Generalized Ideal Point Models for Time-Varying and Missing-Data Inference

Robert Kubinec

July 11, 2019

Abstract

This paper presents an item-response theory parameterization of ideal points that unifies existing approaches to ideal point models while also extending them. For time-varying inference, the model permits ideal points to vary in a random walk, in a stationary autoregressive process, or in a semi-parametric Gaussian process. For missing data, the model implements a two-stage selection adjustment to account for non-ignorable missingness. In addition, the ideal point model is extended to handle new distributions, including continuous, positive-continuous and ordinal data. To enable modeling of datasets with mixed data (discrete and continuous), I incorporate joint modeling of different distributions. Finally, I also address ways of implementing Bayesian inference with big data sets, including variational inference and within-chain MCMC parallelization.

While ideal point models were conceived as a way to measure ideology in the U.S. Congress, this powerful measurement tool has wide applicability as a general tool for analyzing social situations. Political scientists have continued to apply the ideal point model to new datasets and empirical challenges, such as Twitter data, the United Nations security council, campaign finance data and legislative speech. Through these applications, it has become apparent that the ideal point model has utility for many more measurement problems in the social sciences where data rarely exactly capture the concept of interest. To encourage further usage of this model, and to help address difficult measurement problems in political science, this paper extends the existing ideal point model while also unifying recent advances within a single framework.

The aim of generalizing the ideal point model is to construct a tool that can be applied to most distributions that political scientists encounter while incorporating non-ignorable missingness and permitting over-time inference. Although recent developments in the field have increasingly employed diverse outcomes, the standard ideal point framework assumes binary data in a largely static framework. As I show in this paper, there is nothing about the underlying ideal point model that requires a binary outcome, and there is much to be learned from the dynamics of ideal points. As the datasets available to political scientists have grown in size and complexity, capturing measures of interest, especially time-varying measures, provides an avenue for powerful inference of compelling research questions.

Broadly speaking, I propose that the ideal point framework can be best understood as a social choice aggregation method, of which the most common application of estimating political ideology is a special case. The underlying idea of the ideal point model, that individuals make choices over competing alternatives by comparing their distance to the individual's ideal point, applies to many social situations in which people make decisions with regard to a polarizing latent dimension. Given the wide applicability of this simple rational actor model, it can help political scientists incorporate measurement error by directly estimating traits that are impossible to capture directly.

This feature of ideal point models—incorporating measurement uncertainty—could be a considerable boon for empirical research if applied more widely. As I show in this paper, the

ideal point model can be understood as a generalization of the item-response theory framework long employed by psychometricians to study latent psychological traits. The ideal point model makes fewer assumptions than the standard item-response theory model about how observed indicators are related to an underlying latent construct, permitting more flexible estimation of latent constructs while also providing useful inferences about exactly how some indicators are related to the construct. The ideal point model not only permits the estimation of unobserved traits, but also to help researchers learn about the structure of these traits.

The contribution of this paper is to combine existing approaches to ideal points and extensions within a single framework. All the models in this paper can be fit with the R package `idealstan`. Combining missing-data inference, time-varying inference and inference over joint distributions provides scholars with a multi-purpose tool for tackling measurement challenges that can extend the reach and applicability of the ideal point model.

In this paper I first situate the ideal point model in the broader context of latent variable models and dimension-reduction methods. I then turn to each generalization of the ideal point model presented in this paper, first missing-data inference via selection models, then time-varying inference with parametric and semi-parametric methods, and finally the ability to combine distributions to estimate indices from survey and panel data. To demonstrate the wide possibilities for new inferences from this framework, I look at the association between twenty years' of local monthly unemployment data and rollcall voting in the U.S. House.

1 Latent Variables as Statistical Deduction

“There is nothing more deceptive than an obvious fact.”

– Sherlock Holmes

The traditional presentation of applied statistics in political science, borrowing heavily from econometrics, presents models as estimating un-biased quantities given a recipe list of datasets. If an outcome is a discrete variable that can take on any number of values, then

the standard advice is to use a Poisson distribution. Of course, the datasets that political scientists produce rarely fit these clean molds, which leads to numerous statistical fixes to correct for poorly-fitting models, such as zero-inflated Poisson models in the case of counts and other more generic fixes like clustered standard errors. What all of these approaches have in common is a focus on adjusting for measurement issues in the dependent variable, while the data are assumed to represent known facts.

Measurement models flip this perspective on its head. Rather than assuming that only the outcome needs a statistical distribution, a measurement model implements a statistical distribution *for the data*. Making this mental switch enables us to think of the data as only one set of observed indicators generated by a latent, unseen construct. In other words, the data are measured *with error*, where the exact errors are defined by a particular statistical distribution.

