196)
Polarization	-	-	-0.734**
			(0.325)
Constant	55.694***	-0.071	-0.602
	(10.308)	(0.384)	(0.413)
Observations	932	932	932

* $p \leq 0.10$, ** $p \leq 0.05$, *** $p \leq 0.01$

but is highly significant in specifications that do not include economic growth. Interstate conflict has a strong effect that discourages replacing the leader. The probability of voting the leader out of office is reduced by 27.6 percentage points during a conflict, which is consistent with behavioral arguments about a "rally around the flag" effect and with strategic arguments about "gambling for resurrection" or diversionary war (Downs and Rocke, 1995; Chiozza and Goemans, 2011; Debs and Goemans, 2010). Civil war has no significant effect, but the severity of international conflict is associated with significantly increased probability of replacing the leader, which is consistent with the result in the public opinion literature that casualties have cumulative effects that undermine support for war (Mueller, 1994).

In contrast, the estimated utility for MP (first column) represents the difference in utilities between retaining the leader through political manipulation or by reelection. The leader remains in power either way, so the quality of leadership is irrelevant; the relevant variables affect the attractiveness to the electorate of a political struggle over the succession. We focus on personal characteristics of the leader that are relevant to regime stability (duration of tenure and military background); international factors (UN voting and ratification of the Convention Against Torture); and the feasibility of waging a civil war (mountainous terrain and ethno-linguistic fractionalization). The model estimates that with other variables at their means or modes, voters prefer to reelect the leader rather than have the leader retain office through electoral manipulation if the leader's tenure is above average (approximately eight years). Increasing the length of tenure by one standard deviation (to 26 years six months) reduces the estimated probability of voting against the leader by 11 percentage points. The leader's military background increases the probability of voting against the leader by 21 percentage points.

Two international factors have striking effects. Voters appear to be more assertive if their governments have ratified the Convention Against Torture (CAT), voting against the leader approximately 9 percentage points more often. This could be because adopting the CAT increases the cost of using politically repressive tactics, or because ratification facilitates collective action against the regime by creating a focal point for protest (Simmons, 2009). Alternatively, we could observe

Table 2: Marginal Effects

	Vote Down (from a baseline of .78)		Step Down (from a baseline of .69)	
Variable	**Marginal Eff.**	**95% CI**	**Marginal Eff.**	**95% CI**
Lagged outcome	-	-	.235	(.159, .311)
Military background	.214	(.074, .353)	-.226	(-.355, -.096)
Ideal point distance	-.022	(-.165, .120)	.111	(.053, .170)
GDP pc	-.337	(-.462, -.213)	.201	(.149, .253)
GDP pc growth	-.076	(-.149, -.004)	-	-
CAT ratifier	.090	(.001, .179)	-.164	(-.231, -.097)
ln(Tenure)	-.111	(-.143, -.079)	-	-
Interstate MID	-.276	(-.469, -.083)	-	-
Civil conflict	-	-	-	-
Hostility level	.061	(.017, .105)	-	-
Financial crises	.029	(-.006, .063)	-	-
Polarization	-	-	-.119	(-.226, -.012)
% CAT ratification	-	-	-.164	(-.231, -.097)

this because countries that are unlikely to employ torture against the opposition are more likely to ratify the CAT, and this knowledge emboldens the opposition. In any case, we find no evidence to support the hypothesis that repressive dictators sign the CAT in order to signal their type and deter the opposition (Vreeland, 2008); in that case, we would expect CAT ratifiers to experience less opposition, rather than more.

In addition, we find a significant coefficient of close political relations with the United States, measured in terms of similarity of ideal points estimated from United Nations voting records (Bailey, Strezhnev and Voeten, 2017). The marginal effect of ideal point distance on vote choice is insignificant when all variables are at their means, because there is a compound effect: ideal point distance also affects the leader's choice, and this in turn affects the voter's choice. However, the significant coefficient suggests that voters in countries that are closely aligned with the United States are deterred from voting against their leaders. Leaders who are closely aligned with the United States may be more likely to succeed when they attempt political manipulation, because they will be subject to less international pressure and more support.

The intuitive results for voter preferences strengthen our confidence that the model is well identified, and consequently make us more confident in our interpretation of the results for the

leader's choice. The leader only faces a decision in this model if the voters choose to reject her, so the interpretation of the leader's utilities is straightforward: remaining in office through political manipulation is normalized to zero, and the estimated utility of stepping down (SD) is measured relative to that outcome. In other words, the estimated coefficients indicate how the covariates influence the leader's incentive to comply with an adverse electoral verdict, which is exactly our minimal definition of democracy. We focus on four sets of factors: characteristics of the political system (previous electoral outcome, GDP per capita, and polarization), leader characteristics (military background), the feasibility of repression (urbanization, military personnel), and international factors (alignment with the United States and ratification of the CAT in the rest of the world). Only the variables associated with the means of repression were insignificant.

Three systemic variables play a key role in predicting democracy. The previous electoral outcome is a significant covariate that captures the idea of democratic consolidation. As Przeworski (1991) argued, democratic institutions function properly when they represent a self-enforcing equilibrium: alternation in office over the long run makes electoral defeat in the short run tolerable. The best indicator that this is the case is that the previous incumbent surrendered power voluntarily, and we find that this is associated with a 23.5 percentage point increase in compliance. Second, we find that the level of economic development (per capita GDP) is a strong predictor of compliance. This is consistent with the finding of Przeworski et al. (2000) that democracies that had achieved a high enough level of per capita income were unlikely to revert to authoritarianism. Highly developed economies have educated populations that tend to be politically engaged and efficacious, and they provide resources for social groups to mobilize politically. The increased strength of popular opposition makes electoral manipulation less attractive and less likely to succeed. We find that increasing per capita GDP by one standard deviation increases the probability of compliance by 20.1 percentage points. Third, political polarization decreases the incentive to comply. As political elites become increasingly polarized, compliance in the future becomes more uncertain, which undermines the incentive to comply in the present. A one-standard-deviation increase in

the polarization index is associated with an 11.9 percentage point decrease in the probability of compliance.

