Exploring the Dynamics of Latent Variable Models

Kevin Reuning
Miami University of Ohio

Michael R. Kenwick
The University of Pennsylvania

Christopher J. Fariss[1]
The University of Michigan

[1]An earlier version of this paper was presented at the annual meeting of the American Political Science Association in Philadelphia, PA (2016) and the Latent Variable Mini-Conference at the Varieties of Democracy Institute at the University of Gothenburg, Sweden (2016). We would like to thank the participants at these conferences and also James Lo, Suzie Linn, Kyle Marquardt, Ryan McMahon, Dan Pemstein, Kevin Quinn, Brigitte Seim, Jeff Staton, Jane Sumner, Alex Tahk, and Anne Whitesell for helpful comments and suggestions. The estimates from this paper along with the code necessary to implement the models in STAN and R are publicly available at a dataverse repository here: https://doi.org/10.7910/DVN/SSLCFF (Reuning, Kenwick and Fariss 2018).

Abstract

Researchers face a tradeoff when applying latent variable models to time-series, cross-sectional data. Static models minimize bias but assume data are temporally independent, resulting in a loss of efficiency. Dynamic models explicitly model temporal data structures, but smooth estimates of the latent trait across time, resulting in bias when the latent trait changes rapidly. We address this tradeoff by investigating a new approach for modeling and evaluating latent variable estimates: a robust dynamic model. The robust model is capable of minimizing bias and accommodating volatile changes in the latent trait. Simulations demonstrate that the robust model outperforms other models when the underlying latent trait is subject to rapid change, and is equivalent to the dynamic model in the absence of volatility. We reproduce latent estimates from studies of judicial ideology and democracy. For judicial ideology, the robust model uncovers shocks in judicial voting patterns that were not previously identified in the dynamic model. For democracy, the robust model provides more precise estimates of sudden institutional changes such as the imposition of martial law in the Philippines (1972-1981) and the short-lived Saur Revolution in Afghanistan (1978). Overall, the robust model is a useful alternative to the standard dynamic model for modeling latent traits that change rapidly over time.

Political scientists are increasingly focused on improving the measurement of unobservable concepts. Recent methodological and computational advances have led to a flourishing of latent variable modeling techniques that provide researchers with a means of measuring these concepts based on events, ratings, or other observed information that are assumed to be manifestations of the unobservable latent trait (Imai, Lo and Olmsted 2016; Martin and Quinn 2002; Carpenter et al. 2016). These models have been applied across a variety of subfields, encompassing the study of political ideology (Barbera 2015; Martin and Quinn 2002; Caughey and Warshaw 2015; Kōnig, Marbach and Osnabrügge 2013; Pan and Xu 2018; Treier and Hillygus 2009), political attitudes, knowledge, and preferences (Blaydes and Linzer 2008; Pérez 2011; Jesse 2017; Stegmueller 2011, 2013), regime type (Treier and Jackman 2008; Pemstein, Meserve and Melton 2010), UN voting positions (Voeten 2000), human rights abuse (Schnakenberg and Fariss 2014; Fariss 2014), human rights treaty embeddedness (Fariss 2018), judicial independence (Linzer and Staton 2016), and institutional transparency (Hollyer, Rosendorff and Vreeland 2014).

In several recent applications, these models are applied to time-series cross-sectional data (e.g., Treier and Jackman 2008; Pemstein, Meserve and Melton 2010; Schnakenberg and Fariss 2014; Fariss 2014, 2018; Linzer and Staton 2016). Two modeling strategies are commonly applied to accommodate this data structure, but both face limitations when confronted with certain types of temporal dynamics. Until recently, the most commonly used latent trait model for time-series cross section*al data assumes that each observed time period is an independent observation. This type of model is usually referred to as static. For many applications, this assumption is appealing because estimates of the latent trait are determined exclusively by the observed data for a particular unit at a given time point. As a result, static models easily accommodate rapid changes in latent traits over time. Yet, the static approach ignores the time-series nature of the data and imposes a tenuous assumption of local independence with respect to time.

More recently, the "dynamic approach" has emerged as a useful alternative (Jackman

2009, 471-485). Here, the time series structure of the data is accommodated by modeling the latent trait as a random walk or through a local linear model. In Bayesian analysis, this is typically modeled through the prior information assigned to the latent trait for units at each time point. In doing so, these models smooth the estimates over time periods for a given unit. This smoothing is often theoretically appealing and leads to more efficient estimates (Armstrong II et al. 2014, 303-309).

There are limitations to both the static and dynamic strategies when modeling a potentially volatile latent trait. Dynamic models will over-smooth these transitions, causing rapid changes to appear as gradual transitions, thus biasing the resulting estimates. Static models accommodate sudden changes but at the loss of efficiency and without directly modeling the temporal dynamics, which are often of substantive interest. Assuming away the dynamic structure of the model, leads to a loss of important information, which makes evaluating and testing hypotheses involving latent concepts more difficult. These limitations are important for political scientists as many political constructs are characterized by punctuated equilibria or similar theoretical processes where rapid change is possible after long periods of stasis (e.g., Leventoğlu and Slantchev 2007). Political institutions, for example, often persist for generations only to collapse or change suddenly (e.g., Grief and Laitin 2004). Exogenous shocks may also result in concomitant changes in individual level-traits such as political preferences or partisan identification (e.g., Baker et al. 2016).

In this article, we introduce a robust version of the standard dynamic model. We demonstrate that the robust model retains the theoretically useful features of the dynamic model while reducing bias in the estimates that result from over smoothing. This model changes the distributional assumptions conventionally applied to the latent trait while still incorporating temporal information. After a brief discussion of these three models, we evaluate their relative performance. We find that the robust model outperforms conventional modeling choices across a variety of metrics. Further, we find that the robust model performs as well as the dynamic model, even when simulated conditions match the assumptions underlying the dynamic model.

Next, we reproduce and evaluate two common latent variables in political science: judicial voting preferences (Martin and Quinn 2002) and democracy (Pemstein, Meserve and Melton 2010). We find that the estimates from the robust latent variable model reveal previously unobserved shocks and improved fit for both constructs. Substantively, the robust model provides evidence that judicial voting patterns sometimes change rapidly for a subset of Supreme Court justices. We also find that the robust model provides more precise estimates of democracy in response to sudden political events such as the imposition of martial law that occurred in the Philippines from 1972 through 1981 under the regime of Ferdinand Marcos and the short-lived Saur Revolution in Afghanistan in 1978. The analyses indicate that the robust latent variable model yields substantively meaningful insights not obtainable from existing latent variable modeling strategies.

Modeling Strategies and Simulation Analysis

We begin by briefly discussing the static and dynamic modeling strategies and introducing the robust modeling alternative. We also outline a simple data generating process and the priors which we will use in a simulation analysis of the competing models.

