Estimating Smooth Country-Year Panels of Public Opinion

Christopher Claassen*

School of Social and Political Sciences
University of Glasgow

July 19, 2017

Abstract

At the microlevel, comparative public opinion data are abundant. But at the macrolevel, the level where many prominent hypotheses in political behavior are believed to operate, data are scarce. In response, we develop a Bayesian dynamic latent trait modeling framework for measuring smooth country-year panels of public opinion even when data are fragmented across time, space, and survey item. We derive six models based on this framework, and test their accuracy using data on support for democracy. Our best model is reasonably accurate, with predicted responses that deviate from the true response proportions in a hold-out dataset by six percentage points. We show, furthermore, that our smoothed country-year estimates of support for democracy have both construct and convergent validity, with spatiotemporal patterns and associations with other covariates that are consistent with previous research.

*I greatly appreciate the comments provided by Devin Caughey, Roberto Stefan Foa, Duncan Lee, Anthony J. McGann, Jamie Monogan, and Richard Traunmüller on earlier versions of this paper.

1. Introduction

Social scientists are awash in public opinion data. Over a dozen cross-national survey projects are now in existence, regularly asking nationally representative samples in all continents and regions their opinions on a diverse range of social and political topics. Few countries haven't been surveyed at one time or another, and many countries have been polled numerous times, sometimes by several of these survey projects. At the dawn of cross-national public opinion research, when Almond and Verba (1963) completed their pioneering five-country study, researchers could hardly have dreamed of such a vast trove of public opinion survey data.

Yet by another standard, public opinion data are scarce. Many of the classic hypotheses in political behavior actually imply aggregate, country-level relationships between opinion and political outcomes. For example, social trust bolsters the quality of governance (Putnam 1993), policy preferences shape policy choices (Stimson 1991), political intolerance leads to the repression of dissent (Sullivan, Piereson, and Marcus 1982), and support for democracy help sustain a democratic regime (Lipset 1959). Yet at the aggregate level, a typical survey sample of one or two thousand respondents diminishes to a single data point. Thus, although we may have data from millions of respondents worth for any particular topic in public opinion, we might only have a few hundred country-level data points.

Moreover, while such aggregate data may be sufficient to assemble a cross-section of a hundred or so countries which have recently been surveyed, comparing Sweden to Slovakia to Somalia in a cross-sectional fashion affords us little ability to test causal hypotheses. Instead, the fact that that major public opinion projects have been ongoing since the 1990s, if not before, offers a tantalizing possibility of measuring a panel of public opinion that varies across space and time. Country-year panels of public opinion are not only of descriptive interest, they would also allow scholars to incorporate public opinion in studies of comparative political behavior and comparative political economy; in some cases, for the first time.

Unfortunately, however, aggregate public opinion data are not distributed neatly or evenly across space and time. Cross-national surveys are clustered in certain places and times, with fragmented time series and large spatial gaps. To make matters considerably worse, major public opinion concepts are typically measured in multiple ways, with the wording of questions or the nature of response sets varying across survey projects. Any would-be panel of aggregate public opinion thus would appear to have to limit itself to a single survey question. As a consequence, any panel that is assembled out of available data will be highly fragmented, with sparse coverage over space and time.

As such, this paper proposes a method for estimating smooth panels of aggregate public opinion. The idea is to harness existing data, adjust for the bias induced by question wording effects, partially pool across space, and smooth across time. While a number of scholars have developed methods for smoothing single-country time series of aggregate opinion (e.g Beck 1989; McGann 2014; Stimson 1991; Voeten and Brewer 2006), none have focused as yet on *cross-national panels* of opinion. The primary contribution of this paper is the development and validation of such a method. This method will be of interest to scholars in the fields of comparative political behavior and political economy who would benefit from access to country-year panels of opinions on policy mood, social values, political culture, and so on. A second contribution of this paper is then to use the model to measure, for the first time, a smooth panel of opinion on support for democracy for 127 countries and 24 years.

2. Existing Research on Smoothing Aggregate Public Opinion

Political scientists have long been interested in smoothing estimates of aggregate public opinion. A pioneer in this field is Stimson (1991), who estimated "policy mood", or ideology, in the United States over the period from 1956 to 1988. To accomplish this goal, he developed an ingenious dyad-ratios algorithm. This algorithm rests on the realization that while the level of respondent agreement varies idiosyncratically across survey items, the change over time in levels of respondent agreement can be compared across items. The dyad ratios algorithm thus uses the ratio of change over time to standardize survey items. These ratios are then combined using a factor analysis type procedure which weighs each item's ratio of change by the degree to which it correlates with the latent variable. Finally, the estimates are smoothed over time using an exponential smoother.

The dyad ratios algorithm has been phenomenally popular. Stimson and colleagues have used it in several major studies of public opinion (e.g. Stimson, Mackuen, and Erikson 1995; Erikson, Mackuen, and Stimson 2002) and scholars continue to use it to this day to estimate smooth time-series of aggregate opinion (e.g. Baumgartner, De Boef, and Boydstun 2008). However, two years before Stimson, Beck (1989) provided an alternative, arguably more theoretically-grounded, approach to smoothing aggregate opinion. He described a model of dynamic opinion that included a Kalman filter to smooth over time and a measurement model to combine multiple items into one opinion series. Indeed, Beck went even further by showing how the latent opinion estimates could be modeled using a set of covariates. Although he managed to fit and run such a sophisticated model using GAUSS software on a "386-based microcomputer", it appears that Beck was somewhat ahead of his time. It was the much simpler dyads ratio algorithm that became popular.

In recent years, methodologists interested in measuring latent aggregate opinion have moved back to the kind of dynamic measurement models advocated by Beck (e.g. Green, Gerber, and De Boef 1999). This has been made possible, in part, because of the vast increases in computing power at our disposal. It is also, however, due to the rise of Bayesian methods, which not only include prior information to help estimate and identify complex models, but also provide a highly intuitive framework for understanding hierarchical and dynamic models. For example, Jackman (2005) constructs a dynamic Bayesian model of elections polls. He models observed polling marginals as true opinion plus random error, with true opinion then allowed to evolve over the period of the election campaign using a random walk error process. Finally, true opinion is adjusted for "house-effect" bias induced by the methods of particular polling companies.

Voeten and Brewer (2006) present a model that is also in this lineage. They estimate a time series of US public approval of the 2003 war in Iraq. While they do not include sampling error, they follow Beck in adding a measurement model, as they have data from 25 different survey items. In particular, they model observed opinions as a function of true opinion adjusted for item bias and weighted by an item-specific factor loading, plus some random error. They also use Bayesian MCMC methods to estimate their model.

While Jackman (2005) and Voeten and Brewer (2006) use linear models of the proportion of respondents who offer a particular opinion, Linzer (2013) takes a different approach. He uses a binomial specification to model the *number* of respondents supporting a particular party. This neatly allows allows sampling error to be neatly included in the estimates but also allows survey items where almost all or almost no respondents agree (and thus proportions close to 0 or 1) to be more accurately modeled. McGann (2014) follows suit, but also includes a two-parameter item response theory (IRT) measurement model to capture item intercepts and slopes. He further includes a beta prior on the binomial probability parameter to capture the overdispersion in survey data introduced by idiosyncrasies peculiar to survey data such as varying survey modes, methods of sampling, and so on.

This literature has focused on estimating an opinion time series within a single country. No researchers have attempted to extend these models to measure opinion across countries as well over time. If data were abundant, this would be simple. One could simply apply the selected model to each country in turn. Unfortunately, aggregate opinion data are generally not plentiful enough to do so. In particular, in many cases there is simply not enough data to reliably estimate both the country-specific latent effects as well as the item-specific effects – such as item intercepts and slopes – that we would want to include.

Although no scholars have modeled opinion across country and time, a recent paper by Caughey and Warshaw (2015) outlines a model for estimating latent opinion across subnational geographic districts. To do so, Caughey and Warshaw combine a binomial IRT model, a dynamic linear component, and the method of multilevel regression and poststratification (MRP; Park et al (2004)). Their model allows one to disaggregate survey data by subnational district and estimate latent opinion within that district. It also allows for multiple survey items and surveys that are fragmented over time and space. The model is very powerful, but also very complicated.[1] Although it

[1] Indeed Caughey and Warshaw (2015) note in a footnote that one of their models took "several weeks" of computing time to fit.

could be used to measure opinion across countries using complete national samples, much of the complexity comes from allowing the analyst to estimate subnational opinion from disaggregated and unrepresentative samples.