Leaders with a military background have a significantly decreased probability of complying. Military leaders have access to networks of military supporters, which makes military coups in support of the opposition less likely and makes repression easier to organize. According to our estimates, military background is associated with a 22.6 percentage point decrease in the probability of compliance.

International factors again have striking effects. We assumed that voters were concerned about whether their own country had ratified the CAT, because this is what the literature suggests provides protection of human rights. From the perspective of leaders, however, what is more important is the number of other countries that have ratified, because the CAT is enforceable against foreign citizens (including expatriate former dictators) regardless of whether their countries of origin have ratified it. Authoritarian leaders are frequently subject to punishment when they lose office (Chiozza and Goemans, 2011), so they usually flee abroad, and their ability to enjoy a comfortable retirement depends on legal immunity. As the number of CAT ratifiers has expanded and the human rights regime has become more legalized, dictators' outside options have narrowed. Our estimates indicate that this trend has made authoritarian leaders who hold elections significantly less willing to comply when they are defeated at the polls. Increasing the percentage of countries that have ratified the CAT from the average level of 24% by one standard deviation, to 53%, is associated with a decrease of 16.4 percentage points in the probability of compliance. By 2016, 140 countries, or 72.5% of UN member states, had ratified the CAT, representing 1.7 standard deviations, a level that is associated with an estimated decrease in the probability of compliance of 37 percentage points.

Finally, alignment with the United States strongly influences the choices of leaders, as it does those of voters. Leaders of countries with UN voting records similar to that of the United States are less likely to comply when they lose elections – again, presumably, because they are more likely to be shielded from international criticism and provided with material support that strengthens their capacity to repress the opposition. A decrease of one standard deviation in the distance between

a country's estimated ideal point and that of the United States is associated with a decrease in the probability of compliance of 11 percentage points. This is consistent with findings that U.S. foreign aid was associated with longer tenure of authoritarian leaders, at least during the Cold War (Bueno de Mesquita and Smith, 2009; Morrison, 2009; Bermeo, 2016).

So far, we have considered the non-strategic preferences of leaders and voters, but the strategic model also allows us to consider an indirect effect: voters may be deterred from voting to remove the incumbent leader if they anticipate that she will refuse to step down. The consequences of electoral manipulation are generally inferior to the outcome in which the leader is reelected legally, because manipulation may involve repression and civil conflict. Consequently, dictators may be able to masquerade as democratically elected leaders because the voters are afraid of the consequences if they prove otherwise. Leaders of poor, polarized countries that lack a recent experience with a peaceful transition of power, particularly if they have a military background and close relations with the United States, are likely to be reelected simply because the population fears the consequences if they lost. This recalls Machiavelli's famous advice to the prince: it is good to be loved, but it is better to be feared.

In order to test whether strategic voting plays an important role in our model, we compare it with an alternative model that allows for the same distribution of outcomes and employs the same covariates, but assumes that voters are not strategic. Consider a scenario in which voters vote sincerely, rather than strategically. This would be equivalent to a strategic model in which voters believe that the leader will step down for certain upon losing the election. In other words, at their decision node, the voters compare their utility from re-electing the leader to their utility from the leader stepping down, and ignore the possibility that the leader will try to remain in power after losing the election.[5] To compare our model with this non-strategic version, we conduct two non-nested model comparison tests proposed by Vuong (1989) and Clarke (2007). Both test results reject the null hypothesis that the non-strategic model fits the data as well as our strategic election

[5] Voters reelect the leader if $U(SQ) > U(SD)$.

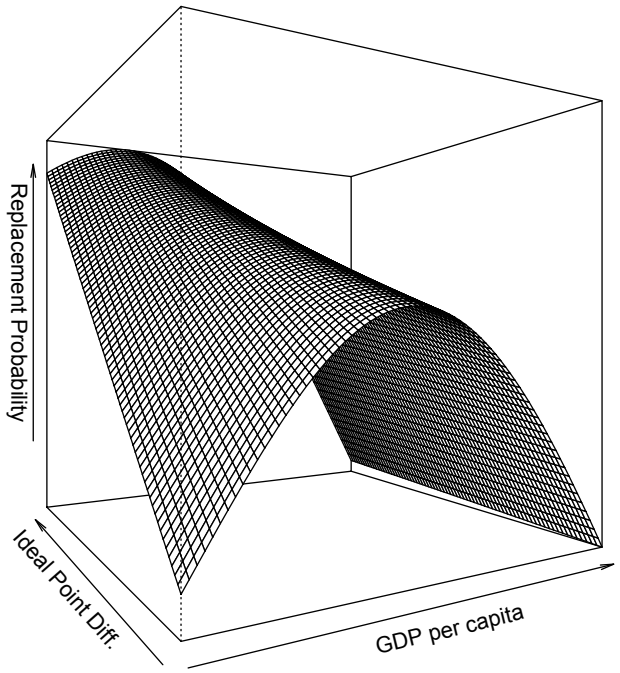

Figure 3: Non-monotonic effect of GDP per capita: Ethiopia in 1995

model to a high degree of confidence.[6] This is a direct test of the hypothesis that voters are strategic, and it indicates that voters are deterred from voting against the incumbent when they believe that she will not step down if she loses.