We begin with a simple item response theory framework (IRT) with binary manifest variables.[1] The latent trait, θ_{it} exists for each unit, $i = 1, \ldots, N$ across each time period $t = 1, \ldots, T$. θ_{it} is not observed directly, but determines the value of a series of manifest variables, or items, y_{kit}, where $k = 1, \ldots, K$ indexes the number of items observed. Thus, y_{itk} is the observed value for item k for unit i at time t. For each item, α_k and β_k are estimated. These are commonly referred to as the "difficulty" and "discrimination" parameters, respectively and are analogous to an intercept and slope in traditional logit regression. The likelihood

[1]The models generalize to other types of manifest variables as well.

function takes the following form, where Λ is the logistic function:

$$\mathcal{L} = \prod_{i,t=1}^{N,T} \prod_{k=1}^{K} \Lambda(\alpha_k - \beta_k\theta_{it})^{y_{itk}} \big(1 - \Lambda(\alpha_k - \beta_k\theta_{it})\big)^{1-y_{itk}}$$

We place weakly-informative, normal priors on the difficulty and discrimination parameters. In addition, β_k, the difficulty parameter, is constrained to be greater than 0. This is a common identification constraint that resolves rotational invariance by preventing "flipping" where the likelihood from $\hat{\beta}_k$ and $\hat{\theta_{it}}$ is equal to the likelihood from $-\hat{\beta}_k$ and $-\hat{\theta_{it}}$. All the items are then coded so that a 1 indicates increased levels of the latent trait, and a 0 decreased levels.[2] Formally our priors are:

$$\alpha_k \sim \mathrm{N}(0, 3)$$

$$\beta_k \sim \mathrm{HN}(0, 3)$$

HN is the half-normal distribution, with support on $[0, \infty)$. All models presented here share these priors on their item parameters.

Static Model

The three modeling strategies we present are differentiated by the prior information assigned to the latent variable. The static model, places a standard normal prior on all units for all time periods:

Static Model Prior

$$\theta_{it} \sim \mathrm{N}(0, 1) \quad \forall i = 1, \ldots, N \quad \& \quad \forall t = 1, \ldots, T$$

[2]When the polarity of an item is not known then other identification constraints must be considered by the analyst (Martin and Quinn 2002).

Estimates for each unit in each time period are therefore differentiated exclusively by the values of the manifest variables for that unit at that time period. This allows for sudden changes in the latent estimates in a unit between time periods. A shortcoming of this model is that it treats each observation as independent. In the case where the manifest variables contain sufficient information on the latent trait, this modeling strategy may not be problematic. Unfortunately, this is seldom the case when using social science data, where indicators are often coarse or missing. As a result, these indicators often do not contain sufficient information to differentiate between theoretically distinct units.

Standard Dynamic Model

To address temporal non-independence in the data, many researchers use a dynamic prior for the latent trait (Martin and Quinn 2002; Schnakenberg and Fariss 2014; Fariss 2014; Caughey and Warshaw 2015; Kōnig, Marbach and Osnabrügge 2013). The choice of a "random walk" prior on the latent variable is particularly common.

Researchers apply a standard normal distribution to the latent trait in the first observation period for every unit. For each subsequent time period, the prior is normally distributed with mean $\theta_{i(t-1)}$, and an innovation standard deviation σ which is either assigned by the researcher or, more commonly, estimated from the data. Here, we assign a weakly informative prior to σ by using a half normal distribution with standard deviation of 3 and mean 0.[3]

Standard Dynamic Model Priors

$$\begin{aligned}
\theta_{i1} &\sim \mathrm{N}(0,1) \quad \forall i = 1,\ldots,N \\
\theta_{it} &\sim \mathrm{N}(\theta_{i(t-1)},\sigma) \quad \forall i = 1,\ldots,N \quad \& \quad \forall t = 2,\ldots,T \\
\sigma &\sim \mathrm{HN}(0,3)
\end{aligned}$$

[3]In our applications, σ is fixed across units, though this assumption can be relaxed to account for unit-specific variation (Imai, Lo and Olmsted 2016) .

This strategy trades the assumption that observations are independent with the assumption that the latent trait will be correlated over time. As a result, estimates from dynamic models typically have less uncertainty because more information is used to estimate each latent variable. This also induces smoothing over time because changes between time periods are constrained. When researchers have theoretical reasons to expect that the latent trait is relatively slow-moving over time, both modeling features can be desirable. If, however, the latent trait is subject to rapid fluctuations or state-changes between time periods, this temporal smoothing can produce biased estimates. The modeling strategy we introduce below is designed to address this problem while still accounting for temporal dynamics.

Robust Dynamic Modeling

Our alternative strategy draws on the robust modeling literature. Robust models, broadly defined, weaken the parametric assumptions common to standard statistical models as a means of accommodating unique data structures and the potential for influential outlying observations. One simple modification commonly used is the substitution of normally distributed errors with Student's t-distributed errors (Gelman et al. 2014; Lange and Sinsheimer 1993; Lange, Little and Taylor 1989). A specific Bayesian approach to the use of the Student's t-distribution has been developed (Geweke 1993; Fonseca, Ferreira and Migon 2008). This strategy has been used to account for heterogenerity in growth models (Zhang et al. 2013), ordinal choice models (Stegmueller 2013), and mixed effects models (Rosa, Gianola and Padovani 2004) among others. Within our context, potential outliers are "shocks" where values of the true latent variable change suddenly within a unit's time series. As a prior on the latent variable θ, the Student's t-distribution is less-restrictive when estimating the temporal data generating process that relates to the latent trait with the observed manifest variables because it has greater density in the tails when compared to the standard normal distribution (for details on the Student's t-distribution see Appendix A).

To develop a robust dynamic model, we set the prior for the first year in a particular time series to a standard normal distribution.[4] For every subsequent year, the prior then follows a Student's t-distribution, with four degrees of freedom. Setting the degrees of freedom to a relatively low value increases the density of the tails of the distribution which allows "extreme values" to be estimated from time period to time period. Thus, the model smooths estimates across time during periods of stability, but also allows for rapid changes in the latent trait during periods of volatility. In the cases presented here we set the degrees of freedom and estimate the scale parameter σ. In Appendix A we discuss an alternative model specification where we estimate the degrees of freedom parameter instead of σ, as well as a model that jointly estimates both parameters. We find that this does not improve model fit while potentially compounding rotational identification concerns in some cases and being computationally more taxing. Because of this, we opt to set the degrees of freedom at a low value of 4, as recommended by Gelman et al. (2014).

Robust Dynamic Model Priors

$$\begin{aligned}
\theta_{i1} &\sim \mathrm{N}(0,1) \quad \forall i = 1,\ldots,N \\
\theta_{it} &\sim \mathrm{T}_4(\theta_{i(t-1)},\sigma) \quad \forall i = 1,\ldots,N \quad \& \quad \forall t = 2,\ldots,T \\
\sigma &\sim \mathrm{HN}(0,3)
\end{aligned}$$

Potential Alternatives

In addition to the robust model there are several alternative modeling strategies that could be used to accommodate temporal shocks. First, given an assumed data generating process that is a combination of a dynamic process and a static process a finite mixture modeling

[4]In practice, one can also substitute a Student's t-distribution with a very high degrees of freedom (e.g. 1,000), which closely approximates the normal distribution.

strategy might be a useful strategy. We assess such a model in Appendix B and find that the mixture model is outperformed by the robust model in almost all cases.

Another potential strategy would be to employ a change-point model (Western and Kleykamp 2004) or regime switching model (Hamilton 2010). These models capture the intuition that there is a discrete change in the underlying process. However, estimating these types of models requires a priori attention to the number of expected change-points and are commonly employed only in time-series analysis (e.g., Pang et al. 2012) but have seen some use in panel data (e.g., Spirling 2007; Joseph et al. 1997). One of the main modeling challenges is estimating the discrete number of change-points. This issue is particularly intractable for the types of punctuated equilibrium processes discussed here where the number of shocks or change-points that are likely to exist in a given time series are not known ex-ante and vary across panels. In the univariate case, there have been some very useful developments in models without a fixed number of change-points (Santifort, Sandler and Brandt 2013). Other alternatives also exist for continuous time latent variables that do not assume discrete time intervals (Tahk 2015) as well as latent growth models which provide a flexible framework for specific theoretically driven assumptions about latent change over time (Duncan and Duncan 2004). Evaluating all of these models is beyond the scope of this paper but each offers important opportunities for generating new inferences about unobservable concepts.