This paper instead focuses specifically on estimating country-year panels of opinion. The assumption is that nationally-aggregated survey marginals are drawn from representative samples – or have been weighted to approximate representativity. The challenge is then to accurately measure opinion despite gaps in time, space, and survey item. The models for doing so are developed and presented in the next section.

3. Modeling Cross-National, Time Series Latent Opinion

What do we require of a model of cross-national, time-series opinion? There are four guiding principles from existing research. First, we should treat opinion as an unobserved, latent trait, with observed survey responses being a function of these latent country-year traits. In effect we should set up a measurement model with latent estimates of country by time, as well as item-specific parameters to adjust the location and scale of the link between observed responses and aggregate opinion (Beck 1989; Caughey and Warshaw 2015; McGann 2014; Voeten and Brewer 2006).

Second, while classic measurement models – whether in the factor analytic or IRT traditions – can be thought of as estimating latent variables by smoothing over (for example) survey items, Beck (1989), Voeten and Brewer (2006), and Caughey and Warshaw (2015) extend these models by additionally smoothing over time. We will follow suit but incorporating a model of temporal dynamics.

Third, we should model the number of respondents – rather than the derived proportion or percentage – offering an affirmative (or dissenting) opinion. This implies a binomial model linking observed responses and an estimated probability parameter (Linzer 2013; Caughey and Warshaw 2015; McGann 2014). Such a model allows for sampling error to be neatly included. It also allows for responses proportions to be non-normally distributed.

Finally, since we are interested in extending models developed for single country time-series to cross national time series, we ought to adjust for heterogenous item functioning across countries, which is unfortunately quite prevalent in cross-national public opinion (Stegmueller 2011). We consider ways of accomplishing this below.

Following these principles, we develop six models of cross-national, time-series opinion (a summary of the six models is provided in Table 1). These models are tested using a real-world application: estimating support for democracy. Using both internal and external validation, we test the accuracy of the six sets of point estimates and variance estimates, and select a preferred model.

3.1. Distributions

The observed number of respondents, y_{ikt}, offering an affirmative opinion (e.g. in support of democracy) for each country, i, year, t, and survey item, k, is modeled as a binomial distributed count:

$$y_{ikt} \sim \text{Binomial}(s_{ikt},\ \pi_{ikt}). \tag{1}$$

We then face a choice, with two ways of proceeding. In the simpler binomial specification we then model the probability parameters, π_{ikt}, directly as a function of item and country-time effects (e.g. Caughey and Warshaw 2015; Linzer 2013). However we might also follow McGann (2014) in utilizing a beta prior on the probability parameter. This allows for some additional dispersion in the observed survey responses beyond that induced by sampling. Indeed, given the numerous sources of error likely to afflict public opinion survey data (e.g., methods of respondent selection, interviewing, questionnaire design, and translation, in addition to sampling), allowing for overdispersion in survey responses would appear to be prudent.

We thus use the simpler binomial specification for three of our six models, and the binomial with beta prior, or beta-binomial, for the other three. The beta-binomial specification then also includes the following step:

$$\pi_{ikt} \sim \text{Beta}(\alpha_{ikt},\ \beta_{ikt}) \tag{2}$$

The two shape parameters of the beta distribution can be reparameterized to an expectation parameter, η, and a dispersion parameter, ϕ:

$$\alpha_{ikt} = \phi \eta_{ikt} \tag{3}$$

$$\beta_{ikt} = \phi(1 - \eta_{ikt}) \tag{4}$$

3.2. Item and Country Parameters

In the case of the binomial specification, the probability parameters are modeled directly as a function of the latent country-year estimates and some item parameters; in the case of the beta-binomial, the beta expectation parameter receives this measurement model. We utilize three variations of this measurement model. The first simply includes country-year latent effects and item intercepts (we show the beta-binomial version here):

$$\eta_{ikt} = \text{logit}^{-1}(\lambda_k + \theta_{it}) \tag{5}$$

$$\lambda_k \sim \text{N}(\mu_\lambda,\, \sigma_\lambda^2) \tag{6}$$

The item intercepts, λ, adjust the location of the latent opinions for the idiosyncrasies of each survey item. They can thus be thought of as item bias effects. We model these intercepts hierarchically, with an expectation, μ_λ, and variance, σ_λ^2, estimated from the data. This hierarchical specification shrinks the item intercepts towards the mean to the extent that data are scarce, which guards against small within-item samples producing extreme estimates.

Survey items are, moreover, likely to have differing effects in different countries, a problem known as lack of equivalence (Stegmueller 2011). For example, one method of measuring support for democracy is to ask respondents for their opinions about having the army govern the country. Respondents in countries with a history of military rule are likely to respond quite differently than respondents in countries without such a history.

Fortunately, each item is asked multiple times in a given country. When replicates of items across units (respondents or countries) are available, analysts may also include parameters capturing item by unit bias (Skrondal, and Rabe-Hesketh 2004). Our second version of the measurement model thus includes a set of item by country effects, δ, to capture the heterogeneity in item bias across countries:

$$\eta_{ikt} = \text{logit}^{-1}(\lambda_k + \delta_{ik} + \theta_{it}) \tag{7}$$

$$\delta_{ik} \sim \text{N}(0,\ \sigma_\delta^2) \tag{8}$$

These item-country intercepts are also modeled hierarchically, which is particularly important as the observed data are especially likely to be sparse when divided by country as well as item. By treating both the item and item-country effects as varying intercepts, or random effects, we can interpret them as error terms (McGraw and Wong 1996). This lends an intuitive understanding to their role in the measurement equation: the λ effects can be seen as the item-level residuals, while the δ effects can be seen as the item-country level residuals, leaving θ as the item and item-country adjusted estimates of latent support for democracy.

Finally, measurement models often also include item slopes, known as factor loadings in the factor analysis framework and discrimination parameters within the IRT approach. Whatever the name, item slopes allow the strength of the relationship between observed responses and latent traits to vary across the items. Where an item shows a weaker relationship with the latent variable, it "loads" to a lesser extent than items with a stronger relationship. We extend the second model by incorporating such item slopes, γ:

$$\eta_{ikt} = \text{logit}^{-1}(\lambda_k + \delta_{ik} + \gamma_k\theta_{it}) \tag{9}$$

With both varying intercepts and varying slopes, this is a type of hierarchical (generalized) linear model, with observed responses nested within both items and countries (ignore time for the moment). As such, it is desirable to model the item-level equations jointly using a bivariate normal (Gelman and Hill 2007; Skrondal, and Rabe-Hesketh 2004). This allows item intercepts and slopes to be correlated, with the ρ parameter capturing the degree of covariation.

$$\begin{pmatrix} \lambda_k \\ \gamma_k \end{pmatrix} \sim \mathrm{N} \left[\begin{pmatrix} \mu_\lambda \\ \mu_\gamma \end{pmatrix}, \begin{pmatrix} \sigma_\lambda^2 & \rho\sigma_\lambda\sigma_\gamma \\ \rho\sigma_\lambda\sigma_\gamma & \sigma_\gamma^2 \end{pmatrix} \right] \tag{10}$$

With three versions of the measurement model coupled with the two response distributions (binomial and beta-binomial), there are six models in total. These are outlined in Table 1.

Table 1. Models

Model number	Response distribution	Item intercepts (λ)	Item-country intercepts (δ)	Item slopes (γ)
1	Binomial	✓		
2	Binomial	✓	✓	
3	Binomial	✓	✓	✓
4	Beta-binomial	✓		
5	Beta-binomial	✓	✓	
6	Beta-binomial	✓	✓	✓

3.3. Temporal Effects

Finally, for all six models, we allow our latent opinion estimates to evolve smoothly over time. Doing so fills in the gaps in each national time series. Following previous research on modeling dynamic latent traits (e.g. Caughey and Warshaw 2015; Jackman 2005), the temporal evolution of latent opinion is specified as a simple local-level dynamic linear model (Durbin and Koopman 2012), where the current level of latent opinion is a function of the previous year's level plus some random noise:

$$\theta_{it} \sim \mathrm{N}(\theta_{i,t-1},\ \sigma_\theta^2) \tag{11}$$

The variance of the noise term, σ_θ^2, is held constant across countries and estimated from the data.

