Estimating 'Smooth' Country-Year Panels of Public Opinion

Christopher Claassen*

July 1, 2017

Abstract

At the microlevel, comparative public opinion data is abundant. But at the macrolevel, the level where many prominent hypotheses in political behavior are believed to operate, data is scarce. In response, we develop a Bayesian dynamic hierarchical 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 illustrate this model with findings from the new smoothed country-year estimates of support for democracy.

Work in progress – comments and suggestions are welcome!

*School of Social and Political Sciences, University of Glasgow (christopher.claassen@glasgow.ac.uk).
Thanks are due to Anthony McGann and Richard Traunmüller for helpful comments.

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 (Diamond 1999). Yet at the aggregate level, a typical survey sample of one or two thousand respondent diminishes to a single data point. Thus, although we may have millions of respondents' worth of data on 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. Even if measures of all the factors that might confound an observed relationship between, for example, social trust and quality of governance existed, we would run out of degrees of freedom to conduct our empirical tests.

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. 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, adjusting 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, none have done so for cross-national panels of opinion. The primary contribution of this paper is the development and validation of such a method, which should be of interest to wide variety of scholars who might be interested in the obtaining cross-national measures of ideology, policy mood, social values, political culture, and so on. A second contribution of this paper is 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 measured "policy mood" or national ideology in the United States over the period from 1956 to 1988. To accomplish this goal, he developed an ingenious dyad-ratios algorithm. Although the average level of respondent agreement varies idiosyncratically across survey items, the degree to which levels of agreement change over time reduces these idiosyncratic items to a common metric. In other words, the dyad ratios algorithm 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 numerous major studies of public opinion (Stimson, Mackuen, and Erikson 1995;

Erikson, Mackuen, and Stimson 2002; Bartle, Dellepiane-Avellaneda, and Stimson 2011), and scholars continue to use it to this day to estimate smooth time-series of aggregate opinion (e.g. Enns 2016). However, two years before Stimson, Beck (1989) provided an alternative, 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 modelled 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.[1]

Yet, in recent years, methodologists interested in measuring latent aggregate opinion have moved back to the kind of dynamic measurement models advocated by Beck. This is partly a function of vastly increases in computing power at our disposal, but also the corresponding rise of Bayesian methods, which use prior information to help estimate and identify such complex models. Jackman (2005) for example, constructs a dynamic Bayesian model of elections polls. Although "essentially a Kalman filter", a Bayesian approach provides an arguable more intuitive window on to dynamic models of latent opinion. Jackman 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, weighted by an item-specific factor loading, an item intercept or bias parameter, and random error. They also use Bayesian MCMC methods to estimate their model.

The literature up to this point focuses on the proportion of survey samples responding a particular way, and models this proportion in a linear fashion. Linzer (2013) takes a different

[1] Other scholars proposed similar models for estimating smooth opinion time series, notably, Green, Gerber, and De Boef (1999), but these are invariably simpler than Beck's proposal.

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. It 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 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 line of research has focused on estimating an opinion time series from a single country. No researchers have attempted to extend these models to measure opinion across countries as well over time. If the data are 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, there would probably not be enough data points in many cases to reliably estimate both the country-specific latent effects, as well as the item-specific effects, such as item intercepts and slopes, that would would want to include.

Although no scholars have modeled opinion across country and time, a recent paper by Caughey and Warshaw (2015) does develop a model for estimating opinions over time and across (subnational) geographic districts. To do so, they combine a binomial IRT model, as used by McGann, with the method of multilevel regression and poststratification (MRP). 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 sophisticated, but also very complicated.[2] Although it 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 models for doing so are developed and presented in the next section.

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

3. Modeling Cross-National, Time Series Latent Opinion

What do we require of a model of cross-national, time-series opinion? There are three 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. 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, scholars such as Beck (1989) and Voeten and Brewer (2006) extend such models by additionally smoothing over time. Our proposal is to extend latent variable models further to, in effect, smooth over survey item, time, and space.

The second principle we adopt from the literature follows from this discussion: to smooth over time, we require a model of temporal dynamics. If we had enough data we could estimate the model anew in each time period. A more general model uses a dynamic specification to overcome sparseness in opinion time series.