The simplest measurement model is the econometric errors-in-variables model where the data were generated by a Normal distribution with a pre-specified standard deviation (Durbin 1954). This model assumes that any noise in the data is evenly distributed across all the data points, and its only effect on the final statistical estimation is to increase the uncertainty of estimates. There are many more measurement models available, however, all of which can be combined with traditional statistical estimation techniques or employed as the primary model of interest.

It is possible to define measurement models in terms of the structure they impose on the distribution underlying the data. As mentioned previously, the errors-in-variables model may be the simplest such distribution in which relatively little is known or assumed about noise in the data. Most other measurement models assume that the data are generated by a reduced-form expression. On one end of the spectrum, in cases where there is little prior information about what underlies the data, reduction methods like factor analysis (Harman 1960), principal components analysis (Hotelling 1933) and correspondence analysis (Greenacre 2017; Barberá 2015) can compress the columns of a data matrix into an eigenvalue-based summary of the variation within the matrix. This kind of information can be thought of a generalization of a correlation matrix of the data, describing how different

variables are related to each other purely through covariation.

On the other end of the spectrum are methods that impose a very specific structure on the data distribution. Structural equation modeling is arguably the most demanding as researchers can stipulate a very precise process through which latent variables are related to each other, permitting complex unobserved feedback networks (Ullman and Bentler 2012). There are other models involving latent variable analysis in time-series that impose a very specific functional form, such as the hidden Markov model (Rabiner 1989) and the state-space model (Roesser 1975), in which the data were generated by an unobserved time-varying construct. The number of latent variable models with a precise structure can be almost endless as it can even include two-stage models like the Heckman selection model (Heckman and Sedlacek 1985).

The ideal point model, and its close kin, the item-response theory framework, fit somewhere in the middle of this spectrum. These approaches put some structure on the latent trait underlying the data by assuming that the data represent a social choice process. By observing the data, we can learn something about people’s latent abilities, in item-response theory terms, or people’s latent ideal points in the political science terminology. Formally, an ideal point model for any person i and any vote j is defined as the difference in utilities between the person’s ideal point α_i and the vote’s Yes position, β_{jY} and No position, β_{jN} , such that i always votes for the bill if:

$$\sqrt{(\alpha_i - \beta_{jY})^2} > \sqrt{(\alpha_i - \beta_{jN})^2}$$

That is, if α_i is closer in Euclidean space to the Yes position than the No position. While Enelow and Hinich (1984) originally defined this model in deterministic terms, most statistical applications add in vote-specific error terms to make the decision stochastic. The two main statistical parameterizations of this model, the DW-NOMINATE family (Poole et al. 2008) and the Bayesian IRT family (Clinton, Jackman, and Rivers 2004), differ primarily in the distribution they assign to that error term (Carroll et al. 2009). While an argument can be made for the preference of one over the other (Carroll et al. 2013), or even to use a fully non-parametric estimator (Tahk 2018), I concentrate on the IRT formulation due to its ease

of implementation in a Bayesian framework and its connection to the IRT literature.

Clinton, Jackman, and Rivers (2004) showed that a certain IRT model, the 2-Pl family, was equivalent to the ideal point model if the actual Yes and No positions of the bill were left unidentified. Instead, this parameterization can estimate the ideal points α_i and a midpoint where the person i is indifferent to voting on the bill j . The IRT 2-Pl model in terms of ideal points α_i , or what are called ability parameters in the IRT literature, discrimination parameters γ_j and difficulty parameters β_j , is as follows for a binary observed vote Y_{ij} :

$$Pr(Y_{ij} = 1) = \prod_{i=1}^I \prod_{j=1}^J \text{logit}^{-1}(\gamma_j \alpha_i - \beta_j) \quad (1)$$

where the midpoint or line of indifference for each vote can be found by re-arranging the linear model so that

$$\alpha_i = \frac{\gamma_j}{\beta_j}$$

In addition to estimating the line of indifference, the bill or item parameters (in IRT speak) have their own meaning. The sign of the discrimination parameter γ_j indicates the “polarity” of the bill/item, or whether voting yes indicates one is more liberal or conservative in the Congressional case. The absolute magnitude of the discrimination parameter indicates how well the bill/item discriminates between persons such that the latent dimension strongly predicts voting. The β_j difficulty parameters, by contrast, have the relatively uninteresting interpretation in representing the propensity of a person to vote yes marginal or independent of the latent dimension, though their incorporation is important to allow for the residual propensity to vote Yes to vary by bill.