The comparative model test also allows us to explore when voting is strategic. Our estimates indicate that the strategic model outperforms the non-strategic version in approximately 60% of observations, but its advantage increases when the probability that the leader complies with the electoral outcome declines. The strategic model outperforms the non-strategic one in 63% of observations if the leader has a military background, compared to only 51% otherwise. In the poorest decile of countries, where GDP per capita is less than $1,000 per year in 2005 dollars, the strategic version outperforms the non-strategic one in 65% of observations. The strategic model outperforms the non-strategic model in 64% of observations in countries with the lowest quartile of ideal point differences with the United States, compared to 59% of observations in other countries.

[6]For our comparison, the Vuong test statistic takes a value 3.04, which implies a p-value of .002. Similarly, the Clarke test results in a value of 541 (out of 932 observations), which implies a p-value less than .0001.

The effect of strategic interaction on vote choices becomes clear when we plot the marginal probabilities of outcomes. Because voting decisions take the leader's strategy into account, they depend on the covariates that enter the leader's utility function. As a result, the reduced form equation to be estimated includes interaction effects between the variables that affect voter choice and the variables that affect the leader's choice. In general, these sorts of strategic interactions lead to non-monotonic effects of covariates on outcome probabilities. Figure 3 provides an illustration of how the probability that the leader is replaced changes as a function of GDP per capita and ideal point distance from the United States. All other variables are fixed, and we chose the values they took during the Ethiopian general election in 1995, so this can be thought of as a counterfactual analysis of how that election would have been different if we were able to vary two parameters.

Focus first on the foreground of the figure, where the ideal-point distance from the United States is small. In this region, the probability of replacement is very low when GDP per capita is near its minimum, because the leader refuses to step down when she loses an election, and is also very low when GDP per capita is near its maximum, because the voters are highly likely to reelect the leader. The probability of replacement is maximized at an intermediate level of national wealth. The ideal point distance from the United States increases as we move towards the back of the figure, and the replacement probability increases because defeated leaders are more likely to step down. This effect is very steep when GDP per capita is low, because in that case voters are likely to vote the incumbent down, so the leader is frequently confronted with the choice of stepping down. The effect of ideal-point distance becomes almost imperceptible when GDP per capita is high, however, because voters rarely vote the leader out of office. The most democratic counterfactual Ethiopia is in the right-rear corner, where GDP per capita and ideal-point distance are both high, and the least democratic case is in the left-front corner, where both are low. The predicted probability of leader replacement is almost identical in these two cases; all of the action occurs in between.

So far, we have discussed the estimates from a single specification of our model, but one of the advantages of the method we propose is its flexibility. Applied researchers can customize the

specification to ask specific questions about the conditions that promote democracy. For example, there is a long-standing debate about whether parliamentary or presidential systems are superior in terms of consolidating democracy. Partisans of presidential systems point to the policy stability provided by veto players, while proponents of parliamentary systems argue that they provide greater responsiveness, avoid deadlock and allow for rapid replacement of unpopular leaders. To test these alternative views, we include an indicator variable for parliamentary/presidential systems in the leader's utility for stepping down. The results are in the appendix. The results indicate that leaders in parliamentary systems are substantially more likely to voluntarily step down after losing an election, so by our definition, parliamentary systems appear to facilitate democratic consolidation. The marginal effect of a parliamentary system is an increase in the estimated probability that a defeated leader complies by stepping down of 0.13 (0.04, 0.22), or about a 20% increase in the odds of compliance over the baseline used in Table 2.

2 Measuring Democracy

According to our limited definition of democracy, countries that do not hold competitive elections are not democratic; those in which elections are followed by a transfer of power are observed to be democratic when the transfer takes place; and those in which incumbent parties are returned to power in competitive elections are democratic with the estimated probability that the incumbent would have stepped down, conditional on losing the election. We code the case of observable non-democracy zero and the case of observable democracy one. For the remaining elections, we estimate the conditional probability of stepping down from our strategic model with partial observability. Between competitive elections, we extrapolate the conditional probability of stepping down using the same model and the prevailing values of the covariates, so our measure varies annually between elections. Between 1950 and 2008, the variable takes an average value of 0.47, with a standard deviation of 0.40.

Figure 4 plots the average level of democracy over time. The data clearly indicate the second and third waves of democratization that have been identified using other measures of democracy.

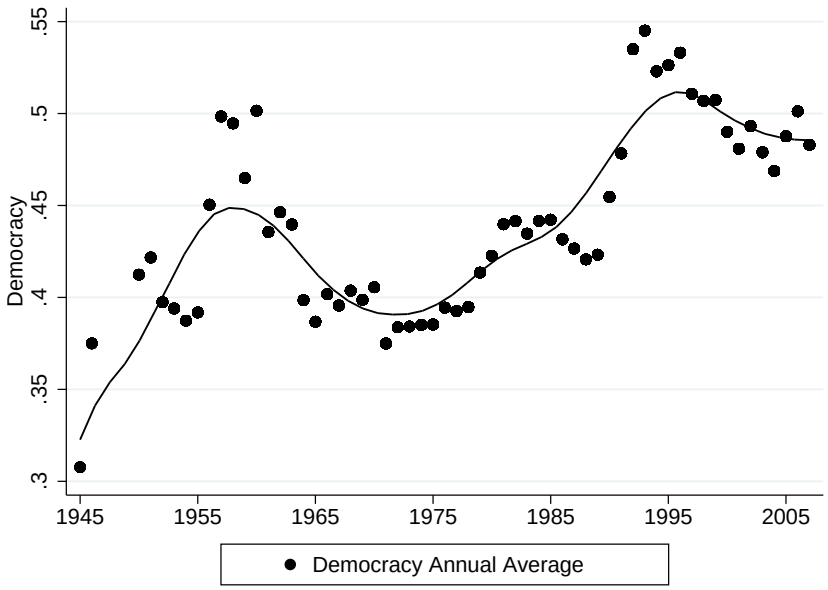

Figure 4: Waves of Democracy

According to our definition, democracy reached a local peak in the late 1950s, declined precipitously in the 1960s as many newly independent post-colonial regimes underwent authoritarian transitions, and began to trend upwards again in the late 1970s. Democracy surged after the collapse of Communism in Eastern Europe in 1990 and of the Soviet Union in 1992, but retreated modestly around the turn of the millenium.