The third principle we take from the literature is that we should model the number of respondents, not the derived proportion or percentage, who offer an affirmative (or dissenting) opinion. This implies a binomial model linking observed responses and an estimated probability parameter. This allows for sampling error to be neatly included. It also allows for responses proportions to be non-normally distributed.

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 political behavior 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 in order to select a preferred model. Using the estimates obtained from this model, we then examine the estimated levels of support for democracy for certain countries and points in time.

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. Given numerous source of possible overdispersion on cross-national survey data, including the biases introduced by survey projects, translation, fieldwork, survey mode, etc., allowing for additional variance may be beneficial.

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 :

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

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

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):

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

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

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 item intercepts hierarchically, with an expectation (μ_λ) and variance (σ^2_λ) to be estimated. This hierarchical specification shrinks the item intercepts towards the mean, to the extent that data are scarce. This guards against small within-item samples producing extreme estimates of item effects.

A recurring theme in the cross-national study of public opinion is that survey items may function differently in different countries (e.g Ariely and Davidov 2011). For example, one method of measuring support for democracy is to ask respondents for their opinions about having the military 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 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:

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

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

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. In addition, by treating the item and item-country effects as varying intercepts, or random effects, we can interpret them as error terms, which lends an intuitive understanding to their role in the measurement equation. Thus 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, classical linear measurement models, such as the common factor analysis model, also include item slopes. These are generally referred to as "factor loadings." These slopes allow the relationship between observed items and latent traits to vary across the item. 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 loadings:

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

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.

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

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.

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; Martin and Quinn 2002), the temporal evolution of latent opinion is specified as a simple local-level dynamic linear model (Durbin and Koopman 2012), where current opinion is a function of last year's opinion 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, I fixed the expectation of the item intercepts, μ_λ, to 0.5, and the expectation of the item slopes, μ_γ, to 1.[3] 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.

[3]For these models λ_1 was not constrained.

4. Application: Support for Democracy

4.1. The Concept of Support for Democracy

Political theorists have long thought that the presence or absence of a democratic political system is somehow related to the attitudes and orientations of the citizenry.[4] In modern times, interest in the cultural requisites of democracy led to the development of the concept of political culture. A democratic political culture is typically conceived of a set of beliefs and values, shared widely among the citizens, that offer support and succor to the institutions and processes of democracy (Eckstein 1966; Inglehart and Welzel 2005).[5]

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. 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 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."[6]

4.2. Data on Support for Democracy

I collected all the available nationally-aggregated responses to questions on support for democracy that have been fielded in cross-national survey projects utilizing representative national

[4]For example, Aristotle argued that societies with egalitarian beliefs provided more fertile terrain for democracy to grow and flourish. Montesquieu similarly thought that the nature of the political regime reflected the dominant orientation of the public, with democracy being congruent with civic orientations.

[5]An alternative point of view holds, instead, that it is the interests of powerful domestic and international actors, not the beliefs of ordinary citizens, that really matter for democratic emergence and survival (e.g. Acemoglu and Robinson 2005; Gleditsch and Ward 2006).

[6]Support for democracy must be distinguished from the concept of satisfaction with democracy, measures of which are also widespread in comparative survey projects. Although some authors (e.g. Armingeon and Guthmann 2014) treat the two as interchangeable, most scholars do not. See (Booth and Seligson 2009) and (Dalton 2004) for a justification.

samples of citizens. Data is collected from nine survey projects including the World Values Survey, Pew Global Attitudes surveys, and all the Global Barometer projects. 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.[7]

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 3,048 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 34 different survey questions clustered within nine broad measurement approaches.[8] 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 sup-

[7]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.

[8]There is an element of subjectivity in determining whether an item can considered to be different because of slight changes in the wording of questions or response sets. We took a conservative route and defined an item as any survey question that was unique *or* was used by a different survey project. Our definition thus allows our item effects to capture variation induced both by question wording and by idiosyncrasies in the methodology of the various survey projects.

port 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 2,745 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. I 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.