3.4. Identification and Priors

To identify latent trait models, analysts must impose restrictions on the location, scale, and perhaps also the direction (or sign) of the parameters (Bafumi et al 2005). Models without item loadings are simpler in this respect as they only require location restrictions. These models were identified by fixing the first item intercept, λ_1, at a value of 1. Models with item loadings (models 3 and 6) additionally require that we identify the scale and direction of the parameters. To do so, we fix the expectation of the item intercepts, μ_λ, to 0.5, and the expectation of the item slopes, μ_γ, to 1.[2] The direction of the item slopes, γ_k, is then identified by constraining these to be positive.

The estimated variances are given weakly-informative half-Cauchy priors, e.g., $\sigma_\lambda \sim \mathrm{C}^+(0, 2)$ (and similarly for σ_δ, σ_γ, and σ_θ). For the models including item slopes as well as intercepts, the variance-covariance matrix of item intercepts and slopes is decomposed into the product of the variances for each vector of parameters and a 2×2 correlation matrix, with ρ being the estimated correlation (Stan Development Team 2017). This correlation matrix is given an LKJ prior (Lewandowski et al 2009).

The expectation of the item intercepts μ_λ (for models 1, 2, 4, and 5) is given a $\mathrm{N}(1,\ 2)$ prior while the dispersion parameter ϕ (for the beta-binomial models), is given a $\Gamma(4,\ 0.1)$ prior. Finally, for all models, the initial value of latent opinion for each country, θ_{i1}, receives a $\mathrm{N}(0,\ 1)$ prior.

4. Application: Support for Democracy

4.1. The Concept of Support for Democracy

Beginning with Aristotle, political theorists have long argued that the presence or absence of a democratic political system is somehow related to the attitudes and orientations of the citizenry. In modern times, interest in the cultural requisites of democracy led to the development of the concept of political culture, with democracy being legitimate, and stable, when it is "congruent" with the

[2] For these models λ_1 was not constrained.

political culture (Inglehart and Welzel 2005; Lipset 1959). Put another way, democracy requires a democratic political culture.

There are in fact two distinct conceptualizations of democratic political culture. According to the first, citizens provide explicit support for democracy when they prefer a democratic regime to some non-democratic alternative (e.g. Fuchs-Schündeln and Schündeln 2015; Mattes and Bratton 2007; Norris 1999; Rose, Mishler and Haerpfer 1998). Here, democracy is legitimate because it is believed to be preferable by the public. According to the second conceptualization, citizens provide implicit support when they subscribe to a broader set of values emphasizing trust, tolerance, and freedom (Inglehart 2003; Inglehart and Welzel 2005). Here, democracy is legitimate because it is consistent with citizen's deeper values and strivings. Although both kinds of democratic political culture have been advocated as providing support for democracy, or perhaps even spurring democratization, the focus of this paper is on the first kind, explicit support for democracy, often simply referred to as "support for democracy."[3]

4.2. Data on Support for Democracy

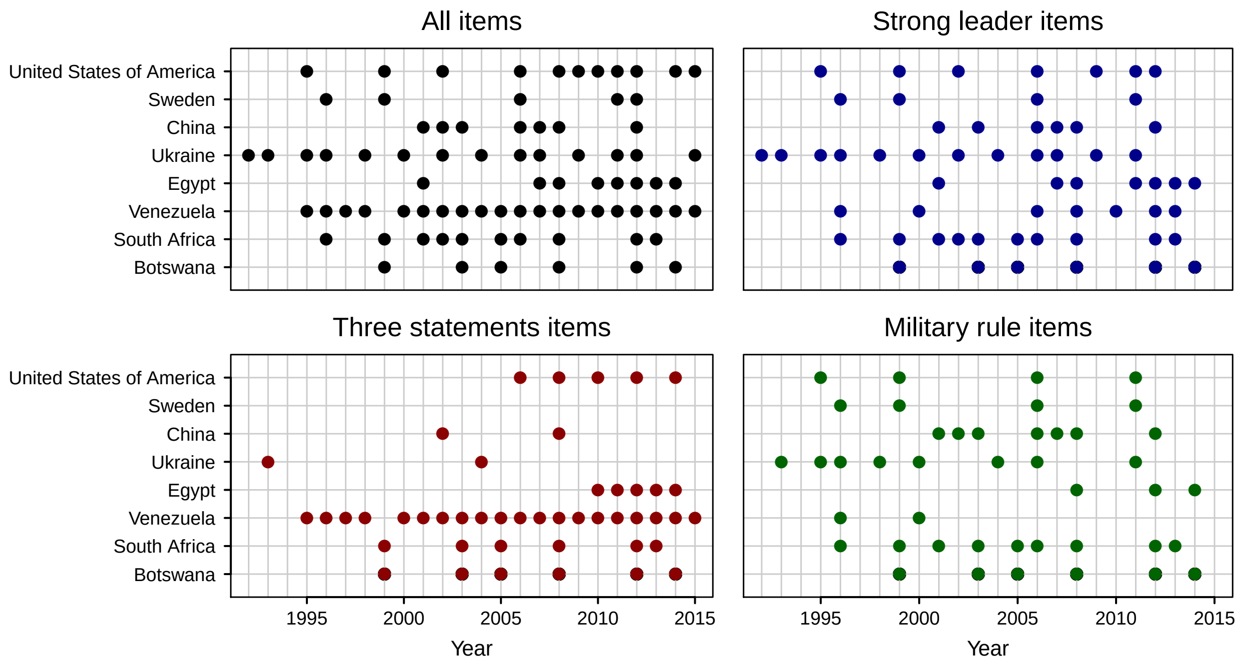
We collect all the available nationally-aggregated responses to questions on support for democracy that have been fielded in cross-national survey projects utilizing representative national samples of citizens. Data are gathered from ten survey projects including the World Values Survey and all the Global Barometer projects (see Appendix A). Surveys with relevant items were fielded in 145 countries over a 24 year period between 1992 and 2015. There are 2,978 nationally aggregated responses, obtained from 1,126 separate national survey samples.[4]

These data epitomize the challenges of measuring cross-national time-series opinion. First, they are sparse over time and space. If we restrict our focus to the 127 countries that were surveyed at least twice on explicit support over the years from 1992 to 2015, we have a potential dataset of

[3]Scholars such as Canache, Mondak, and Seligson (2001) and Norris (1999) demonstrate that support for democracy must be distinguished from the concept of "satisfaction with democracy," measures of which are also widespread in comparative survey projects. We follow suit.

[4]We excluded data from the World Values Survey for some countries and items due to known problems in some cases, and suspicious response levels in others. See Kurzman (2014) for further discussion and the appendix for details.

Figure 1. Sparseness of Aggregate Support for Democracy By Country, Year, and Survey Item



The first panel shows the availability of at least one survey item across a selection of eight countries and all 24 years. The other three panels indicate the availability of data for the three most common question themes.

3,048 country-years. However, surveys were conducted in just over a third of these country-years. We present a visualization of the sparseness of this data in Figure 1. The top panel shows the fragmented supply of support for democracy measures across time for eight countries, chosen to represent variance in regions and availability of data.

To take the example of South Africa, questions on explicit support for democracy were asked in 11 national surveys over the period in question: by the World Values Survey in 1996, 2001, 2006 and 2013, the AfroBarometer project in 1999, 2003, 2005, 2008 and 2012, and Pew Global Attitudes in 2002 and 2013. Data on democratic support are thus only available for ten out of the 24 years in South Africa – and this is a case that has above average survey coverage.

Second, to compound the problem, researchers have not settled on standard survey questions for measuring democratic political culture. Indeed, we find an extraordinary diversity of approaches to measuring support for democracy: as many as 36 different survey questions clustered within nine broad measurement approaches.[5] The lower panels of Figure 1 show the supply of support for democracy data within the three most prominent of of these measurement approaches. Once disaggregated in this way, the data are clearly even more sparse across the country-year matrix.

To continue the South African example, although 11 surveys fielded questions on support for democracy in this country, each used several different items, resulting in 41 data points in total. These 41 data points are however fragmented across seven different questions. If we were to focus on a single survey question to obtain a meaningful time-series, we would discard most of the data. Even the most popular survey item, the question asking respondents the extent to which they support or oppose having a strong but undemocratic leader, is asked only in ten out of the 24 years.