Our measure is highly correlated with existing measures of democracy, including the Przeworski et al. (2000) democracy measure (ACLP), the Polity score, the Freedom House political rights and civil liberties indices, and the V-Dem high-level indicators. The correlation between our measure and ACLP is .78; between our measure and *Polity* is .80; between our measure and the inverse of the Freedom House measure of *political rights* is 0.78; and between our measure and the inverse of *civil liberties* is 0.76.[7] Since empirical work often uses dichotomized measures of democracy, we also compared a dichotomous version of our measure to ACLP and to a dichoto-

[7]The reported correlation with ACLP is polychoric, because ACLP is a dichotomous indicator. Taking into account uncertainty in our measure, the 95% confidence interval for the correlation with ACLP is (.782, .786), and the 95% confidence interval for the correlation with Polity is (.794, .798).

mous version of Polity. The tetrachoric correlation between our measure and ACLP is maximized when the cut-point 0.5 is used for dichotomizing our measure, and is 0.85. The tetrachoric correlation between our measure and Polity is maximized when we use the cut-point 0.5 for our measure and 7 (Polity>6) for Polity, and is 0.87 (see appendix).

The V-Dem project has generated five main variables that are designed to capture distinct features of democratic governance. *Electoral democracy* is conceptually closest to our definition, but includes features that we do not such as the inclusiveness of the electorate. *Liberal democracy* measures respect for human rights, rule of law and the rights of minorities. *Deliberative democracy* measures the quality of democratic governance, while *Participatory democracy* captures access to government and participation. Finally, *Egalitarian democracy* measures economic equality. Each of these measures represents a latent variable derived from surveys of country experts using a Bayesian IRT model. Despite the broad conceptual differences across these indicators, the correlations between our measure and the V-Dem variables range from 0.76 for *Electoral democracy* to 0.69 for the most distinct concept, *Egalitarian democracy*.[8] The fact that our spare definition of democracy is strongly correlated with these broader measures is consistent with the hypothesis that credible elections are the key to democracy.

2.1 Out-of-Sample Prediction

We use out-of-sample prediction to address the possibility of over-fitting. We exclude a randomly-drawn 20% of the sample from our estimation sample and replicate our analysis, generating predicted Democracy measures for the excluded observations from the estimation sample. When we compare the out-of-sample predictions to the predictions for the same observations from the model

[8]The bivariate correlations are 0.76 for *Electoral democracy*; 0.74 for *Liberal democracy*; 0.73 for *Deliberative democracy*; 0.76 for *Participatory democracy*; and 0.69 for *Egalitarian democracy*. The V-Dem variables are themselves correlated at a level of 0.95-0.98, which suggests that the various dimensions of democratic governance that they measure are closely related. A principal-components analysis of the five main V-Dem measures indicates that all of them load positively (and almost identically) on a single dimension, which in turn explains 97% of the variation in the five indicators.

using the complete data set, 95% of the correlations in repeated runs exceed 0.99. (See appendix.) This indicates that our estimation results are stable across randomized sub-samples of the data, and it is very unlikely that we have over-fitted the data. When we regress our estimates from the full sample on our estimates from the out-of-sample estimation, we retrieve a coefficient very close to one, indicating the absence of forecasting bias.

In addition, out-of-sample prediction allows us to explore how the conditions for democratic consolidation have changed over time. We have emphasized that the problem of false elections has become more important since the end of the Cold War, as the number of authoritarian regimes that conduct elections has increased. We are able to quantify this claim by using out-of-sample prediction, and we can also learn something about how the effects of the covariates of democratic compliance have changed. We split our sample at 1990 and ask how well the historical data are able to predict democratic outcomes after the Cold War.[9] The results are in the appendix. Some coefficients change significantly when we use only the historical data, and the correlation between the out-of-sample democracy scores and the full-sample measures drops to 0.91. Most importantly, excluding the more recent data leads to a significant bias: on average, the out-of-sample predictions are substantially more optimistic. The mean level of democracy after 1990 is estimated to be 0.62 using the full sample, and this figure jumps to 0.72 in the out-of-sample predictions. On average, the full-sample estimates are only 86% as optimistic as the out-of-sample estimates. This implies that something important has changed about the process of democratic consolidation. Indeed, one might conclude that observers at the end of the Cold War were justified in being more optimistic about the fate of democracy than observers today, because the best data available to them predicted substantially more favorable outcomes.

The most substantial over-predictions of democracy occur in sub-Saharan Africa and in the Russian Federation in the years of Vladimir Putin's first administration. The full-sample estimates

[9]Dating the endpoint of the Cold War is almost as controversial as dating its beginning, but a convenient reference point is the Malta Summit in December 1989, when George H. W. Bush and Mikhail Gorbachev declared that it was over.

for the Russian Federation for 2000-2004 range between 0.36 and 0.44, which is close to the consensus of country experts at the time. Putin was a leader who probably would not conduct a fair election or step down if he lost one, and his regime did not meet the 50% threshold for a democracy, but the political system was still considerably more competitive than it had been under the Soviet Union. The out-of-sample estimates, however, range from 0.70 to 0.76, which is solidly democratic territory. They imply that early in his rule, Putin would have been highly likely to conduct fair elections and step down if he lost. Comparing the coefficients of the variables that generate these different predictions in the two samples is instructive. The full sample places a higher weight on former military background, and Putin is coded as having such a background because he was a former KGB Colonel. The post-1990 experience has taught us to be more skeptical of the democratic credentials of former military officials, and that information makes our estimates of Putin's regime more pessimistic. The full sample places a lower weight on urbanization than the Cold War sample, and a higher weight on the proportion of countries that have ratified the CAT. The Russian case has played an important role in tempering expectations about the spread of democracy, and this exercise suggests that the theoretical lessons we should learn from it are that democratic consolidation is more fragile than we previously believed and that the personal characteristics of leaders are more important.