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 500 samples each, with the first 250 samples in each chain used for warm up, and discarded, and the remaining 1,000 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.

To conduct the external validation, the full dataset of aggregate survey responses was randomly split into an 75% training set and a 25% test set. The six models were rerun using Stan and the training set. Posterior estimates of model parameters were extracted and used to predict the proportion supporting democracy for the 733 test data points.

4.4. Empirical strategy

To select a model, we compare the accuracy and efficiency of the six models using a number of 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 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

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

Although internal validation is fairly simple to conduct, it favors more complex models. As such, relying solely on internal validation could lead to the selection of a model that overfits to the data set at hand, rather than providing a more general best choice. In response to this problem, analysts often use some sort of information criterion. These penalize models for the number of parameters. As such, they are attempts to measure the model that would have the lowest out-of-sample predictive error. We use the "Leave-One-Out" (LOO-IC) information criteria, which Vehtari, Gelman, and Gabry (2017) argue to be superior to alternatives such as DIC and WAIC.

Better still, however, is to test models and select the best using external validation (Hastie, Tibshirani, and Friedman 2009). We thus split the dataset into an 75% training set and a 25% test set. We fit the models using the training set, and use their parameter estimates to predict observed proportions supporting democracy in the 25% test set. We again calculate the mean absolute error, but now in predicting the error in the j held-out data. We also examine the efficiency of our models by calculating their credible interval coverage. To do so, we count the number of times each of the J observed survey proportions is covered by the 80% credible interval of the corresponding estimate

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

5. Results and Validation

5.1. Validation

These results of our internal and external validation tests are displayed in Table 2. Beginning with the tests of internal validation, four findings are apparent. First, all six models offer a substantially better fit to the observed proportions supporting democracy than the two naive "models" of item or country means. This is hardly surprising given the crudeness of 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.

Perhaps more interesting is the second finding, which is that models 1 and 4 – the models including only item 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 with the LOO information criterion. Although this criterion 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.[9] There is however little to distinguish between the models with item slopes (3 and 6) and the models without (2 and 5).

Finally, the results of the internal validation MAE suggest that the simpler binomial models (1–3) are slightly better fitting than the beta-binomial models (4–6).

The tests of model fit using the 25% hold-out sample offer a better means of gauging model fit as the models are fit and tested using different datasets. We see that the six models continue to offer improvements in accuracy when compared with the naive item- or country-mean methods of estimating support for democracy. The gap, however, has diminished. Our

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

models estimate far more parameters than the naive methods. This test of external validation suggests that part of the improvement we saw in the internal tests of validation were simply due to overfitting.

Yet all six models are substantially more accurate than the naive 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. Adding item slopes or factor loadings also does not improve predictive accuracy at all (and perhaps slightly harms it).

In addition, the tests of external validation diverge again when it comes to comparing the binomial (models 1–3) and beta-binomial (models 4–6) specifications. The latter are now slightly more accurate than the former.

Finally, I turn to the tests of credible interval coverage. These measure the accuracy of the estimates of uncertainty produced by each model. To do so, we find the percentage of times each of the 733 observed survey proportions, in the test sample, is covered by the 80% credible interval of the corresponding estimate. The best estimates here should converge on the nominal level of the credible interval, which are 80% in this case. Coverage substantially below this nominal level shows inaccurate and overly precise standard errors; coverage substantially above this level indicates inaccurate and inefficient uncertainty estimates.

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%. One interpretation of this results is that we have only explicitly modeled a few of the many sources of error in cross-national survey data. The beta-binomial specification, however, proves to have substantially better uncertainty estimates than the simpler binomial specification. In addition, including item-country random effects results in better credible interval coverage. Models 5 and 6, with both item-country intercepts and a beta-binomial distribution, show by far the most accurate estimates of uncertainty.