As such, once we separate available support for democracy data by survey item as well as country and year, the data begin to look very sparse indeed. If survey questions from every one of the main nine themes were asked annually in each country, we would have 27,432 country-level data points. Yet, only around ten percent of these potential item-country-year cells actually feature data. Classical methods of measuring public opinion using multiple items, such as factor analysis of the complete cases, are not an option here. The data are simply too sparse across survey item, country, and year. Indeed, there are no complete cases unless one ignores the temporal dimension.

Given these limitations, analysts have tended to abandon any temporal variation in support for democracy and focus only on cross-sectional variation. However, given the presence of numerous cross-national confounding factors, such research designs do not permit causal conclusions to be drawn with any confidence. we aim to overcome these limitations by estimating support for democracy over 127 countries and over 24 years, creating a time-series cross-sectional dataset with 3,048 observations.

[5] We took a conservative route in categorizing survey items by always classing two items as distinct when they were fielded by different project even if their wording appeared to be identical. Doing so allows our item effect parameters to capture variation induced both by question wording and by idiosyncrasies in the methodology of the various survey projects.

4.3. Estimation

The six models are estimated using Bayesian Markov-Chain Monte Carlo (MCMC) methods via Stan software, which implements Hamiltonian Monte Carlo sampling (Carpenter et al 2017; Stan Development Team 2017). Four parallel chains were run for 1,000 samples each, with the first 500 samples in each chain used for warm up, and discarded, and the remaining 2,000 samples of the posterior density saved and analyzed further. This number of iterations proved to be more than sufficient for convergence, with the Gelman-Rubin diagnostic reaching a value of between 0.95 and 1.05 for all parameters.

4.4. Empirical strategy

To select a best model out of the six, we compare their accuracy and efficiency using several metrics. First, we test the accuracy of the models using the same data used to fit the model, known as internal validation (Hastie, Tibshirani, and Friedman 2009). In particular, we use the mean absolute error (MAE) to measure the average discrepancy between the observed proportions of each sample offering a pro-democratic attitude, and the proportions predicted by each model given the estimated values of latent opinion for each country and year:

$$\mathrm{MAE} = \frac{1}{J} \sum_{j \in ikt}^{J} \left| \frac{s_{ikt}}{n_{ikt}} - \hat{\pi}_{ikt} \right| \tag{12}$$

Although internal validation is fairly simple to conduct, it favors more complex models, and thus could lead to the selection of a model that overfits the dataset at hand. Instead, analysts tend to rely on information criteria, which attempt to estimate out-of-sample predictive error by penalizing models as their parameters increase in number. An good choice of information criterion in for Bayesian MCMC methods is the “Leave-One-Out” (LOO-IC) information criterion, which Vehtari, Gelman, and Gabry (2017) argue to be superior to alternatives such as the Deviance and Watanabe-Akaike information criteria.

Better still, however, is to select models using a source of data that was not previously used to fit the model – also known as external validation (Hastie, Tibshirani, and Friedman 2009). To do so, we randomly split our full dataset into an 75% training set and a 25% test set. We refit our six models using the training set, and use the resulting parameter estimates to predict the national proportions offering a supportive (i.e. pro-democratic) response for each of the 733 survey items comprising the test dataset. We again calculate the mean absolute error, but now in predicting the error in the test data. We also examine the efficiency of our models by calculating their credible interval coverage (CIC). To do so, we count the percentage of times each of the $J = 733$ observed survey proportions is included in the 80% credible interval of the corresponding estimate:

$$\text{CIC} = \frac{100}{J} \sum_{j=1}^{J} \left(\frac{s_j}{n_j} \in \text{CI}_{\alpha}(\hat{\pi}_j) \right) \tag{13}$$

We also include two naïve methods of estimating the out-of sample proportions: firstly using only the item means in the training dataset to predict the proportions in the test set, and secondly, using only the country means. These will serve as baseline measures to guage the relative accuracy of all six of our models.

5. Results

5.1. Internal and External Validation

The results of our internal and external validation tests are displayed in Table 2. Beginning with the tests of internal validation, three findings are apparent. First, all six models offer a substantially better fit to the observed proportions supporting democracy than the baseline item or country means. This is hardly surprising given the difficulty in estimating a particular aggregate response to a support for democracy survey item knowing nothing other than either the survey item or the country. Yet it does serve as a first check that are models are adding value.

Table 2. Internal and External Validation Tests

	Internal validation		External validation	
Model	MAE	LOO-IC	MAE	CIC
1	.050	143491	.080	16.6
2	.022	59816	.065	38.6
3	.022	58118	.071	43.8
4	.062	35951	.072	36.3
5	.032	34353	.060	60.8
6	.031	34329	.063	60.4
Country means	.106		.113	
Item means	.096		.098	

Internal validation includes tests of model fit using the same data for model fitting and validation. External validation creates two separate datasets: models are fit to training set and validated using test / hold-out set. MAE: mean absolute error; LOO-IC: leave-one-out information criterion; CIC: 80% credible interval coverage.

Perhaps more interesting is the second finding, which is that models 1 and 4 – the models including item intercepts, but not item slopes or item-country intercepts – offer substantially worse fit than the the other, more complex models. The error rate is roughly halved when adding item-country intercepts (which are incorporated in models 2, 3, 5, and 6). This result is confirmed by the LOO-IC measures. Although the LOO-IC penalizes the log-likelihood for the number of estimated parameters, models 2 and 3 offer a better fit than model 1, as do models 5 and 6 when compared with model 4.[6] There is however little to distinguish between the models with item slopes (3 and 6) and the models without (2 and 5).

Finally, the internal validation mean absolute errors results suggest that the simpler binomial models (1–3) fit slightly better than the beta-binomial models (4–6).

We now turn to the external validation tests, which offer a better means of gauging model fit because the models are estimated and tested using different datasets. Looking at the external validation MAE results (Figure 2), we see that the six models continue to offer improvements

[6]The use of different distributions means that one cannot compare the LOO-IC across the binomial and beta-binomial specifications.

in accuracy when compared with our baseline, naïve estimates. The difference, however, has diminished. Part of the gap between our models and the baseline estimates were thus due to our models overfitting the data.

However, all six models offer gains in accuracy over naïve methods. Whereas the latter leads to an average error of 10-11 percentage points, the former result in errors of 6-8 percentage points. In addition, the external validation test confirms that the simpler item-intercept only models are the worst fitting, with the additional complexity added by including item-country intercepts leading to a reduction in error. Finally, adding item slopes or factor loadings does not improve predictive accuracy at all.

A second finding from the tests of external validation – and one that stands in contrast to the internal validation tests – is that the beta-binomial (models 4–6) specifications are slightly more accurate than the corresponding binomial models (1–3). The additional dispersion added by the beta-binomial enhances the accuracy of the estimates, perhaps by capturing some of the non-sampling error endemic in public opinion data.

Finally, the tests of credible interval coverage. This metric measures the accuracy of the estimates of uncertainty produced by each model. A model with accurate uncertainty estimates should have similar empirical and nominal levels of credible interval coverage. Since we use 80% credible intervals, we expect 80% empirical coverage. Coverage substantially below this nominal level shows inaccurate and overly precise standard errors; coverage substantially above this level indicates inaccurate and inefficient uncertainty estimates.[7]

The uncertainty estimates generated by the six models are all overly precise. None of the rates of empirical credible interval coverage come appreciably close to the nominal level of 80%. There are of course, many sources of error in cross-national public opinion data, and we have explicitly modeled only a few. Moreover, some sources of error – such as the translation problems in the World Values Survey identified by Kurzman (2014) – are impossible to model.

[7]For example, an estimated interval between negative and positive infinity would have 100% coverage but it would tell us nothing of consequence.

However, the beta-binomial specification, which includes an overdispersion parameter, ϕ, proves to have substantially better uncertainty estimates than the simpler binomial specification. While the binomial CICs are very poor, ranging from 17 to 44%, the beta-binomial CICs are much closer to nominal 80% level, ranging from 36 to 61%. Including item-country random effects also produces substantially better credible interval coverage, in both the binomial and beta-binomial specifications. Models 5 and 6, with both item-country intercepts and beta-binomial distributions, have the most accurate estimates of uncertainty.