Simulation Analysis

To evaluate model performance we simulate conditions where the true data generating process is known. We focus in particular on situations where there is some time-series structure to the underlying latent trait. We do not evaluate any models in which there are no temporal dependencies in the underlying data generating process (e.g. the static model) given our focus on improving modeling techniques for data with dynamic temporal structures. While latent traits are sometimes temporally independent and identically distributed, we focus on more common conditions when this is not the case. In all our simulations, the latent trait is drawn from a standard normal distribution in its first time period and follows a random-walk process with the probability of a "shock" thereafter. When a unit experiences a shock, it is re-drawn from a standard normal distribution, thus inserting "breaks" in the time-series such that the latent trait is subject to the possibility of rapid change. From these "latent" variables, we estimate binary manifest indicators that are recorded with error for each observation. With these data in place, we estimate variants of the static, dynamic, and robust models and evaluate their performance under a variety of conditions. Information on these analyses, along with additional information on the underlying data generating process are available in the appendix.[5] Here, we provide a brief overview of our findings.

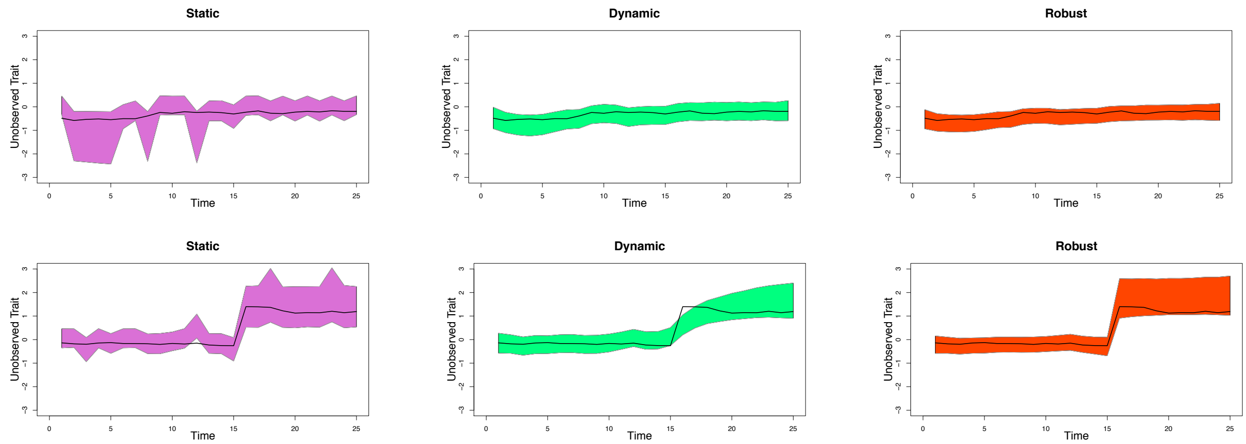
Figure 1 displays the differences in estimates of the latent trait from the static (left column), dynamic (center column), and robust (right column) for two sample units. These units were simulated with the probability of shock was set to 0.01, and the innovation standard deviation set to 0.05.[6] The first of these units (top panels) is representative of cases where the latent trait is relatively stable across time and does not experience a shock. The second (bottom panel) represents a case where a shock to the latent trait occurs after

[5]We provide a more detailed discussion of the data generating process for the simulation in Appendix C. We assess model performance in the time surrounding shocks to the latent trait in Appendix C.2, accuracy in ranking observations in Appendix C.3, within-unit rank correlations in Appendix C.4, cross-validated accuracy in Appendix C.5, and differences between time periods in Appendix C.6. We also consider models that estimate the degrees of freedom parameter from the student's t-distribution instead of the innovation standard deviation in Appendix C.7, and the estimation of both of these parameters in Appendix C.8.

[6]The simulation analyses in the appendix explore model performance with these parameters set to a variety of values. We chose 0.01 and 0.05 for this example because they represent plausible conditions for political science data where temporal variation is often relatively low relative to cross-sectional variation, and where "shocks" occur, but are relatively rare.

period 20.

Figure 1: Model Estimates Across Two Sample Unites

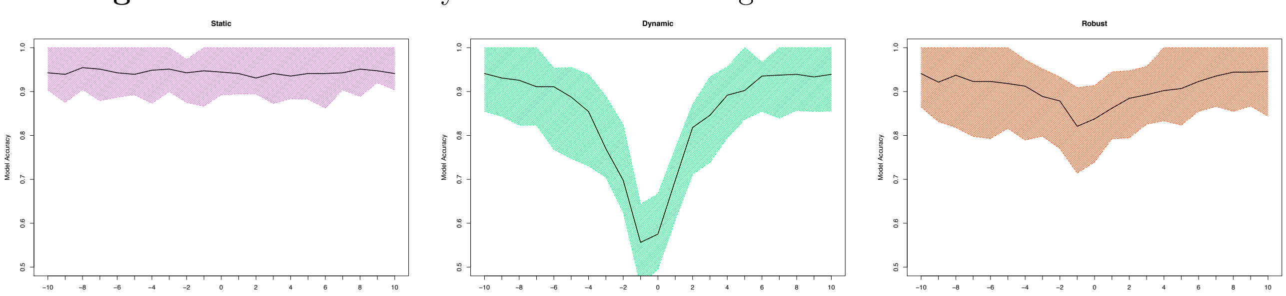


Note: 95% Credible Intervals for two sample units using the four different modeling strategies. The true latent value is displayed with a dark line, and is the same in each column.

The efficiency loss in the static model is apparent for the unit that did not experience a shock. By efficiency loss, we mean a decrease in the amount of information contained in the estimate of the latent variable which translates into a flatter (and less informative) estimated probability density for each unit within the latent space. Although the true latent variable is within the credible interval of the static model across all time points, the credible intervals vary widely around it. This is because estimates at each time period are generated independently and so there is relatively limited information used in each time period. As expected, the dynamic and robust model generate lower estimates of uncertainty by leveraging time-series information in the data. Of particular note is that the use of the Student's t-distribution by the robust model does not lead to a significant increase in estimates of uncertainty relative to the dynamic model. The advantages of the robust model are apparent for the unit experiencing a shock. As before, the static model's credible intervals are wide, but almost always capture the true estimate. By contrast, the dynamic model over-smooths the sudden changes in the time series and therefore fails to produce accurate credible intervals over several time periods before the shock. The robust model, by contrast, produces relatively small credible intervals (in comparison to the static model) and is also capable of accounting for sudden changes to the true latent trait when the shock happens.