2.2 Democracies with Hegemonic Parties: Japan and India

The most difficult regimes to classify according to our definition are those that appear to conduct competitive multiparty elections, but in which one party consistently wins. Przeworski et al. (2000) adopt a conservative solution and apply a retrospective coding rule: if a hegemonic party ever gives up power after losing an election, it is coded retrospectively as a democracy for its entire history; if it ever repudiates an electoral defeat, it is coded retrospectively as a dictatorship for its entire history. They acknowledge that this "is not a very satisfactory solution" (Przeworski et al. 2000, 24), but they use this rule to code India and Japan as democracies from the moment of independence.

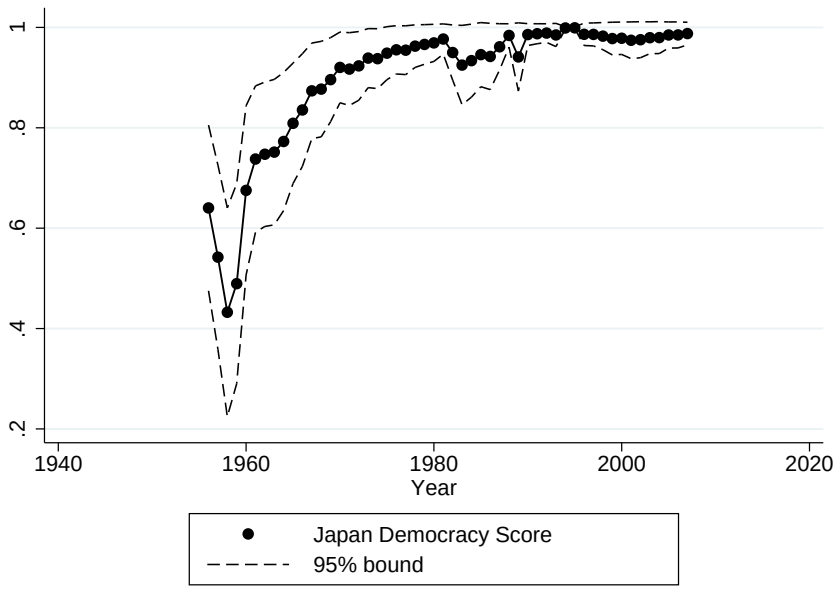

Figure 5: Japan's Democracy over time

Our estimates offer a middle-ground approach for such cases. They suggest that there was a substantial degree of uncertainty about democratic institutions in both countries, particularly in the early years. The Japanese political system was competitive under the American occupation, which ended in 1952, and involved several transitions of power, but Japan was ruled continuously by governments led by the Liberal Democratic Party (LDP) from its formation in 1955 until 1993. By the end of this period, our estimates (see Figure 5[10]) confirm the conventional wisdom that the Japanese government was democratic and the LDP was continuously reelected because it enjoyed broad public support, but the estimated probability that the LDP would actually give up power if it lost an election did not come to exceed 90% until the 1980s.[11]

[10]For presentational purposes, Figures 5 and 6 plot raw democracy scores and their confidence intervals from the estimation process, before we code the country democracy with certainty when the incumbent steps down after elections.

[11]V-Dem estimates of *Electoral democracy* also show some change over time, but the changes are small and the resulting graph is much flatter. This suggests the possibility of ex post evaluation bias, since the V-Dem surveys were conducted long after the period in which we find Japan had consolidated its democracy.

Our estimates show a sharp drop in probabilistic democracy in 1957, when Nobusuke Kishi became Prime Minister, and the estimated probability of conceding an electoral defeat remains below 50% until he leaves office in 1960. This drop is driven largely by Kishi's military background, and in this case the leader's personal characteristics appear to justify skepticism. Kishi was Minister of Commerce and Industry in the Tojo government during the Second World War, responsible for wartime mobilization, and was subsequently imprisoned as a war criminal. Before the war, he had been an outspoken nationalist and had been responsible for the notorious economic management of Manchuria. As Prime Minister, Kishi attempted to rebuild Japan's international influence and pushed a deeply unpopular security treaty with the United States through parliament, replacing the one adopted during the U.S. occupation. This provoked the largest post-war political demonstrations in Japan, which eventually compelled him to resign. At the same time, other U.S. allies in East Asia were authoritarian (Taiwan, South Vietnam) or moving in that direction (Korea). Jung-hee Park, for example, led a coup in South Korea in 1961, was elected president in 1963, and declared martial law in 1972 when he lost the next election. It appears plausible that Kishi might have led Japan in the same direction, had circumstances been different. By the 1980s, however, Japanese society had been transformed by rapid economic growth, and the incentive to play by the rules of electoral democracy had become compelling.