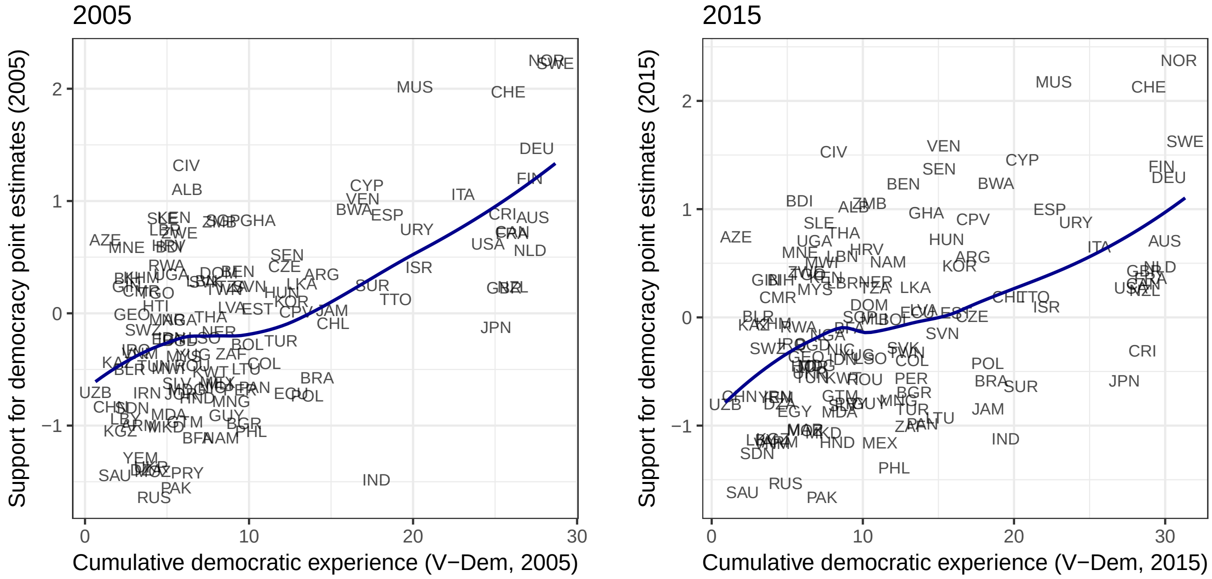
Weighing up all the evidence from our tests of external validation, it would appear that model 5, with a beta-binomial specification and random item and item-country effects, is the best overall. This model shows the lowest error in predicting survey responses in the test dataset and has the best credible interval coverage. It is also considerably simpler to code and fit than the runner-up, model 6, which requires the covariance of the item effects to be estimated. We conclude that model 5 is a good choice for estimating fragmented cross-national public opinion.

5.2. Construct and Convergent Validation

Our external validation tests demonstrate that our modeling framework – and in particular, the beta-binomial model with item-country intercepts – is reasonably accurate in predicting observed survey responses in a hold-out sample. To further build confidence in our approach we examine in this subsection whether the estimated latent opinion corresponds to the theoretical construct of support for democracy. We examine, in other words, whether our country-year measures of support for democracy behave as previous scholars have suggested this variable *should* behave.

In particular, we consider the cross-national distribution of estimated support for democracy at certain points in time and the across-time evolution of support for democracy for certain countries. This analysis will allow us to say something about the construct validity of our estimates. We also examine the correlation, at certain points in time, between cross-national point estimates of support for democracy and a measure of a country's cumulative democratic experience. This will allow us to test the convergent validity of our measures.

Figure 2. The Relationship Between Support for Democracy and Democratic Experience



Estimates of support for democracy, from Model 5, for all countries, plotted against cumulative democratic experience, in 2005 (top) and 2015 (bottom). Cumulative democratic experience is the sum of annual democracy scores discounted by 2% a year. A LOWESS line is added to each plot. The Pearson's correlations are 0.55 in 2005 and 0.46 in 2015.

We begin with the latter, convergent validity. To assess our model according to this standard, we examine the covariation between support for democracy and cumulative experience with democracy at two points in time, 2015 and 2005. Cumulative experience with democracy is the sum of the democracy scores for a given country between the year in question and 1950, with each year's score discounted by two percent.[8] Scholars have previously demonstrated that supportive attitudes toward democracy are linked with the length of time a country has been democratic (Fuchs-Schündeln and Schündeln 2015; Mattes and Bratton 2007). Scholars have also argued – although perhaps not yet empirically demonstrated – that support for democracy helps democratic institutions to survive (Lipset 1959; Norris 1999). Since either of these processes would lead to a correlation between support for democracy and democratic experience, we regard this as a test of the convergent validity of our measures. To carry out this test, we plot point estimates of support for democracy against democratic experience in both 2015 and 2005 (Figure 2).

[8] We use the "Liberal democracy index" from the Varieties of Democracy project (Lindberg et al 2014)

In both 2005 and 2015, a robust and positive relationship is evident between the two variables (in 2005 the Pearson's correlation is 0.55; in 2015, 0.46). The greater a country's experience with democracy, the higher the support that its citizens express for a democratic versus an autocratic system. These correlations show that our estimates of opinion do in fact behave as theories of democratic political culture have suggested.

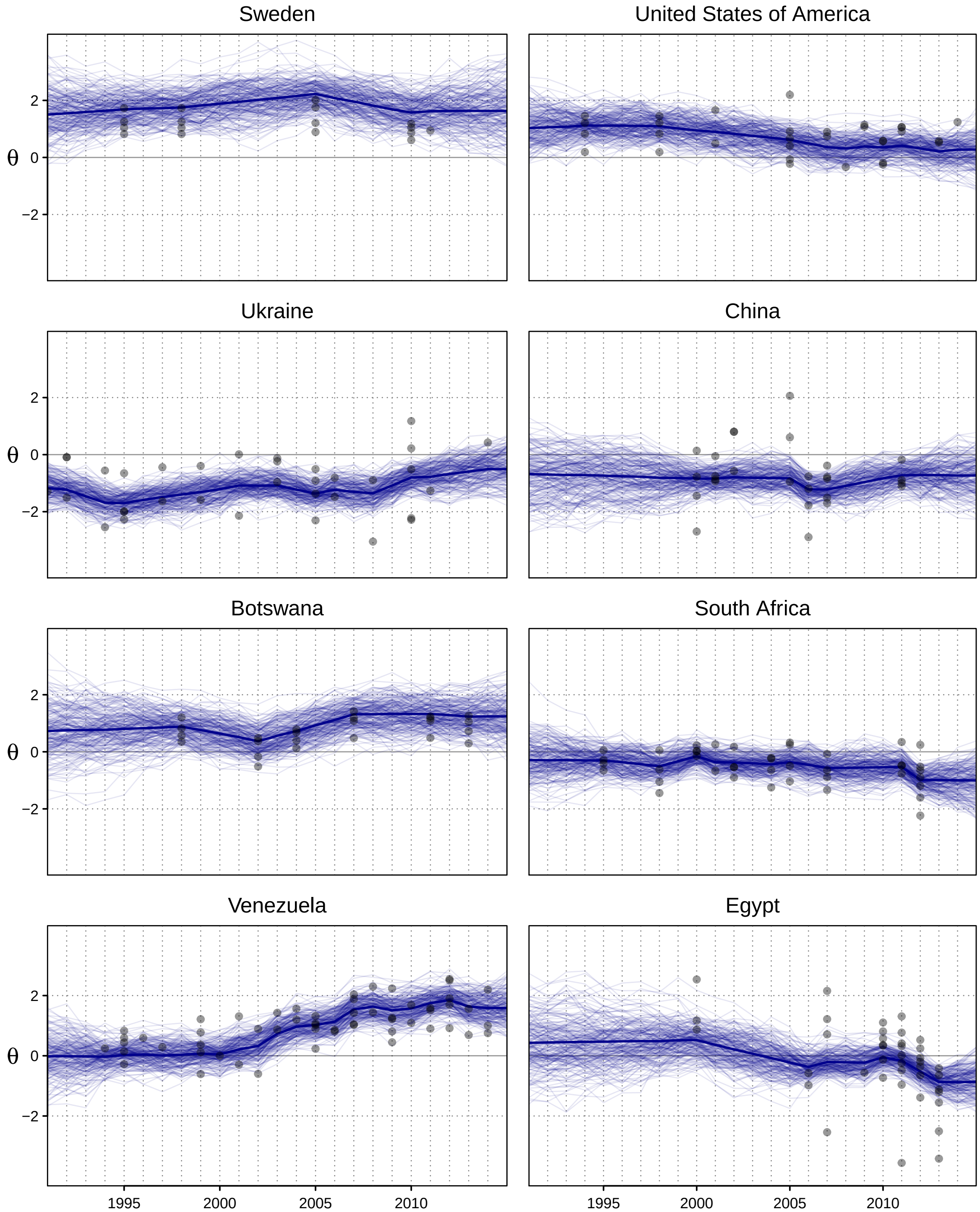
Moving on to construct validity, we examine the estimated levels of support for democracy for a selection of eight countries. These estimates are displayed in Figure 3. Each plot shows the latent support for democracy estimates for a particular country over all 24 years. The dark bold line shows the mean value of θ in each year; the light lines indicate 200 random draws from the posterior density of θ, which shows uncertainty in the estimates. we have also included the observed data on these plots, indicated using points.[9]