Next we show how these results generalize to many simulated units drawn from the described data generating process. In Figure 2 we plot the percentage of observations where the true latent trait value is contained within the 95% credible interval generated for each model as a function of the time before and after a shock to the latent trait (experienced at time $t = 0$). As expected, the static model performs particularly well using this metric due to its large credible intervals. More relevant for our purposes, the model performs equally well across all time periods, regardless of whether and when a unit experienced a shock. By contrast, the dynamic model performs well for units that have not recently experienced a shock, and poorly in the time of or proximate to a shock. This dip in accuracy is significantly mitigated, though not reduced entirely by the robust model. Taken together, the simulation analyses reported here and in the appendix indicate that the robust model is a viable alternative to standard dynamic modeling choices, particularly for latent traits subject to temporal volatility.

Figure 2: Model Accuracy in Time Surrounding Shocks to the Latent Variable



Note: The percentage of times the true latent variable is within 95% credible intervals. The horizontal axis is the distance away from the nearest shock. We omit a small number of observations that were within three time periods of two shocks. A distribution of values is estimated by generating 50 different simulated datasets – the bounds show the 20$^{\text{th}}$ and 80$^{\text{th}}$ percentile.

Substantive Applications

We now assess how the robust latent variable model performs using data from two important applications of latent variable models: a model of judicial ideology from Martin and Quinn (2002) who originally used a dynamic model; and a model of democracy from Pemstein, Meserve and Melton (2010) who originally used a static model. For each, we estimate the original models and compare these original modeling strategies to the robust latent variable model. We assess model fit using the Widely Available Information Criterion (WAIC), posterior predictive checks, and visual assessments of well-known cases.[7]

Martin and Quinn (2002)

Martin and Quinn (2002) design a latent variable model to measure United States Supreme Court justice ideology. The units are Supreme Court justice years and the items are the votes that each justice cast in a particular case. Each vote is coded zero or one depending on whether it overturns or affirms a lower court decision. Justices serve for multiple years, and Martin and Quinn (2002) use a dynamic model with random walk priors to model the position of the justices over time. We replicate this model using the data made available on their website which extends the timespan of the original model to contain the years 1937-2015.

The sparsity of data along with a lack of clear polarity for cases-votes required Martin and Quinn (2002) to achieve identification through two modeling decisions. First, a subset of justices are assigned strong priors to allow them to act as anchors within the policy space. In addition, because of a sparsity of data they–after some experimentation–fix the value of the innovation variance (σ^2) to 0.1.[8]

Fixing the value of σ^2 warrants some additional discussion before proceeding. Often

[7]All models are estimated in Stan (Carpenter et al. 2016) using 4 chains and 2,000 iterations. Convergence diagnostics are in Appendix G.

[8]The authors also set Douglas to 0.001, as he was present at very few cases near the end

researchers estimate this parameter directly from the data, but sparse data structures such as this sometimes lead research to fix σ^2. This value must be chosen carefully. When $\sigma^2 \to \infty$, the model will produce estimates for each period of time independently. Assigning σ^2 a value that is too large may produce a model that is not rotationally identified in cases where there are no expectations about the direction of the relationship between items and the latent trait. When $\sigma^2 = 0$, the dynamic model will produce a single time-invariant estimate of the latent trait. Setting σ^2 to a value that is too small therefore ignores temporal variation of the latent trait. For a standard dynamic model, this is equivalent to fitting a flat line to a time-series.

The robust model is no less sensitive to these modeling decisions and shares many of the same properties as the standard dynamic model. One important exception, is that as σ^2 approaches zero, the robust model may fit an approximately stepwise function to smooth linear trends in the latent trait. This would result in the appearance of shocks in a time series where none is present.[9] We therefore recommend practitioners estimate the value of σ^2 directly from the data when possible. If this parameter must be fixed, we recommend researchers evaluate model performance using multiple values of σ^2. At each step, practitioners should evaluate model performance through a variety of metrics including those employed here: fit statistics such as WAIC; posterior predictive checks (particularly in times surrounding the "shocks" identified by the robust model); and visual inspection of the data.

We use a prior assignment and identification strategy similar to Martin and Quinn (2002), including setting the innovation variance to 0.1 for all justices except Douglas and providing informed priors around a subset justices for their initial year on the court. In addition,

of his term and is therefore much less likely to change positions over time.

[9]Appendix D contains a simulation analysis comparing the dynamic and robust models with fixed values of σ. In Appendix E we evaluate the effect of either inflating or estimating σ^2 for the Martin and Quinn (2002) replication and find no substantive differences.

because we use a more expansive dataset than originally employed by Martin and Quinn (2002) we impose two additional priors on justices at the very end or very beginning of the time series: Samuel Alito is given a prior on θ centered at 1.5 and Hugo Black is given a prior on θ centered at -3. To ensure rotational identification we constrain some justices to be strictly positive or negative.[10] Finally, as in Martin and Quinn (2002) we find that Douglas's extremely limited and liberal voting pattern leads to near separation so we tighten the innovation variance to 0.0001, slightly tighter than the 0.001 imposed by Martin and Quinn (2002). In the robust model we set the scale parameter in the Student's t-distribution so that the overall innovation variance is 0.1 given a Student's t-distribution with 4 degrees of freedom.[11]

We evaluate model fit through posterior predictive checks (Gelman et al. 2014; Gelman and Hill 2007). We sample from the posterior distribution of the parameter estimates generated from each model and use these values to make predictions about the votes cast by justices in each of the cases heard during their tenure. For every draw, we compute the proportion of justice decisions correctly predicted by each model. The robust model accurately predicts Supreme Court decisions 72.95% of the time, while the dynamic model generates accurate estimates 72.77% of the time. Although this is a substantively small difference, it is probabilistically distinguishable from 0. The fact that the robust model's improvement is modest should be unsurprising, as the comparative advantage of this model pertains to the

[10]We constrain Marshall and Brennan to be strictly negative (liberal) and Rehnquist, Scalia, and Thomas to be strictly positive (conservative).

[11]The scale parameter in the Student's t-distribution is similar but not equivalent to the variance parameter in the normal distribution. For the normal distribution and the three parameter Student's t-distribution to have the same variance, the scale parameter of the Student's t distribution is set to $\frac{\nu-2}{\nu}\sigma_{\text{norm}}$ (ν is the degrees of freedom and σ_{norm} is the standard deviation of the normal distribution). For the majority of justices we set this to 0.22 (for Douglass this is set to 0.007).

relatively small subset of instances where a Supreme Court justices' voting position suddenly changes between years. Nevertheless, the improvement in predictive power is substantively meaningful for the subset of cases that seem to experience rapid fluctuations.

We estimate WAIC as a second means of comparing the models. WAIC is a fully bayesian alternative to DIC and is asymptotically equal to Bayesian Leave-One-Out Cross-Validation. WAIC estimates the expected log pointwise predictive density[12] by estimating the log pointwise predictive density (LPPD)[13] and correcting it by the number of effective parameters (Furr 2017; Vehtari, Gelman and Gabry 2016). When estimating the expected model performance on new data, it is necessary to determine the form that hypothetical out-of-sample data will take (Gelman et al. 2014). In this context, we evaluate out-of-sample performance in terms of new realizations of a particular justice's decision on existing court cases. The LPPD is therefore calculated using the likelihood of a judicial decision, given each judge's latent preference. The effective numbers of parameters correction is estimated using the summation of the variance of the log prediction density for each vote.[14]

The difference in the WAIC scores provides evidence that the robust model is outperforming the original dynamic model. The estimate of WAIC for the dynamic replication is 46,852, reported on the deviance scale. The robust model improves on this, with a WAIC of 46,647. This is a difference of 205 with a standard error of 15.6 – the robust model better fits the data than does the dynamic model.