Like Japan, India was ruled continuously by various incarnations of the Congress Party for the first several decades after its independence in 1947. Congress was undeniably popular as the organization that had led India's quest for independence from Britain. In addition, it enjoyed tremendous organizational advantages and was the only truly national party. It faced stiff opposition from various ethnic parties in particular states, but no single party was able to mount an effective national challenge. Despite Congress' popularity, we estimate that for most of India's first three decades of independence the probability that an incumbent leader would voluntarily step down after losing an election hovered around 50%. India was impoverished, with high levels of illiteracy and subsistence farming, so electoral manipulation was easy to carry out and difficult

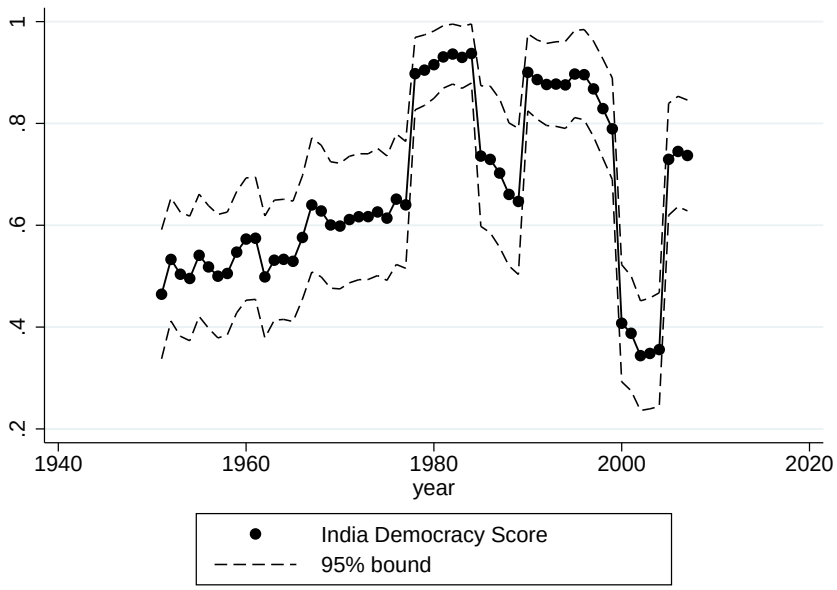

Figure 6: India's Democracy over time

to prove. Congress was able to count on substantial rural support, so it took a long time for the nascent urban middle class to begin to feel a sense of efficacy.

In three elections, 1951, 1962 and 1971, our estimates indicate that the Congress Party benefited from strategic voting. In these cases, we estimate that voters ranked replacing the government over reelecting it if Congress could be persuaded to concede electoral defeat, but regarded a disputed election as worse than maintaining the status quo. We estimate relatively low probabilities that Congress would have acquiesced in defeat (46.5% in 1951, 49.9% in 1962, and 61.1% in 1971), which drove down the expected utility of voting against the ruling party enough to ensure its victory. In short, our model suggests that India appeared to be more democratic than it really was because voters internalized the risk of political destabilization that could ensue if Congress lost, and voted for Congress because they did not trust it to step down if it were defeated.

The most dramatic event to underscore the weak rule of law in India was the declaration of martial law from 1975 to 1977. The crisis was provoked when an Indian court ruled Prime Minister Indira Gandhi's election to parliament to be illegal on grounds of electoral malpractice for using government resources for campaigning, effectively stripping her of her position as Prime Minis-

ter. She refused to resign, and when the opposition protested, she asked the President of India to declare martial law, and then proceeded to rule by decree, imprison opposition leaders, dissolve the opposition-led governments of two Indian states, and impose press censorship. She ruled under the state of emergency for two years, and then called elections that she expected to win. The opposition slogan was that this was the last chance to choose "democracy or dictatorship." It was a surprise when the opposition prevailed, and Indira Gandhi stepped down. Our estimates of Indian probabilistic democracy increase substantially after 1977, because lagged alternation in office predicts compliance with electoral outcomes; in conventional terms, precedents contribute to the norm of compliance. However, India remained a very poor country, with hundreds of millions of subsistence farmers living below the international poverty line, and the volatility of its democracy estimates reflects the fact that it is anomalous for a country this poor to be a stable democracy.

2.3 Comparative Applications

We perform two comparative tests of our measure of democracy against two existing measures, Polity and ACLP. The results and discussion are in the appendix. First, we compare the predicted effects of joint democracy on international conflict. The democratic peace result holds in several specifications for all three measures considered: democratic pairs are less likely to be engaged in conflict than other pairs of states. However, models using probabilistic democracy find significant results using country fixed effects, because probabilistic democracy exhibits more over-time variation than existing approaches. In addition, an interesting difference in results arises in an interactive model, which allows us to differentiate between the decisions of democratic and non-democratic conflict initiators. In the interactive model using probabilistic democracy, the result is driven by the disinclination of democratic initiators to attack democratic targets, which is consistent with the canonical interpretation: democracies are disinclined to fight with other democracies. In the interactive models using Polity and ACLP, however, the lower occurrence of conflict between two democracies is driven primarily by the increased tendency of non-democratic initiators to target democratic states, which is inconsistent with the standard interpretation.

Second, we compare the predicted effects of democracy on economic growth, conditional on GDP per capita. According to all three measures, democracy reduces economic growth, but only in relatively wealthy countries. The differences between the measures regard the level of wealth that is necessary for democracy to impede growth. According to probabilistic democracy, only high-income countries (GDP per capita of $12,000 or more in 2005 dollars) are subject to the effect. According to Polity and ACLP, in contrast, middle-income countries also grow more slowly under democracy. These differences imply different inferences about what mechanisms might be at work to cause the effect and have different implications for what kinds of institutions are most likely to help poor countries to develop their economies. Thus, for example, our results would not encourage proponents of authoritarianism in Chile in the 1980s or Russia after the end of the Cold War.

3 Conclusions

Democracy is a critical concept in political science, and measures of democracy are involved in most quantitative studies in comparative politics and international relations. Nevertheless, we are unsettled both on conceptual and empirical grounds. Conceptually, we have no clear consensus on what we mean by democracy. Is it an attribute of institutions, of a political equilibrium, or of government policies? Is it procedural or substantive? Empirically, how do we measure such an elusive concept? Are we satisfied with multi-dimensional indices that combine disparate ingredients in arbitrary proportions?