When data are abundant in a particular country (e.g. Venezuela), the estimates are fairly precise (the y-axis is calibrated on the z-score scale). When data are scarce (e.g. Egypt and China before 2000), the estimates are noisier, and become increasingly so the larger the gap there is in the data. The beta-binomial model is fairly aggressive in smoothing across time, especially compared with the binomial specification (not shown), which produces a more jagged, rapidly changing pattern of opinion. Such a pattern is not particularly plausible for a slow-moving orientation such as democratic political culture (Inglehart and Welzel 2005), which suggests again that the beta-binomial specification should be favored.

Figure 3 focuses on eight time series that were selected as representing a range of levels of support as well as some interesting dynamics. First, we see that established democracies, such as Sweden, show high levels of support for democracy. This is consistent with previous research focusing on particular subsets of the available survey data (e.g. Klingemann 1999). An important exception to this rule is the United States (and, indeed, other Anglophone democracies), where we find declining support for democracy. This is in fact also consistent with new research by Foa and

[9] The observed data (which are national proportions offering support for democracy) are measured on a different scale to the latent estimates (which are unit-normal standardized). We thus standardized the observed responses by centering by survey item and dividing by the standard deviation of all responses. This places the observed data on approximately the same scale as the latent estimates.

Figure 3. Estimated Support for Democracy for 8 Countries over 24 Years



Estimates of support for democracy, from Model 5, for eight selected countries over 24 years. Each plot shows 200 random draws from the posterior distribution of θ for a particular country. The posterior means are indicated using bold lines. Observed survey responses for each country are plotted using points; these are unit-normal standardized within survey item so as to display them on roughly the same scale as the latent estimates.

Mounk (2016).

Second, newer democracies show divergent trends. We examine a pair of cases from Southern Africa: Botswana has high and increasing support for democracy, which echoes previous case study research (Hjort 2009). In contrast, Botswana's neighbor, South Africa, shows fairly low (and declining) support, which resonates with earlier research that analyzes different data (Gibson 2003). Third, we see that countries with a long history of autocratic rule, such as China and the Ukraine, have low levels of support, as existing research on authoritarian legacies would lead us to expect (Fuchs-Schündeln and Schündeln 2015; Rose, Mishler and Haerpfer 1998). Finally, Egypt and Venezuela are two countries where one can see public support reacting to political events, albeit in divergent ways. Venezuelan support for democracy steadily increases after Chavez began dismantling democratic checks and balances in 2000, while Egyptians reacted to the tumult of the Arab Spring by turning away from democracy.

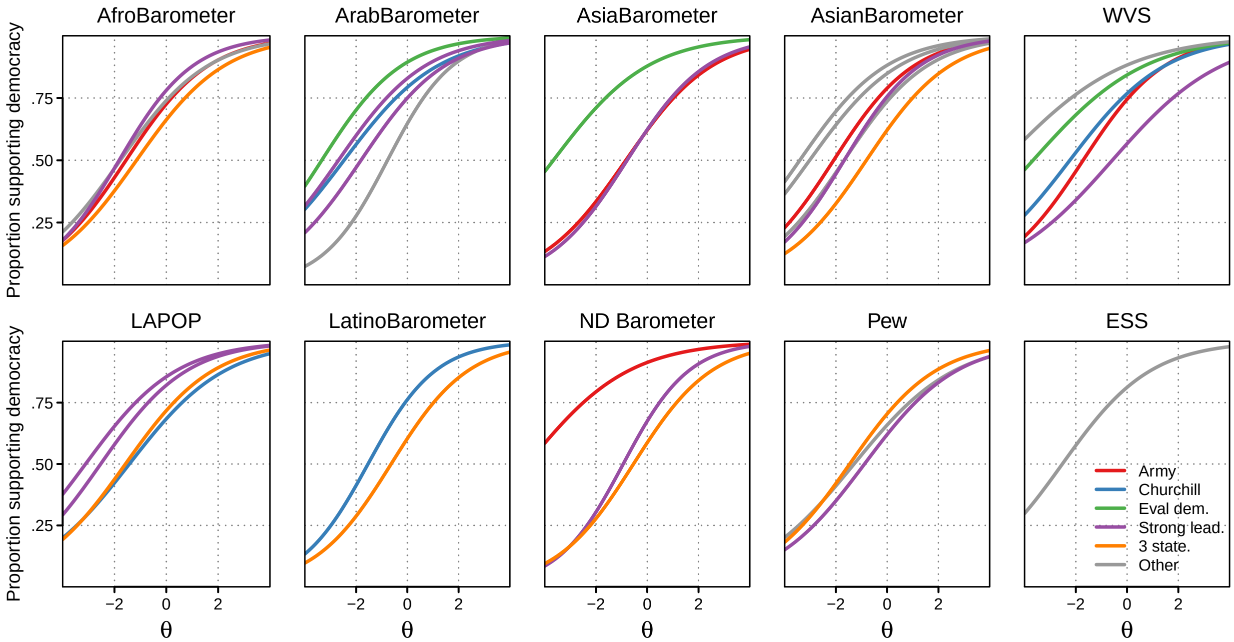
5.3. Item Analysis

Our final analysis is to examine the item parameters in more detail. Doing so will enable us further comment on the validity of our estimates. Moreover, the fact that we are able to analyze item functioning also illustrates an advantage of our modeling approach.

Our discussion will focus on the item characteristic curves (ICCs), which we plot in Figure 4. ICCs display the the relationship between the proportion of a national sample responding supportively toward democratic institutions (y-axis) and the latent estimates of support for democracy (x-axis). The vertical alignment of the curves is governed by λ, the item intercepts, while the steepness of the curves is governed by γ, the item slopes. To aid in interpretation, we group the items by their survey project, and color-code the curves by the question theme that was used. Finally, we use the parameter estimates from Model 6. Although not our selected model, it included both item intercepts and slopes; the latter, in particular, are a useful diagnostic tool.

Turning to Figure 4, one can see that all 36 items display a positive relationship between the latent quantity and the observed responses. All items, in other words, have positive slopes,

Figure 4. Item Characteristic Curves for All Items, Grouped By Project



Lines show the item characteristic curves (ICCs) for each item, with parameter estimates drawn from Model 6. These depict the relationship between the proportion of a national sample responding supportively toward democratic institutions (y-axis) and the latent estimates of support for democracy (x-axis). The vertical alignment of the curves is governed by λ, the item intercepts; the steepness of the curves is governed by γ, the item slopes. The ICCs are grouped by survey project and are color-coded to show question themes.

and indeed, most the items show similar slopes. This is a welcome finding, as it indicates that all items do indeed measure the latent construct. It is not a particularly surprising finding, however, as the items were pre-selected on the basis of previous analyses of microlevel survey data (e.g. Klingemann 1999; Mattes and Bratton 2007). Items that bore some superficial resemblance to support for democracy, but which did not display a deeper empirical relationship with this latent variable, were not included in the dataset in the first place.[10]

There are nonetheless a few items with weaker slopes, and thus, more tenuous relationships with latent support for democracy. First is the "army rule" item from the New Democracies Barometer, and second is the "evaluate democratic political system" item from the AsiaBarometer. Another three come from the World Values Surveys: the importance of living in a democracy

[10]The similar item slopes also suggests why the inclusion of these parameters in Models 3 and 6 did not improve their fit.

item, and the evaluations of a "strong leader" and "democratic political system" items. This latter item, in particular, has previously been criticized as offering only "lip service", rather than actual support, for democracy (Inglehart 2003). However, when we consider all the items from the various survey projects that use similar wording, we do not find them to be generally weak, or "lip service." Rather, we notice that such items tend to have larger intercepts – more respondents tend to agree with these items overall. This can actually be a useful feature as it allows such items to discriminate among societies that are generally opposed to democracy (and thus situated on the left hand side of the latent space in Figure 4.