---

[12]This is $\sum_i^N \mathrm{E}_f\big(\log p_{\text{post}}(\tilde{y}_i)\big)$ where E_f is the expectation over the distribution of data from the true data generating process and $\tilde{y}_i$ is a single new data point. The true data generating process is unknown, and so f must be estimated, but this leads to the model appearing to fit the data better than it does.

[13]LPPD can be calculated by using draws from the posterior distribution: $\sum_i^n \log\big(\frac{1}{S}\sum_{s=1}^S p(y_i|\theta^s)\big)$ where θ^s is a single draw from the posterior distribution of θ.

[14]We use the summation of variance correction p_{WAIC2} calculated as $\sum_{i=1}^n \mathrm{V}_{s=1}^S(\log p(y_i|\theta^s))$ where $\mathrm{V}_{s=1}^S$ is the variance of S draws from the posterior (Gelman et al. 2014, p 173).

To demonstrate the substantive value of the robust model, we conduct a brief investigation of estimates for a subset of Supreme Court justices using visual evidence in Figure 3.We include (in the shaded area) the output from the dynamic model estimated by Martin and Quinn (2002). One important contribution generated from the original model was confirmatory evidence that Supreme Court justice preferences changed meaningfully over time (Martin and Quinn 2002, 2007). Our results corroborate this finding, and for most justices there is little difference in the dynamic and robust model estimates.

Yet, the smoothing pattern imposed by the dynamic modeling structure employed by the original authors also precludes the possibility that justice preferences may change rapidly. That a justice may suddenly become more liberal or conservative in their decision-making may seem unlikely. Nevertheless, we find evidence of such changes in a small, but substantively interesting set of cases. A brief examination suggests these rapid fluctuations are not simply a modeling artifact. For example, William Rehnquist's score suddenly becomes more liberal in 1987. This was the first full year in which Rehnquist served as Chief Justice and it is plausible that this change to the composition of the court led to a commensurate change in Rehnquist's voting patterns. As an additional validity check, we also consulted a criterion variable generated by Epstein et al. (1996) that records a justice's voting pattern on a similar liberal-conservative spectrum. These authors also identify a rapid change in Rehnquist's voting pattern, which indicates that this shift reflected real-world changes and is not simply an artifact of our modeling assumptions.

Though we do not offer causal interpretation or explanation for these changes, investigating rapid shifts in voting patterns is a promising area for future research. For example, strategic explanations for Supreme Court justice behavior suggest that voting patterns are a function of both a justice's underlying political ideology and the anticipated actions and preferences of their colleagues and external political entities (Epstein and Knight 2013). Scholars adhering to this framework might argue that Rehnquist changed voting patterns after becoming chief justice were a result of strategic incentives. Prior to becoming Chief

Justice, Rehnquist was known for consistently voting with Chief Justice Burger, in part to maintain their positive collegial relationship (Woodward and Armstrong 1979). As Chief Justice, however, Rehnquist selected the majority opinion writer for all cases in which he was also in the majority. This provides him with an avenue for substantively impacting court decisions, even in cases where he may have otherwise been inclined to vote with the minority. We are agnostic to the validity of these potential explanations. Nonetheless, we find it heartening that the new patterns uncovered by the robust model appear to corroborate some existing arguments. In short, the robust model estimates help bring evidence about such cases to the forefront by accommodating rapid changes – an insight that would remain unobserved due to the smoothing induced by standard dynamic modeling procedures.

Pemstein, Meserve and Melton (2010)

Pemstein, Meserve and Melton (2010) generate country-year estimates of democracy using a static latent variable model that uses ten indicators of democracy. These items include: Polity (Marshall, Jaggers and Gurr 2006); Freedom House (2007); polyarchy (Coppedge and Reinicke 1991); Bollen (2001); Vanhanen (2003); Arat (1991); Political Regime Change (PRC) (Gasiorowski 1996); Bowman, Lehoucq and Mahoney (2005) (BLM); and Przeworski et al. (2000) (PACL). Each of these items are assumed to be a unique approximation of a given country's latent level of democracy. While the items are measured on different scales, the authors treat each as ordinal when linking the items to the latent level of democracy through an ordinal probit function. The model is identified through the application of a standard normal prior distribution on the latent level of democracy and uninformative, uniform priors applied to the cut-point parameters for each item. We depart from the original model slightly by linking the items to the latent trait using the ordered logit, rather than ordered probit function. We also use item-specific slope parameters, instead of item-specific variances, to account for the differences in precision of each item. These modifications do not lead to substantial changes in model estimates – the correlation between our replicated estimates of democracy and the original estimates is 0.987. In addition, we also estimate dynamic and robust variants of the original model.

As with the Martin and Quinn (2002) replication, we also calculate WAIC for each of the three models. The WAIC is 93,267, 79,237, and 65,082 for the static, dynamic, and robust models respectively. The difference between the static model and robust model is 28,185 with a standard error of 358.8. The difference between the dynamic model and the robust model is 14,029 with a standard error of 247.7. In terms of predictive accuracy, the robust model is a substantial improvement upon both alternatives.

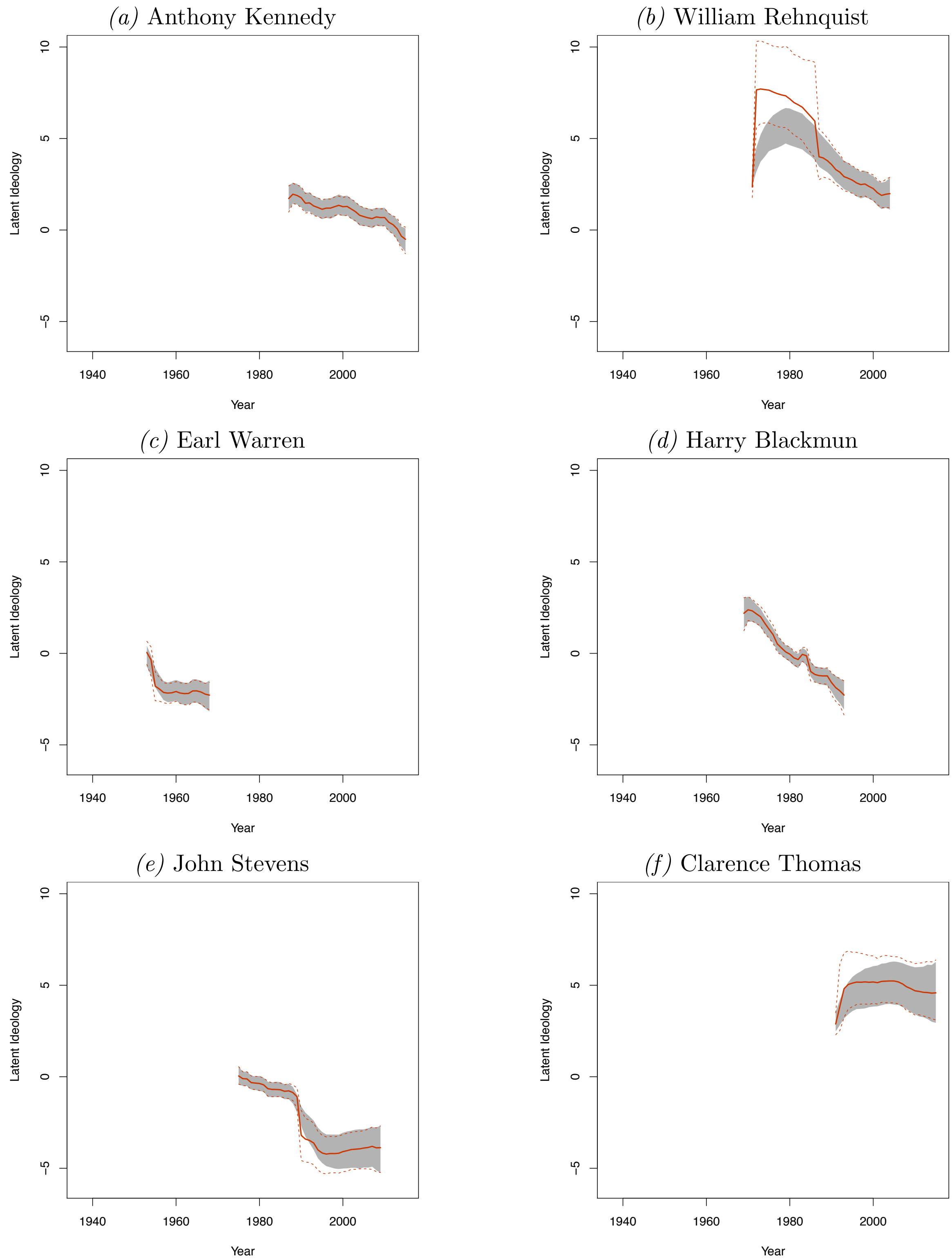
Figure 4 reports the posterior predictive checks used to assess model performance. With some exception, we find that each of the models produce less accurate predictions for indicators with many categories, and more accurate predictions for indicators with few categories. The difference in model performance is greatest for the polity indicator – here the static model correctly predicts polity scores for 41 percent of observations, the dynamic model does so for 61 percent, and the robust model does so for 90 percent. The static model does, however, fair best when predicting Polyarchy, Bollen, Arat, and BLM, though the differences in accuracy are less stark for these variables. The models produce similar accuracy figures for the remaining indicators (Freedom House, Vanhanen, PRC, and PACL). In Appendix F, we evaluate how the models perform when predicting annual change for seven democracy indicators. The robust model performs as well or better for every indicator. It also continues to outperform the static model for Polity.