Weighing up all the evidence from Table 2, it is clear that model 5 with a beta-binomial specification and random item and item-country effects, is the best overall. It shows joint lowest error in predicting aggregate survey responses in the 25% hold-out sample, 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. Descriptive Results

Finally, we examine the results of the selected model in more detail. In particular, we examine the estimated levels of support for democracy for a selection of eight countries. These estimates are displayed in Figure 2. Each plot shows the latent support for democracy estimates for a particular country over all 24 years. The dark blue line shows the mean value of θ in each year; the light blue lines show 200 random draws from the posterior density of θ, and indicate uncertainty in the estimates. I have also included the observed data on these plots, indicated using grey points.[10]

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 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, which reduces the difference in the precision of estimates for years with data and years without. The binomial specification (not shown) results in more precise estimates for years with data. This produces a more jagged, rapidly changing pattern of estimated opinion. Such a pattern is not particularly plausible for a measure of political culture, such as support for democracy, as these are typically thought to be slow-moving (at the individual level) orientations. As we have already noted, the binomial specification also leads to lower predictive accuracy and worse credible interval coverage.

Figure 2 focuses on eight time series that were selected as representing a range of levels of support and interesting dynamics. For example, in Venezuela support for democracy appears to have risen after Chavez began entrenching his power in the mid-2000s. Venezuelans now

[10]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). I 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.

exhibit one of the highest levels of support for democracy in the world. In contrast, the Egyptian results suggest that support for democracy *fell* substantially in the years before and after the Arab Spring. Finally, our estimates from the United States appear to confirm the claim made recently by Foa and Mounk (2016), based on an analysis of a subset of the survey data, that support for democracy has declined, and now is only at middling levels.

To complement this view of the data, I also provide a cross section in the most recent year for which we have survey data, 2015. These estimates for all 127 countries are plotted in Figure 3. In the figure, the points indicate the mean estimate of θ for each country in 2015, with 90% credible intervals shown using horizontal bars. These results indicate that support for democracy is highest among Northern European bastions of democracy such as Norway and Germany, which is consistent with existing research focusing on particular subsets of the available survey data (e.g. Klingemann 1999). The figure also shows that some of the more-democratic African countries, such as Mauritius and Senegal, now have among the highest levels of support for democracy in the world. Support is low in Russia, the Philippines, and Pakistan, partially-democratic countries with checkered records of maintaining democratic procedures, but also Saudi Arabia and Vietnam, which are longstanding autocracies.

6. Conclusion

This paper develops and tests models for estimating smooth country-year panels of opinion. We find that a beta-binomial specification is preferable to a simpler binomial model. Both should of course be preferable to modeling the proportions directly under the assumption of a linear model. We also find that item-slopes or factor loadings do not increase the accuracy of our models, but do increase the computational complexity, so we recommend avoiding these. However, when a particular item is asked several times in a particular country, we recommend including item-country effects, as these increase accuracy dramatically.

Our selected model is fairly accurate, with predicted responses in the held-out dataset that deviate from the true proportions by six percentage points. This compares favorably to a baseline error rate of 11 percentage points, obtained by using only the item or country means to predict.

Our selected model also shows uncertainty interval coverage that approaches the nominal 80% level. The coverage for all our models falls short of the nominal level, sometimes dramatically, indicating variance estimates that are too precise. This shows the challenges of estimating cross-national public opinion given the numerous sources of bias that are present in such data.

Future research might consider modeling additional sources of bias and variance. In particular, analysts could investigate the inclusion of spatial effects, which, although computationally challenging, allow unobserved heterogeneity to be modeled among neighbors or within regions.

References

Acemoglu, Daron, and James A. Robinson. 2005. *Economic Origins of Dictatorship and Democracy*. Cambridge University Press.

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

Ariely, Gal, and Eldad Davidov. 2011. "Can we Rate Public Support for Democracy in a Comparable Way? Cross-National Equivalence of Democratic Attitudes in the World Value Survey." *Social Indicators Research* 104(2): :271–286.

Armingeon, Klaus, and Kai Guthmann. 2014. "Democracy in Crisis? The Declining Support for National Democracy in European Countries, 20072011". *European Journal of Political Research* 53(3): 423–442.

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.

Bartle, John, Sebastian Dellepiane-Avellaneda, and James Stimson. 2011. "The Moving Centre: Preferences for Government Activity in Britain, 1950–2005." *British Journal of Political Science* 41(2): 259–85.