We propose a minimalist definition because of its conceptual clarity: a government is democratic if the leader steps down conditional on losing an election (Przeworski, 1991; Przeworski et al., 2000). We argue that electoral accountability captures a necessary condition for democratic governance, contributes to beneficial features often associated with democracy, and remains an important dimension of variation in contemporary regimes. In some cases, this is not a quantity that can be directly measured, however, because it involves estimating a counterfactual. We provide a straightforward algorithm for estimating democracy that can be implemented as a structural model, so there is no conceptual blurring between our definition and our measure.

There are lessons to be learned from the estimation process. Estimating democracy allows us to learn which conditions facilitate peaceful power transitions; to test the hypothesis that voters succumb to electoral intimidation; to use out-of-sample prediction of democracy scores to pose questions about how the process of democratic consolidation has changed over time; and to customize the estimation process to draw new lessons and test new hypotheses.

The estimation results provide several important substantive findings. On the voters' side, we find evidence consistent with economic voting, a rally effect during international disputes, and voter intimidation in countries that have not signed the Convention Against Torture (CAT). On the incumbents' side, we find several clues to what makes elections work. The strongest predictor of a peaceful transfer of power is what happened in the last election, which confirms the intuition that democracy becomes consolidated through experience of peaceful transitions. The incumbent is much more likely to step down voluntarily in a wealthy country (modernization), and in one where the political system is not excessively polarized (political culture). In addition, however, three striking findings concern the leader's personal incentives. First, leaders with military experience are less likely to step down when they lose elections, presumably because they are advantaged at organizing repression and surviving coup attempts. Second, leaders that have close relations with the United States are less likely to step down, presumably because they do not have to fear international reprisals. Third, leaders become more reluctant to step down as the number of countries that have ratified the CAT increases, presumably because this makes their retirement abroad less secure. These results suggest that democracy is a feature of the international system as much as of the domestic political system. Because of the limits of what we can observe, a structural model helps us identify these hypotheses and distinguish incumbents' from voters' choices.

Further, our model allows for the possibility that voters offer support strategically in order to avoid conflict with the incumbent, and we find that voters do in fact vote strategically in favor of governments that they do not support. We are able to use our structural model to conduct a direct test of this hypothesis by comparing the fit of the full model to the fit of a model that is restricted to prevent voters from conditioning their strategies on their expectations about whether the leader

will step down if defeated. We reject the null hypothesis of no electoral intimidation with a high degree of confidence, and we can show that it is most frequent in poor countries and those with ex-military leaders. We argue that this kind of electoral intimidation is an important part of the toolkit of authoritarian leaders, and helps to explain the puzzle of why so many authoritarian regimes conduct elections.

Out-of-sample prediction confirms that our model is predictive, and is not simply over-fitting the data. Substantively, estimation with sub-samples of the data allows us to explore how the conditions for peaceful transitions of power have changed over time. By splitting the sample at the end of the Cold War, we find that leaders with military backgrounds became less likely to give up power after the Cold War, that urbanization has weaker effects of consolidating democracy after the Cold War, and that the effect of prior democratic transitions is weakened after the Cold War. These changes drive the fact that the Cold War-era estimates significantly over-predict democracy during the post-Cold War period. We conclude that the optimism often expressed about democracy in the early post-Cold War period appeared to be justified by the historical evidence available up to that point, although it appears naive in retrospect.

It is a strength of our approach that it does not depend upon an irrevocable set of modeling choices. The estimation procedure is flexible and can include additional covariates or make different identification assumptions. The code and data are publicly available, and skeptics are invited to customize the estimation to answer their own questions. As an example, we added a variable for presidential/parliamentary system to our initial specification. The results confirm the intuition that parliamentary systems are more conducive to democratic consolidation than presidential systems.

We compare the estimated values of probabilistic democracy to established measures such as Polity, ACLP, Freedom House and V-Dem scores, and we find that they are highly correlated with all of these measures. Our measure comes with a few desirable properties, including conceptual clarity, estimates of uncertainty, and avoiding potential expert bias. The fact that a minimalist definition of democracy generates estimates that accord so well with a range of multi-dimensional measures confirms Przeworski et al.'s intuition that electoral accountability is close to the core of

democratic governance. Our measure helps capture this concept in some of the ambiguous cases due to the dominance of hegemonic parties. For instance, we conclude that both Japan and India were less democratic before the 1980s than is generally supposed.

Like the V-Dem scores, our measure is estimated, and therefore comes equipped with estimates of uncertainty. This makes it possible for subsequent analyses to correct for measurement error by sampling from the distribution of the estimates. Unlike V-Dem, our measure avoids potential expert bias by not relying on expert assessments. Experts coding contemporary cases are embedded in cultural and professional contexts that may influence their judgments; those coding historical cases may have access to a common set of case histories and influential interpretations. Our approach allows us to avoid such subjective judgments.

To illustrate the use of our measure, the appendix provides estimates of the effect of joint democracy on international conflict and of the effect of democracy on economic development. We are able to confirm the findings that democracies are unlikely to fight wars against each other and that democracies grow more slowly than non-democracies, but with important caveats. The findings using our measure fit the logic of the democratic peace hypothesis better than those using some of the existing measures, because we find that it is the behavior of democracies that is affected by joint democracy, rather than the behavior of autocratic initiators that is influenced by democratic governance in potential targets. The substantive import of the effect of democracy on development is also different using our measure: unlike some of the existing measures, the effect applies only to upper-income countries, excluding middle-income countries.

As long as democracy has to be measured, it will be measured with error in some cases. The coding rules for the commonly-used measures of democracy in political science do not always agree with each other, and expert coding may lead to different pitfalls and face difficulties in some cases. As an alternative, an estimation procedure is transparent and more amenable to improvement as knowledge accumulates. In addition, the process of estimating the model creates new knowledge and better understanding of the means by which elections are manipulated in authoritarian countries.

References

Bailey, Michael A, Anton Strezhnev and Erik Voeten. 2017. "Estimating Dynamic State Preferences from United Nations Voting Data." *Journal of Conflict Resolution* pp. 430–456.