A brief examination of historical cases that experience shocks–displayed in Figure 5–offers additional support for the robust model. While the dynamic model produces narrower credible intervals than the static model, but over-smooths and anticipates the sudden changes to the level of democracy in these two countries. For the Philippines, the robust model, identifies a steep decrease in the level of democracy in 1972 when Ferdinand Marcos declares martial law and the onset of an autocratic regime. The robust model also estimates a notable increase after 1981, when marital law ended, and another dramatic increase in 1987 with the revolution that expelled Marcos from power. The robust dynamic model is also able to capture sudden changes for the case of Afghanistan. The model even captures the Saur Revolution, which occurred in 1978 and was quickly followed by the Soviet invasion in 1979. This sudden and short-lived institutional change is not observable in any of the other models which again highlights the pitfalls associated with the over-smoothing produced by the standard dynamic approach and constitutes an important validity check.

Discussion

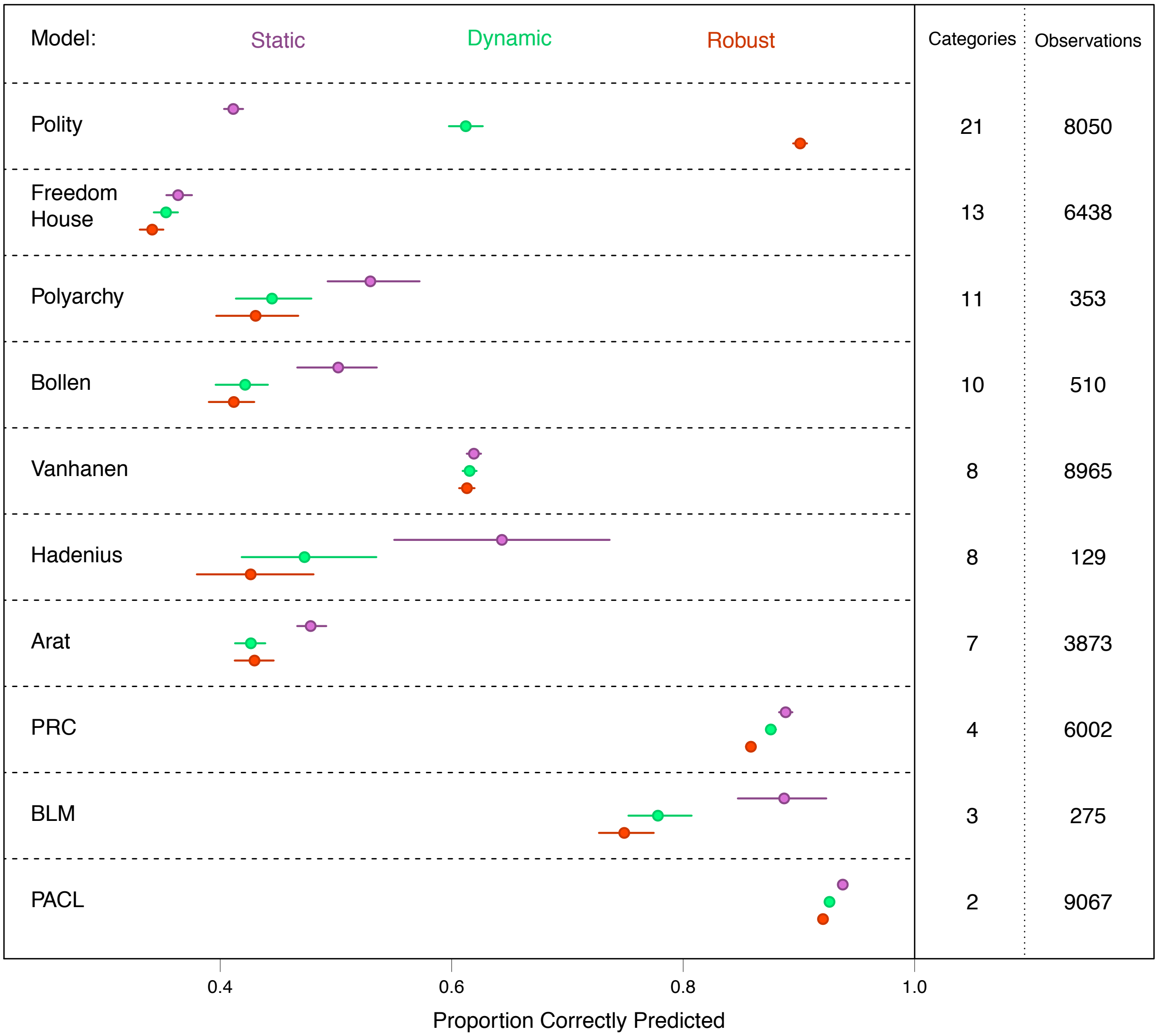
Through simulated data and the reproduction of two existing latent variable models, we have explored how a relatively simple modification to the dynamic latent variable model increases the validity of the estimates of these variables when temporal shocks are present. Across a variety of metrics, the robust dynamic model outperforms existing modeling techniques when the latent traits are subject to volatility. Nevertheless, there is no guarantee that any single modeling strategy will be equally well suited for use with all data types or for estimating different latent concepts. The assumptions of the measurement model will influence the conclusions researchers draw both about the underlying theoretical concept of interest, as well as the empirical linkages between these concepts and other phenomena. We therefore suggest researchers estimate both dynamic and robust models and then assess the relative validity of the latent estimates with as many different evaluation tools as possible. Fit statistics, posterior predictive checks, and visual analysis of the temporal patterns of several well-known cases allowed us to evaluate the competing models without relying on a single statistical tool. Overall, the robust latent variable model reduces the bias associated with dynamic models without sacrificing efficiency, which has lead to new insights about the dynamic patterns of judicial ideology and democracy.