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

Booth, John A., and Mitchell A. Seligson. 2009. *The Legitimacy Puzzle in Latin America: Political Support and Democracy in Eight Nations*. Cambridge: Cambridge University Press.

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

Claassen, Christopher. 2015 "Measuring University Quality." *Scientometrics* 104(3): 793–807.

Dalton, Russell. 2004. *Democratic Challenges, Democratic Choices: The Erosion of Political Support in Advanced Industrial Democracies*. Oxford: Oxford University Press.

Diamond, Larry. 1999. *Developing Democracy: Towards Consolidation*. Baltimore: Johns Hopkins University Press

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

Eckstein, Harry. 1966. *A Theory of Stable Democracy*. Princeton University Press.

Enns, Peter K. 2016. *Incarceration Nation: How the United States Became the Most Punitive Democracy in the World.* New York: Cambridge University Press.

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

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

Gelman, Andrew. 2006. "Prior Distributions for Variance Parameters in Hierarchical Models." *Bayesian Analysis* 1(3): 515–33

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

Gleditsch, Kristian, and Michael D. Ward. 2006. "Diffusion and the International Context of Democratization." *International Organization* 60(4): 911–33.

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.

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.

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

Martin, Andrew D. and Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999." *Political Analysis* 10(2): 134–153.

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.

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

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.

Table 1. Models

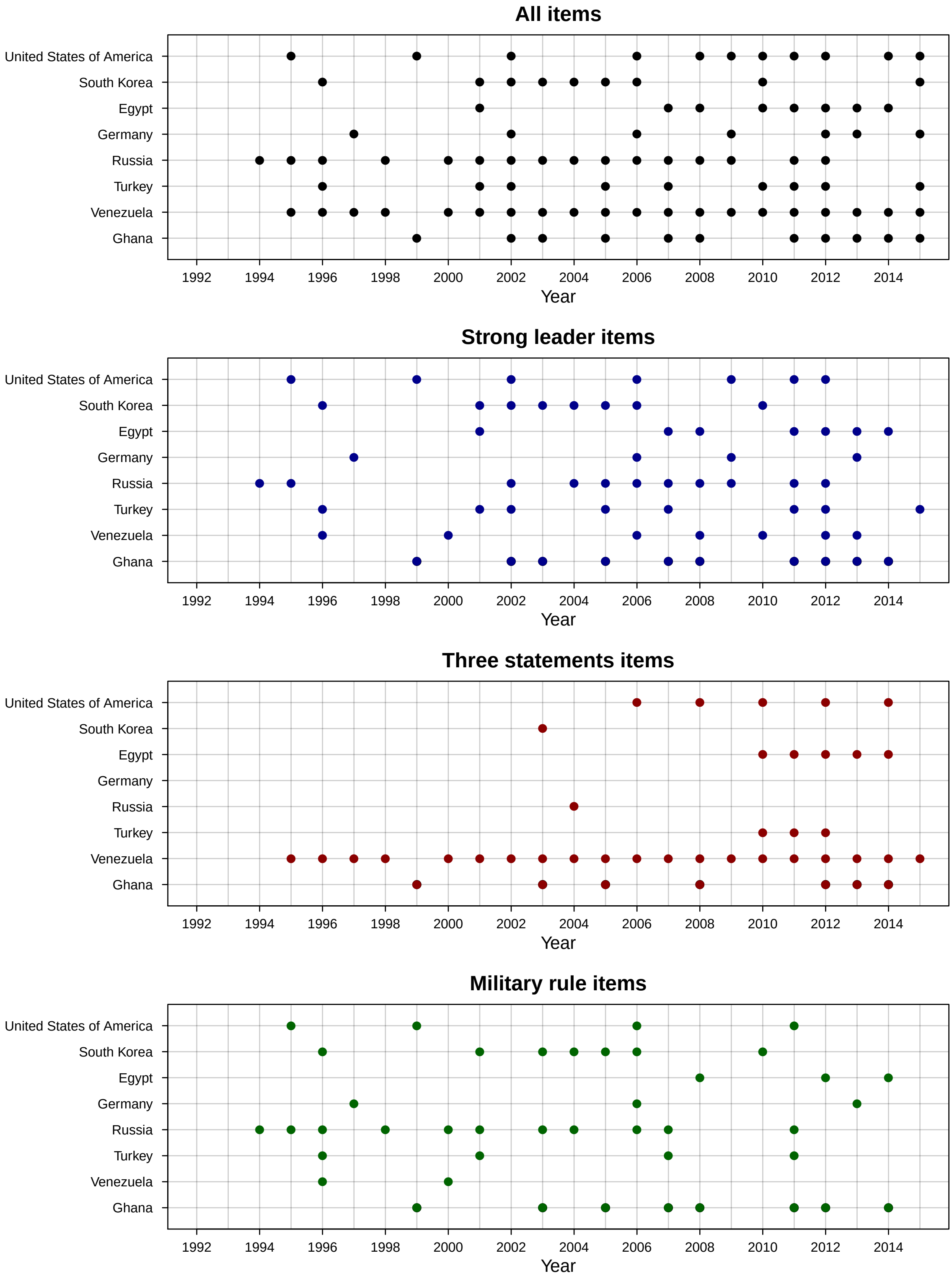
Model number	Response distribution	Item intercepts (λ)	Item-country intercepts (δ)	Item slopes (γ)
1	Binomial	✓		
2	Binomial	✓	✓	
3	Binomial	✓	✓	✓
4	Beta-binomial	✓		
5	Beta-binomial	✓	✓	
6	Beta-binomial	✓	✓	✓