Our main finding from the item analysis is that all our items show a marked, positive relationship between the latent variable and the observed responses. Indeed, most of our items show a strong relationship, indicating that they are sound measures of the latent construct of support for democracy.

6. Conclusion

The availability of smooth panels of cross-national public opinion would be a great asset to scholars of comparative political behavior and comparative political economy. Yet assembling and estimating such panels is far from straightforward. Public opinion data are fragmented over space and time. They are also fractured across the numerous survey items that are used to measure any given opinion construct. To make matters worse, cross-national opinion data are gathered by a variety of survey projects using a variety of methodologies and in dozens of languages and countries, threatening their equivalence across countries.

Despite these challenges, our models perform reasonably well in tests of external validation. Our best model predicts aggregate survey responses that, on average, deviate by six percentage points from the observed percentages. This same model provides modestly accurate estimates of measurement uncertainty, with empirical credible interval coverage falling 20 percentage points short of the nominal level. We furthermore find that our estimated panel of opinions on support for democracy displays spatiotemporal patterns and associations with other variables that are consistent with previous research, suggesting both construct and convergent validity. Given the problems endemic in such data, we think that our results warrant optimism.

They also warrant further application, refinement, and testing. For scholars interested in applying these models to other contexts and to other opinions, we find, firstly, that a beta-binomial specification should be selected rather than the simpler binomial. Both are preferable, however, to modeling the proportions directly using a linear model. Secondly, we find that item-slopes or factor loadings do not increase the accuracy of our models but do increase its computational complexity. However, they have a diagnostic utility, as we demonstrate in our item analysis, so analysts might consider running models both with and without these item slopes. Finally, when item-country replicates are available due to a particular item being fielded more than once in a particular country, we recommend including item-country effects, because these increase accuracy dramatically.

References

Almond, Gabriel A., and Sidney Verba. 1963. *The Civic Culture: Political Attitudes and Democracy in Five Nations*. Boston: Little & Brown.

Bafumi, Joseph, Andrew Gelman, David K. Park, and Noah Kaplan. 2005. "Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation." *Political Analysis* 13(2): 171–87.

Baumgartner, Frank R., Suzanna L. De Boef, and Amber E. Boydstun. 2008. *The Decline of the Death Penalty and the Discovery of Innocence.* New York: Cambridge University Press.

Beck, Nathaniel. 1989. "Estimating Dynamic Models Using Kalman Filtering." *Political Analysis* 1: 121–156

Canache, Damarys, Jeffery J. Mondak, and Mitchell A. Seligson. 2001. "Meaning and Measurement in Cross-National Research on Satisfaction with Democracy." *Public Opinion Quarterly* 65(4): 506–528.

Caughey, Devin and Christopher Warshaw. 2015. "Dynamic Estimation of Latent Opinion Using a Hierarchical Group-Level IRT Model." *Political Analysis* 23(2): 197–211.

Carpenter, Bob, Andrew Gelman, Matthew Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76(1): 1–32

Durbin, James and Siem Jan Koopman. 2012. *Time Series Analysis by State Space Methods*, Second Edition. Oxford: Oxford University Press.

Erikson, Robert S., Michael MacKuen, and James A. Stimson. 2002. *The Macro Polity.* New York: Cambridge University Press.

Foa, Roberto Stefan and Yascha Mounk. 2016. "The Democratic Disconnect." *The Journal of Democracy* 27(3): 5–17.

Fuchs-Schündeln, Nicola and Matthias Schündeln. 2015. "On the Endogeneity of Political Preferences: Evidence from Individual Experience with Democracy." *Science* 347(6226): 1145–1148.

Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel / Hierarchical Models.* New York: Cambridge University Press.

Gibson, James L. 2003. "The Legacy of Apartheid: Racial Differences in the Legitimacy of Democratic Institutions and Processes in the New South Africa." *Comparative Political Studies* 36(7): 772–800.

Green, Donald P., Alan Gerber, and Suzanna L. De Boef. 1999. "Tracking Opinion Over Time: A Method for Reducing Sampling Error." *Public Opinion Quarterly* 63: 178–92.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd Edition).* New York: Springer.

Hjort, Jonas. 2009. "Pre-Colonial Culture, Post-Colonial Economic Success? The Tswana and the African Economic Miracle." *The Economic History Review* 63(3): 688–709.

Inglehart, Ronald. 2003. "How Solid is Mass Support for Democracy – and How Can We Measure It?" *PS, Political Science and Politics* 36(1): 51–57

Inglehart, Ronald and Christian Welzel. 2005. *Modernization, Cultural Change, and Democracy: The Human Development Sequence*. Cambridge University Press.

Jackman, Simon. 2005. “Pooling the Polls over an Election Campaign.” *Australian Journal of Political Science* 40(4): 499–517.

Klingemann, Hans-Dieter. 1999. “Mapping Political Support in the 1990s: A Global Analysis.” In *Critical Citizens: Global Support for Democratic Governance*, ed. Pippa Norris. Oxford: Oxford University Press.

Kurzman, Charles. 2014. “World Values Lost in Translation.” *Washington Post*. September 2, online only. https://www.washingtonpost.com/news/monkey-cage/wp/2014/09/02/world-values-lost-in-translation/

Lewandowski Daniel, Dorota Kurowicka, and Harry Joe. 2009. “Generating Random Correlation Matrices Based on Vines and Extended Onion Method.” *Journal of Multivariate Analysis* 100(9): 1989–2001.

Lindberg, Staffan I., Michael Coppedge, John Gerring, Jan Teorell, Daniel Pemstein, Eitan Tzelgov, Yi-ting Wang, Adam Glynn, David Altman, Michael Bernhard, Steven Fish, Alan Hicken, Matthew Kroenig, Kelly McMann, Pamela Paxton, Megan Reif, Svend-Erik Skaaning, and Jeffrey Staton. 2014. “V-Dem: A New Way to Measure Democracy.” *Journal of Democracy* 25(3): 159–169.

Linzer, Drew. 2013. “Dynamic Bayesian Forecasting of Presidential Elections in the States.” *Journal of the American Statistical Association* 108(501): 124–134.

Lipset, Seymour Martin. 1959. “Some Social Requisites of Democracy: Economic Development and Political Legitimacy.” *American Political Science Review* 53(1): 69–105.

Mattes, Robert and Michael Bratton. 2007. “Learning about Democracy in Africa: Awareness, Performance, and Experience.” *American Journal of Political Science* 51(1): 192–217.

McGann, Anthony J. 2014. “Estimating the Political Center from Aggregate Data: An Item Response Theory Alternative to the Stimson Dyad Ratios Algorithm.” *Political Analysis* 22(1): 115–129.

McGraw, Kenneth O. and S. P. Wong. 1996. “Forming Inferences About Some Intraclass Correlation Coefficients.” *Psychological Methods* 1(1): 30–46.

Norris, Pippa. 1999. “Institutional Explanations for Political Support.” In *Critical Citizens: Global Support for Democratic Governance*, ed. Pippa Norris. Oxford: Oxford University Press.

Park, David K., Andrew Gelman and Joseph Bafumi. 2004. “Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls.” *Political Analysis* 12(4): 375–385.

Putnam, Robert D. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton: Princeton University Press.

Rose, Richard, William Mishler, and Christian Haerpfer. 1998. *Democracy and Its Alternatives: Understanding Post-Communist Societies*. Baltimore: Johns Hopkins University Press.

Skrondal, Anders, and Sophia Rabe-Hesketh. *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: CRC Press.

Stan Development Team. 2017. *Stan Modeling Language: User's Guide and Reference Manual: Stan 2.15.0.* Stan Development Team

Stegmueller, Daniel. 2011. "Apples and Oranges? The Problem of Equivalence in Comparative Research." *Political Analysis* 19(4): 471–487.