Figure: 3: Sample of Judicial Ideology - Robust and Martin and Quinn Scores



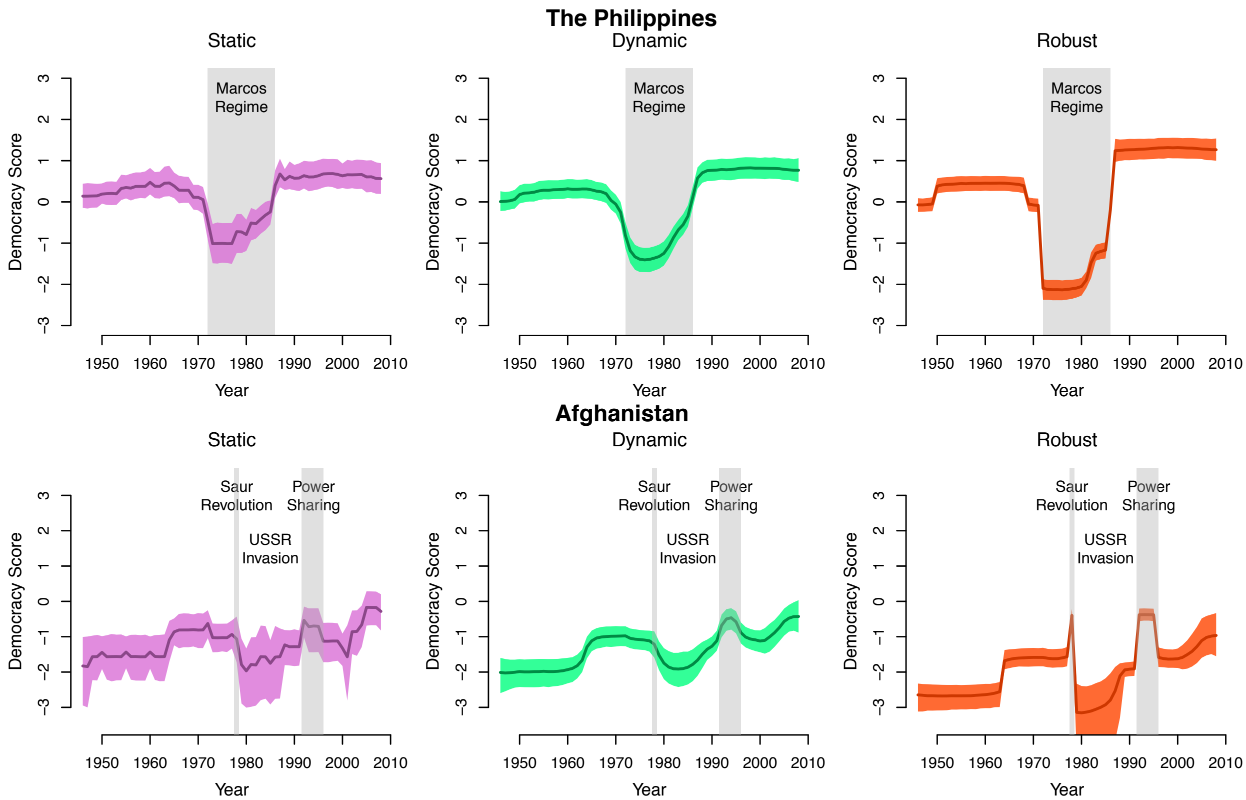
Note: Two sets of judicial ideology scores based on data from Martin and Quinn (2002). These models are estimated using two different versions of the latent variable model (robust and dynamic). The robust model is represented by the solid black line for the median value and the dashed lines for the 2.5 and 97.5 credible interval. The dynamic model is represented by the shaded area, which is the 2.5 to 97.5 credible interval.

Figure 4: Posterior Predictive Checks – Pemstein, Meserve and Melton (2010) Replication



Note: This plot displays each models performance in accurately predicting the values of the ten democracy indicators used to generate model estimates. 500 sets of predictions were randomly drawn from the posterior distributions produced by each model. The horizontal axis reports the percent of cases correctly predicted. Models are displayed along the vertical axis. Dots correspond to the median value from the set of predictions, while solid lines denote the 2.5 and 97.5 percentile values of each models performance. The static model estimates are colored purple, the dynamic model is colored green, and the robust model orange.

Figure: 5: The Philippines and Afghanistan, Latent Democracy Score – Pemstein, Meserve and Melton (2010) Replication



Note: Estimates of latent democracy level are displayed for the Philippines (first row) and Afghanistan (second row). Mean estimates and 95 percent credible intervals are indicated with a solid line and shaded region. For the case of the Philippines, the robust dynamic model is able to clearly distinguish between the periods before, during, and after the Marcos regime. For the case of Afghanistan, the robust dynamic model is able to clearly capture the short-lived Saur Revolution in Afghanistan that occurred in 1978 and was quickly followed by the Soviet invasion in 1979.

References

Arat, Zehra F. 1991. *Democracy and Human Rights in Developing Countries*. Lynne Reinner.

Armstrong II, David A., Ryan Bakker, Royce Carroll, Christopher Hare, Keith T. Poole and Howard Rosenthal. 2014. *Analyzing Spatial Models of Choice and Judgement with R*. CRC Press.

Baker, Andy, Anand E Sokhey, Barry Ames and Lucio R Renno. 2016. "The Dynamics of Partisan Identification When Party Brands Change: The Case of the Workers Party in Brazil." *The Journal of Politics* 78(1):197–213.

Barbera, Pablo. 2015. "Birds of the Same Feather Tweet Together. Bayesian Ideal Point Estimation Using Twitter Data." *Political Analysis* 23(1):76–91.

Blaydes, Lisa and Drew A. Linzer. 2008. "The Political Economy of Women's Support for Fundamentalist Islam." *World Politics* 60(July):576–609.

Bollen, Kenneth A. 2001. *Cross-National Indicators of Liberal Democracy, 1950-1990*. 2nd ICPSR version ed. Inter-university Consortium for Political and Social Research.

Bowman, Kirk, Fabrice Lehoucq and James Mahoney. 2005. "Measuring Political Democracy: Case Expertise, Data Adequacy, and Central America." *Comparative Political Studies* 38:939–70.

Carpenter, Bob, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A Brubaker, Jiqiang Guo, Peter Li and Allen Riddell. 2016. "Stan: A probabilistic programming language." *Journal of Statistical Software* 20.

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

Coppedge, Michael and Wolfgang H. Reinicke. 1991. "Measuring Polyarchy". In *On Measuring Democracy: Its Consequences and Concomitants*, ed. Alex Inkeles. New Brunswick pp. 47–68.

Duncan, Terry E and Susan C Duncan. 2004. "An introduction to latent growth curve modeling." *Behavior therapy* 35(2):333–363.