Table 2. Internal and External Validation Tests

Model	Internal validation		External validation	
	MAE	LOO-IC	MAE	CIC
1	.050	143676	.080	17.1
2	.023	60561	.065	37.7
3	.022	58849	.076	45.7
4	.062	35956	.072	36.3
5	.032	34375	.060	61.7
6	.032	34354	.062	60.6
Country means	.106		.113	
Item means	.096		.098	

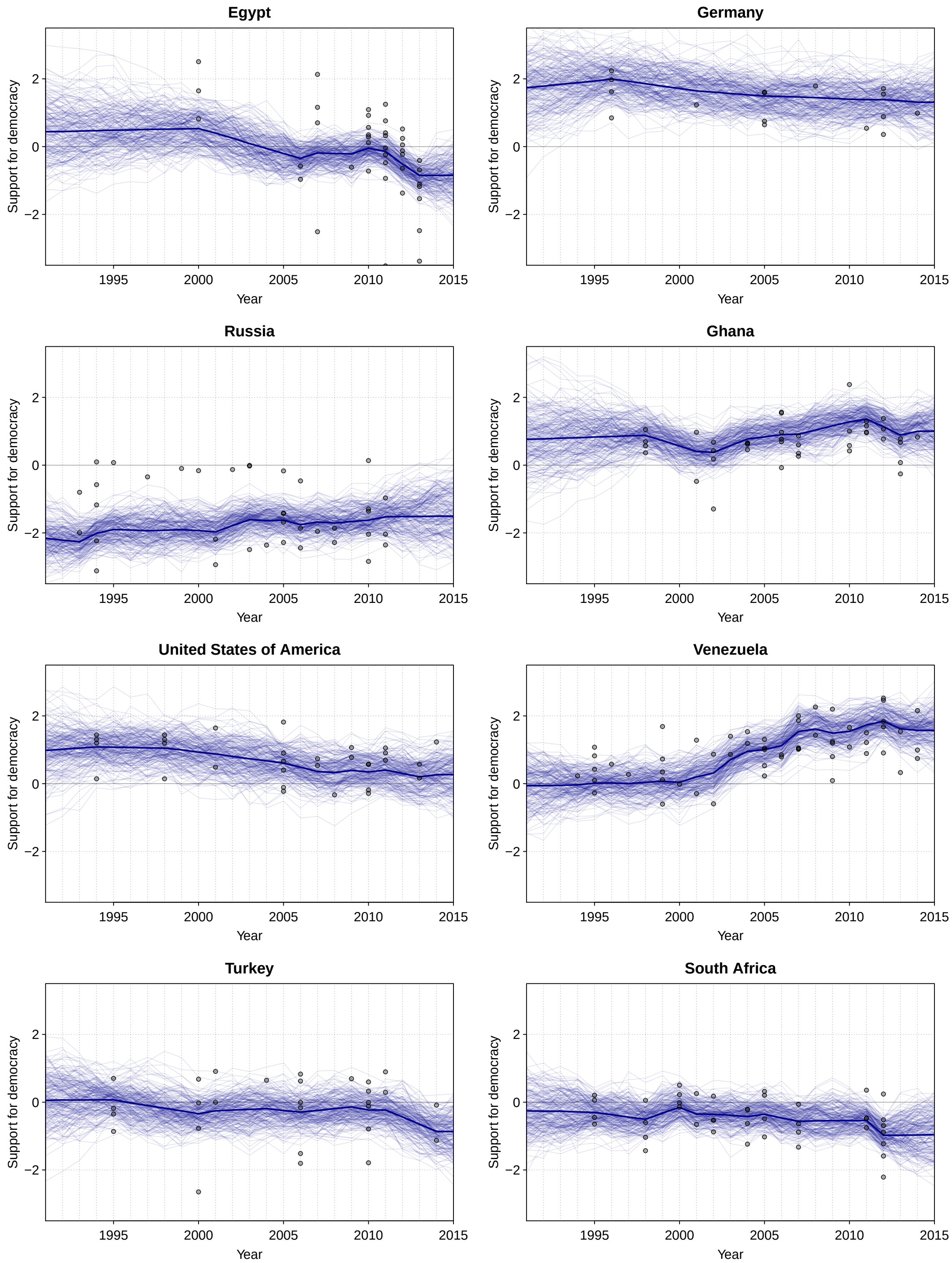
Internal validation includes tests of model fit using the same data for model fitting and validation. External validation creates two separate datasets: models are fit to training set and validated using test / hold-out set. MAE: mean absolute error; LOO-IC: leave-one-out information criterion; CIC: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