Stimson, James A. 1991. *Public Opinion in America: Moods, Cycles, and Swings, Transforming American Politics.* Boulder, CO: Westview Press.

Stimson, James A., Michael B. Mackuen, and Robert S. Erikson. 1995. "Dynamic Representation." *American Political Science Review* 89(3):543–65.

Sullivan, John L., James E. Piereson, and George E. Marcus. 1982. *Political Tolerance and American Democracy*. Chicago: University of Chicago Press.

Vehtari, Aki, Andrew Gelman, and Johan Gabry. 2017. "Practical Bayesian Model Evaluation Using Leave-One-Out Cross-Validation and WAIC." *Statistics and Computing* 27(5): 1413–1432

Voeten, Erik, and Paul R. Brewer. 2006. "Public Opinion, the War in Iraq, and Presidential Accountability." *Journal of Conflict Resolution* 50(6): 809–830.

Estimating Smooth Country-Year Panels of Public Opinion: Online Supplementary Materials

Survey Items

Three statements items

1. Which of these three statements is closest to your own opinion?: democracy is preferable to any other kind of government, under some circumstances, an authoritarian government can be preferable to a democratic one, for someone like me, it does not matter what kind of government we have (Pew Global Attitudes).

2. Which of these three statements is closest to your own opinion?: democracy is preferable to any other kind of government, under some circumstances, an authoritarian government can be preferable to a democratic one, for someone like me, it does not matter what kind of government we have (AfroBarometer).

3. Which of the following statements do you agree with most? Democracy is preferable to any other kind of government. In certain situations, an authoritarian government can be preferable to a democratic one. To people like me it doesn't matter whether we have a democratic government or a non-democratic government (LatinoBarometer).

4. Which of the following statements do you agree with most? Democracy is preferable to any other kind of government. In certain situations, an authoritarian government can be preferable to a democratic one. To people like me it doesn't matter whether we have a democratic government or a non-democratic government (European Social Survey).

5. With which of the following phrases are you in most agreement: for people like me, it doesnt matter whether a regime is democratic or non-democratic, democracy is preferable to any other type of government, under some circumstances an authoritarian government can be preferable to a democratic one (Latin American Public Opinion Project).

6. Which of the following statements comes closest to your own opinion? For people like me, it does not matter whether we have a democracy, under some circumstances, an authoritarian government can be preferable, democracy is always preferable to any other kind of government (AsianBarometer)

Churchill items

7. Democracy may have its problems, but it is better than any other form of government. To what extent do you agree or disagree? (World Values Survey)

8. Democracy may have its problems, but it is better than any other form of government. To what extent do you agree or disagree? (ArabBarometer)

9. Do you strongly agree, agree, disagree or strongly disagree with the following statements: Democracy may have problems but it is the best system of government (LatinoBarometer)

10. With which of the following phrases do you most agree: in general, despite its problems, democracy is the best form of government, there are other forms of government that can be just as good or even better than democracy, dont know. (Latin American Public Opinion Project)

Strong leader items

11. Best to get rid of Parliament and elections and have a strong leader who can quickly decide everything. What do you think? (New Democracies Barometer)

12. Best to get rid of Parliament and elections and have a strong leader who can quickly decide everything. What do you think? (AsianBarometer)

13. On some occasions, democracy doesn't work. When that happens there are people that say we need a strong leader who doesn't have to be elected through voting. Others say that even if things don't function, democracy is always the best. What do you think? (Latin American Public Opinion Project)

14. There are people who say that we need a strong leader that does not have to be elected. Others say that although things may not work, electoral democracy, or the popular vote, is always best. What do you think? (Latin American Public Opinion Project)

15. There are many ways to govern a country. Would you disapprove or approve of the following alternatives? Elections and Parliament are abolished so that the president can decide everything. (AfroBarometer)

16. I'm going to describe various types of political systems and ask what you think about each as a way of governing this country. For each one, would you say it is a very good, fairly good, fairly bad or very bad way of governing this country? Having a strong leader who does not have to bother with parliament and elections (World Values Survey)

17. Some feel that we should rely on a democratic form of government to solve our country's problems. Others feel that we should rely on a leader with a strong hand to solve our country's problems. Which comes closer to your opinion? (Pew Global Attitudes)

18. I will describe different political systems to you, and I want to ask you about your opinion of each one of them with regard to the countrys governance for each one would you say it is very good, good, bad, or very bad? (ArabBarometer)

19. I will describe different political systems to you, and I want to ask you about your opinion of each one of them with regard to the countrys governance for each one would you say it is very good, good, bad, or very bad? (ArabBarometer)

20. I'm going to describe various types of political systems. Please indicate for each system whether you think it would be very good, fairly good or bad for this country. Governance by a powerful leader without the restriction of parliament or elections (AsiaBarometer)

Military rule items

21. The army should govern the country. What do you think? (New Democracies Barometer)

22. The army should govern the country. What do you think? (AsianBarometer)

23. There are many ways to govern a country. Would you disapprove or approve of the following alternatives? The army comes in to govern the country (AfroBarometer)

24. I'm going to describe various types of political systems and ask what you think about each as a way of governing this country. For each one, would you say it is a very good, fairly good, fairly bad or very bad way of governing this country? Having the army rule (World Values Survey)
25. I'm going to describe various types of political systems. Please indicate for each system whether you think it would be very good, fairly good or bad for this country. Military government (AsiaBarometer)

One party rule items

26. There are many ways to govern a country. Would you disapprove or approve of the following alternatives? Only one political party is allowed to stand for election and hold office (AsianBarometer)
27. There are many ways to govern a country. Would you disapprove or approve of the following alternatives? Only one political party is allowed to stand for election and hold office (AfroBarometer)

Evaluate democracy items

28. I will describe different political systems to you, and I want to ask you about your opinion of each one of them with regard to the countrys governance for each one would you say it is very good, good, bad, or very bad? A democratic political system (public freedoms, guarantees equality in political and civil rights, alternation of power, and accountability and transparency of the executive authority). (ArabBarometer)
29. I will describe different political systems to you, and I want to ask you about your opinion of each one of them with regard to the countrys governance for each one would you say it is very good, good, bad, or very bad? A democratic political system (public freedoms, guarantees equality in political and civil rights, alternation of power, and accountability and transparency of the executive authority). (World Values Survey)
30. I will describe different political systems to you, and I want to ask you about your opinion of each one of them with regard to the countrys governance for each one would you say it is very good, good, bad, or very bad? A democratic political system (public freedoms, guarantees equality in political and civil rights, alternation of power, and accountability and transparency of the executive authority). (AsiaBarometer)

Democracy suitable items

31. Here is a similar scale of 1 to 10 measuring the extent to which people think democracy is suitable for our country. If 1 means that democracy is completely unsuitable for [name of country] today and 10 means that it is completely suitable, where would you place our country today? (AsianBarometer)
32. Suppose there was a scale from 0-10 measuring the extent to which democracy is suitable for your country, with 0 meaning that democracy is absolutely inappropriate for your country and 10 meaning that democracy is completely appropriate for your country. To what extent do you think democracy is appropriate for your country? (ArabBarometer)

Importance to you items

33. How important is it to you to live in a country where honest elections are held regularly with a choice of at least two political parties? Is it very important, somewhat important, not too important or not important at all? (Pew Global Attitudes)

34. How important is it for you to live in a country that is governed democratically? On this scale where 1 means it is “not at all important” and 10 means “absolutely important” what position would you choose? (World Values Survey)

35. How important for you to live in democratically governed country? (European Social Survey)

Desire for democracy items

36. To what extent do you want our country to be democratic now? (AsianBarometer)

WVS Items Dropped from the Dataset

Survey responses from the following items-year-country combinations were not included in the analysis due to evidence of, or suspicion of, poor translations and severe bias (see Kurzmann 2014):

- Vietnam: Army rule 2001; Strong leader 2001
- Albania: Army rule 1998
- Indonesia: Army rule 2001 & 2006
- Iran: Army rule 2000; Strong leader 2000 & 2005
- India: Strong leader, all years.
- Pakistan: Army rule 1996 & 2001; Strong leader 1996 & 2001
- Kyrgyzstan: Strong leader 2003 & 2011
- Romania: Strong leader 1998, 2005 & 2012
- Egypt: Strong leader 2012