Epstein, Lee and Jack Knight. 2013. "Reconsidering Judicial Preferences." *Annual Review of Political Science* 16:11–31.

Epstein, Lee, Jeffrey A. Segal, Harold J. Spaeth and Thomas G. Walker. 1996. *The Supreme Court Compendium: Data, Decisions, and Developments.* 2 ed. Congressional Quarterly Inc.

Fariss, Christopher J. 2014. "Respect for Human Rights has Improved Over Time: Modeling the Changing Standard of Accountability." *American Political Science Review* 108(2):297–318.

Fariss, Christopher J. 2018. "The Changing Standard of Accountability and the Positive Relationship between Human Rights Treaty Ratification and Compliance." *British Journal of Political Science* 48(1):239–272.

Fonseca, Thaís CO, Marco AR Ferreira and Helio S Migon. 2008. "Objective Bayesian analysis for the Student-t regression model." *Biometrika* 95(2):325–333.

Freedom House. 2007. "Freedom in the World." *http://www.freedomhouse.org* .

Furr, Daniel C. 2017. Bayesian and frequentist cross-validation methods for explanatory item response models PhD thesis University of California, Berkeley.

Gasiorowski, Mark J. 1996. "An Overview of the Political Regime Change Data Set." *Comparative Political Studies* 29:469–83.

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

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari and Donald B. Rubin. 2014. *Bayesian Data Analysis*. 3 ed. CRC Press.

Geweke, John. 1993. "Bayesian treatment of the independent student-t linear model." *Journal of Applied Econometrics* 8(S1):S19–S40.

Grief, Avner and David D. Laitin. 2004. "A Theory of Endogenous Institutional Change." *American Political Science Review* 98(4):633–650.

Hamilton, James D. 2010. Regime switching models. In *Macroeconometrics and time series analysis*. Springer pp. 202–209.

Hollyer, James R., B. Peter Rosendorff and James Raymond Vreeland. 2014. "Measuring Transparency." *Political Analysis* 22:413–434.

Imai, Kouske, James Lo and Jonathan Olmsted. 2016. "Fast Estimation of Ideal Points with Massive Data." *American Political Science Review* 110(4):631–656.

Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences*. Wiley.

Jesse, Stephen A. 2017. "Don't Know Responses, Personality and the Measurement of Political Knowledge." *Political Science Research and Methods* 5(4):711–731.

Joseph, Lawrence, David B Wolfson, Roxane Du Berger and Roseann M Lyle. 1997. "Analysis of panel data with change-points." *Statistica Sinica* pp. 687–703.

Kōnig, Thomas, Mortiz Marbach and Mortiz Osnabrügge. 2013. "Estimating Party Positions Across Countries and Time - A Dynamic Latent Variable Model for Manifestos Data." *Political Analysis* 21(4):468–491.

Lange, Kenneth and Janet S. Sinsheimer. 1993. "Normal/Independent Distributions and Their Applications in Robust Regression." *Journal of Computational and Graphical Statistics* 2(2):175–198.

Lange, Kenneth L., Roderick J. A. Little and Jeremy M. G. Taylor. 1989. "Robust Statistical Modeling Using the t Distribution." *Journal of the American Statistical Association* 408(84):881–896.

Leventoğlu, Bahar and Branislav L. Slantchev. 2007. "The Armed Peace: A Punctuated Equilibrium Theory of War." *American Journal of Political Science* 51(4):755–771.

Li, Longhai, Shi Qiu, Bei Zhang and Cindy X Feng. 2016. "Approximating cross-validatory predictive evaluation in Bayesian latent variable models with integrated IS and WAIC." *Statistics and Computing* 26(4):881–897.

Linzer, Drew and Jefferey K. Staton. 2016. "A Global Measure of Judicial Independence, 1948–2012." *Journal of Law and Courts* 3(2):223–256.

Marshall, Monty G., Keith Jaggers and Ted Robert Gurr. 2006. "Polity IV: Political Regime Characteristics and Transitions, 1800-2004." *http://www.cidcm.umd.edu/polity/* .

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):134–53.

Martin, Andrew D. and Kevin M. Quinn. 2007. "Assessing Preference Change of the US Supreme Court." *Journal of Law, Economics and Organization* 23(2):365–385.

Pan, Jennifer and Yiqing Xu. 2018. "China's Ideological Spectrum." *Journal of Politics* 80(1):254–273.

Pang, Xun, Barry Friedman, Andrew D. Martin and Kevin M. Quinn. 2012. "Endogenous Jurisprudential Regimes. Political Analysis." *Political Analysis* 20(4):417–436.

Pemstein, Daniel, Stephen A. Meserve and James Melton. 2010. "Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type." *Political Analysis* 18(4):426–449.

Pérez, Efrén O. 2011. "The Origins and Implications of Language Effects in Multilingual Surveys: A MIMIC Approach with Application to Latino Political Attitudes." *Political Analysis* 19:434–454.

Przeworski, Adam, Michael Alvarez, José Cheibub and Fernando Limongi. 2000. *Democracy and Development: Political Regimes and Economic Well-Being in the World, 1950-1990*. Cambridge: Cambridge University Press.

Reuning, Kevin, Michael R. Kenwick and Christopher J. Fariss. 2018. "Replication Data for: Exploring the Dynamics of Latent Variable Models." *https://doi.org/10.7910/DVN/SSLCFF* .

Rosa, GJM, D Gianola and CR Padovani. 2004. "Bayesian longitudinal data analysis with mixed models and thick-tailed distributions using MCMC." *Journal of Applied Statistics* 31(7):855–873.

Santifort, Charlinda, Todd Sandler and Patrick T Brandt. 2013. "Terrorist attack and target diversity: Changepoints and their drivers." *Journal of Peace Research* 50(1):75–90.

Schnakenberg, Keith E. and Christopher J. Fariss. 2014. "Dynamic Patterns of Human Rights Practices." *Political Science and Research Methods* 2(1):1–31.

Spirling, Arthur. 2007. "Bayesian approaches for limited dependent variable change point problems." *Political Analysis* 15(4):387–405.

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

Stegmueller, Daniel. 2013. “Modeling Dynamic Preferences: A Bayesian Robust Dynamic Latent Ordered Probit Model.” *Political Analysis* 21:314–333.

Tahk, Alexander M. 2015. “A continuous-time, latent-variable model of time series data.” *Political Analysis* 23(2):278–298.

Treier, Shawn and D. Sunshine Hillygus. 2009. “The Nature of Political Ideology in the Contemporary Electorate.” *Public Opinion Quarterly* 73(4):679–703.

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

Vanhanen, Tatu. 2003. *Democratization: A Comparative Analysis of 170 Countries*. New York: Routledge.

Vehtari, Aki, Andrew Gelman and Jonah Gabry. 2016. “Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC.” *Statistics and Computing* pp. 1–20.

Voeten, Erik. 2000. “Clashes in the Assembly.” *International Organization* 54(2):185–215.

Western, Bruce and Meredith Kleykamp. 2004. “A Bayesian change point model for historical time series analysis.” *Political Analysis* 12(4):354–374.

Woodward, Bob and Scott Armstrong. 1979. *The Brethren: Inside the Supreme Court*. Simon and Schuster.

Zhang, Zhiyong, Keke Lai, Zhenqiu Lu and Xin Tong. 2013. “Bayesian inference and application of robust growth curve models using student’s t distribution.” *Structural Equation Modeling: A Multidisciplinary Journal* 20(1):47–78.