Barber, Benjamin R. 2003. *Strong Democracy: Participatory Politics for a New Age*. Univ of California Press.

Bas, Muhammet A, Curtis S Signorino and Taehee Whang. 2014. "Knowing ones future preferences: A correlated agent model with Bayesian updating." *Journal of Theoretical Politics* 26:3–34.

Bas, Muhammet A and Randall W Stone. 2014. "Adverse Selection and Growth under IMF Programs." *The Review of International Organizations* 9(1):1–28.

Bermeo, Sarah Blodgett. 2016. "Aid is not Oil: Donor Utility, Heterogeneous Aid, and the Aid-Democratization Relationship." *International Organization* 70(01):1–32.

Bueno de Mesquita, Bruce and Alastair Smith. 2009. "A Political Economy of Aid." *International Organization* 63(02):309–340.

Bueno de Mesquita, Bruce, Alastair Smith, Randolph M. Siverson and James D. Morrow. 2003. *The Logic of Political Survival*. MIT press.

Calvin, Jean. 1536. *Institutes of the Christian Religion: 1536 Edition*. Vol. 1 Wm. B. Eerdmans Publishing.

Chiozza, Giacomo and Hein Erich Goemans. 2011. *Leaders and International Conflict*. Cambridge University Press.

Clarke, Kevin A. 2007. "A Simple Distribution-free Test for Non-nested Model Selection." *Political Analysis* 15(3):347–363.

Coppedge, Michael, John Gerring, Staffan I Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, Michael Bernhard, M Steven Fish, Adam Glynn, Allen Hicken et al. 2017. "V-dem dataset v7.".

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

De Tocqueville, Alexis. 2003. *Democracy in America*. Vol. 10 Regnery Publishing.

Debs, Alexandre and Hein E Goemans. 2010. "Regime Type, the Fate of Leaders, and War." *American Political Science Review* 104(03):430–445.

Downs, George W and David M Rocke. 1995. *Optimal Imperfection?: Domestic Uncertainty and Institutions in International Relations*. Princeton University Press.

Hamilton, Alexander, James Madison, John Jay and Lawrence Goldman. 1788. *The Federalist Papers*. Oxford University Press.

Hyde, Susan D. 2011. *The Pseudo-Democrat's Dilemma: Why Election Observation Became an International Norm*. Cornell University Press.

Jaggers, Keith and Ted Robert Gurr. 1995. "Tracking Democracy's Third Wave with the Polity III Data." *Journal of Peace Research* 32(4):469–482.

Lewis, Jeffrey B. and Kenneth A. Schultz. 2003. "Revealing preferences: Empirical estimation of a crisis bargaining game with incomplete information." *Political Analysis* 11(4):345–367.

Locke, John. 1689. *Second Treatise of Government: An Essay Concerning the True Original, Extent and End of Civil Government*. John Wiley & Sons.

Marshall, Monty and Keith Jaggers. 2007. "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2006." *Center for Systemic Peace Data Manual* .

Morrison, Kevin M. 2009. "Oil, Nontax Revenue, and the Redistributional Foundations of Regime Stability." *International Organization* 63(01):107–138.

Mueller, John. 1994. *Policy and Opinion in the Gulf War*. University of Chicago Press.

Poirier, Dale J. 1980. "Partial Observability in Bivariate Probit Models." *Journal of Econometrics* 12(2):209–217.

Przeworski, Adam. 1991. *Democracy and the Market: Political and Economic Reforms in Eastern Europe and Latin America*. Cambridge University Press.

Przeworski, Adam and James Raymond Vreeland. 2000. "The Effect of IMF Programs on Economic Growth." *Journal of development Economics* 62(2):385–421.

Przeworski, Adam, Michael E. Alvarez, Jose Antonio Cheibub and Fernando Limongi. 2000. *Democracy and Development: Political Institutions and Well-being in the World, 1950-1990*. Cambridge University Press.

Reinhart, Carmen M and Kenneth S Rogoff. 2009. *This Time is Different: Eight Centuries of Financial Folly*. Princeton University Press.

Riker, William H. 1982. *Liberalism against Populism: A Confrontation between the Theory of Democracy and the Theory of Social Choice*. Waveland Press.

Rozenas, Arturas. 2016. "Office Insecurity and Electoral Manipulation." *The Journal of Politics* 78(1):232–248.

Signorino, Curtis. 2002. "Strategy and Selection in International Relations." *International Interactions* 28(1):93–115.

Signorino, Curtis S. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *American Political Science Review* 93(2):279–297.

Simmons, Beth A. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics*. Cambridge University Press.

Stokes, Susan C, Thad Dunning, Marcelo Nazareno and Valeria Brusco. 2013. *Brokers, Voters, and Clientelism: The Puzzle of Distributive Politics*. Cambridge University Press.

Treier, Shawn and Simon Jackman. 2008. "Democracy as a Latent Variable." *American Journal of Political Science* 52(1):201–217.

Verba, Sidney and Gabriel A Almond. 1963. "The Civic Culture: Political Attitudes and Democracy in Five Nations.".

Vogt, Manuel, Nils-Christian Bormann, Seraina Rüegger, Lars-Erik Cederman, Philipp Hunziker and Luc Girardin. 2015. "Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Data Set Family." *Journal of Conflict Resolution* 59(7):1327–1342.

Vreeland, James Raymond. 2003. *The IMF and Economic Development*. Cambridge Uni. Press.

Vreeland, James Raymond. 2008. "Political Institutions and Human Rights: Why Dictatorships Enter into the United Nations Convention Against Torture." *International Organization* 62(01):65–101.

Vuong, Quang H. 1989. "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses." *Econometrica: Journal of the Econometric Society* pp. 307–333.