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



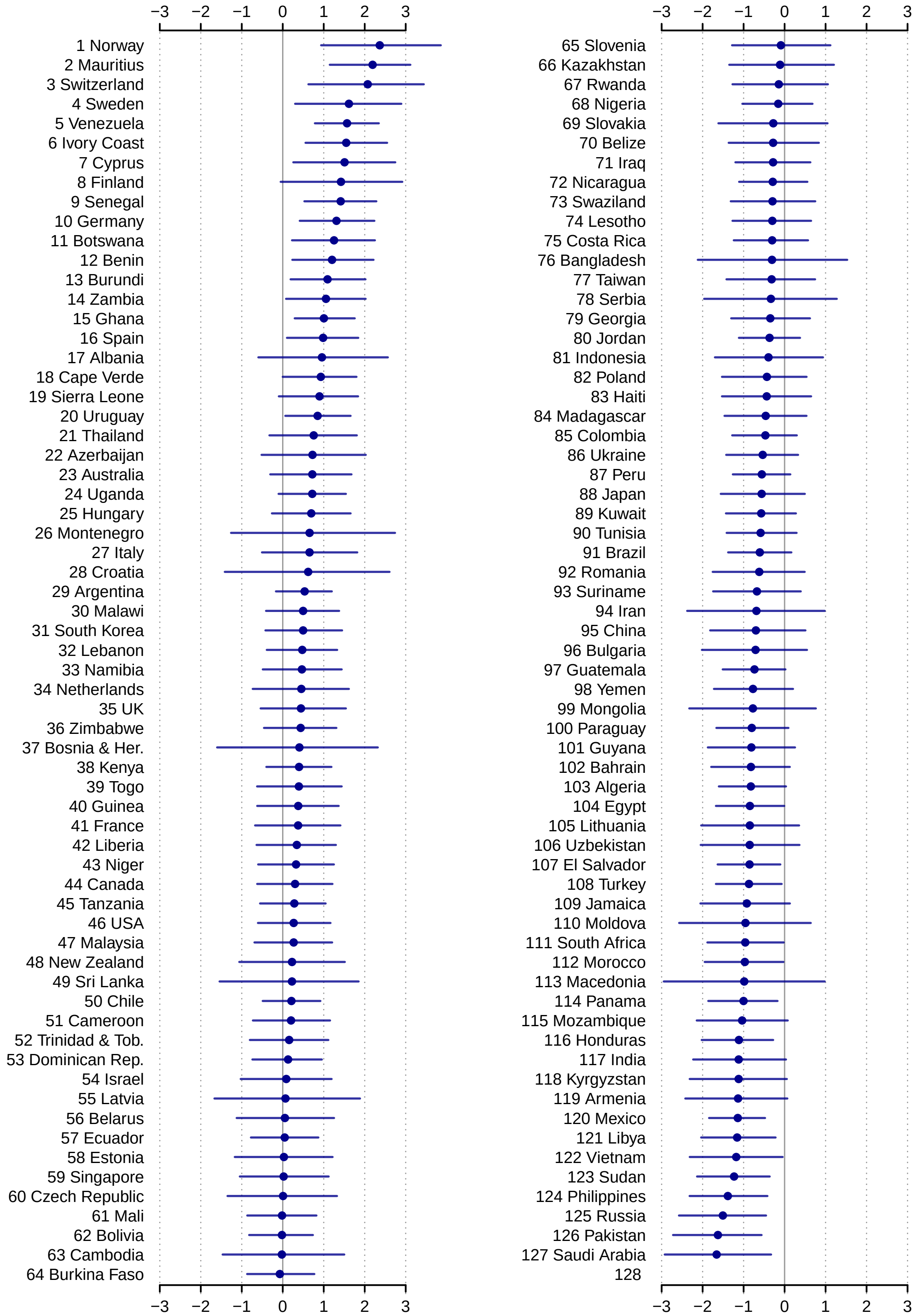
The top panel shows the availability of at least one survey item for a particular year and country. The lower three panels indicate the availability of data for the three most common question themes. The seven countries are selected to show variation by region and data availability.

Figure 2. Estimated Support for Democracy for 8 Countries over 24 Years



Estimates of support for democracy, from Model 5, for eight selected countries and 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.

Figure 3. Estimated Support for Democracy for 127 Countries in 2015



Estimates of support for democracy, from Model 5, for all countries in 2015. The points show the mean of the posterior density of θ_i for country i; the horizontal bars, the central 90% intervals of θ_i.

Appendix A: 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).
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 (AfroB).
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 (LatinoB).
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 (New Democ B).
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 (LAPOP).
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 (AsianB)

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? (LAPOP)
8. Democracy may have its problems, but it is better than any other form of government. To what extent do you agree or disagree? (WVS)
9. Democracy may have its problems, but it is better than any other form of government. To what extent do you agree or disagree? (ArabB)
10. 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 (LatinoB)
11. 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 (AsianB)
12. 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. (LAPOP)

Strong leader items

13. Best to get rid of Parliament and elections and have a strong leader who can quickly decide everything. What do you think? (NDB)
14. Best to get rid of Parliament and elections and have a strong leader who can quickly decide everything. What do you think? (AsianB)
15. 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? (LAPOP)
16. 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? (LAPOP)

17. 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. (AfroB)
18. 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 (WVS)
19. 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)
20. 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? (ArabB)
21. 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? (ArabB)
22. 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 (AsiaB)

Military rule items

23. The army should govern the country. What do you think? (NDB)
24. The army should govern the country. What do you think? (AsianB)
25. 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 (AfroB)
26. 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 (WVS)
27. 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 (AsiaB)

One party rule items

28. 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 (AsianB)
29. 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 (AfroB)
30. No opposition party should be allowed to compete for power (AsianB)

Evaluate democracy items

31. 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). (ArabB)
32. 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). (WVS)
33. 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). (AsiaB)

Democracy suitable items

34. 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? (AsianB)
35. 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? (ArabB)

Importance to you items

36. 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)
37. 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? (WVS)
38. How important for you to live in democratically governed country? (ESS)

Desire for democracy items

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