The crucial difference between this parameterization and the standard IRT model is that the discrimination parameters γ_j are unconstrained. In an IRT 2-Pl model the discrimination parameters are normally constrained to be positive so that the ideal points/ability parameters α_i can have the interpretation of reflecting the ability of a students to answer correctly more difficult test questions. In other words, the observed indicators are always positively related to the latent trait. The ideal point model can be considered a generalization of the IRT 2-Pl

model because it does not assume that this is the case and in fact can estimate the direction of items from the data. This difference, though it would seem minute, is important when analysts are uncertain of the directional relationship between the data and the latent trait, as often occurs in practice.

This flexibility does, though, create some additional complications. The always-positive constraint in a standard IRT model ensures identification in terms of the rotation of the ideal points, so abandoning that assumption requires more work to identify the model. As has become well-known, constraints must be built in to the IRT ideal point estimation in order to identify a unique rotation of the ideal points (Bafumi et al. 2005). Usually the polarity of a subset of ideal points or bills/items is constrained so that persons with high or low scores on the latent variable will map onto positive or negative values in the estimated scale. Then the model can learn the polarity of unconstrained ideal points/items given these a priori constraints. While identification is often considered a hassle, it is in fact a feature, not a bug, as the constraints permit inference on the unconstrained items.

It is important to note that there is nothing about this model that necessarily implies that the latent scale is political ideology. That is a useful interpretation to assign ideal points based on data from the U.S. Congress (Poole and Rosenthal 2007), although even in that domain the interpretability of ideal points in that sense has been recently questioned (Krehbiel and Peskowitz 2014). The ideal point model represents a social choice process where the construct by which people make choices is unobserved. As such, it *can* estimate ideology as a latent construct, but it does not need to. Interpreting what the latent variable represents is a necessary part of any application of the model. What can be said is that it always represents the aggregation of a social choice process with competing alternatives.

In this paper I employ a Bayesian interpretation of this model (Clinton, Jackman, and Rivers 2004) that I estimate using Hamiltonian Markov Chain Monte Carlo with Stan (Carpenter et al. 2017). I am interested in posterior inference on the set of unobserved parameters $\{\alpha_i, \gamma_j, \beta_j\} \in \theta$ that is conditional on the observed data Y_{ij} , or $Pr(\theta|Y_{ij})$. To do so, I define independent priors on θ , $Pr(\theta)$, which is multiplied by the likelihood $L(\cdot)$ that is defined by (1):

$$Pr(\theta|Y_{ij}) \propto Pr(\theta)L(Y_{ij}|\theta) \quad (2)$$

I use the proportional symbol $\propto$ to reflect the fact that the term in the denominator representing the probability of the data $Pr(Y_{ij})$ is normally omitted in computational Bayesian analysis as it is fixed with respect to the parameters (Gelman et al. 2013). We can further specify the priors in terms of the distributions assigned to each parameter in θ :

$$\alpha_i \sim N(0, 1)$$

$$\gamma_j \sim N(0, 5)$$

$$\beta_j \sim N(0, 2)$$

The ideal points are given a tight $N(0, 1)$ prior to constrain the scale for identification purposes, while the discrimination and difficulty parameters have weakly informative priors on the logit scale. The effect ϕ of a matrix of covariates X measured on the persons i can also be added to this model for ideal points as shown in Bafumi et al. (2005) and Clinton, Jackman, and Rivers (2004) by including them as hierarchical predictors:

$$\alpha_i \sim N(X\phi', 1)$$

I also consider the possibility of including hierarchical covariates to predict the discrimination of the J items, which has not been estimated previously in the literature though it is straightforward in a Bayesian framework:

$$\gamma_j \sim N(X\phi', 1)$$

Given this definition of the model, I next show how this framework can incorporate missing data when not all of Y_{ij} is observed, time-varying inference when α_i is measured at different points in time and the estimation of diverse distributions for Y_{ij} whether singly or jointly.

2 Missing Data

Missing data is a tricky problem to address in latent variable models because latent variables are themselves defined as missing data. For this reason, standard imputation techniques cannot be easily applied. Because ideal point models are representations of social aggregation processes, a straightforward approach of imputing data conditional on the model assumes that missingness is orthogonal to observed ideal points (Little and Rubin 2002). In many situations, this assumption may not be realistic when there is reason to believe that social actors, such as legislators, are absent for reasons that are related to their own ideal points.

The most rigorous treatment of the problem of missing data and ideal points can be found in Rosas, Shomer, and Haptonstahl (2015), who put forward a model in which non-ignorable missingness occurs when legislators disagree with the position of their party. Their analysis shows that missingness in ideal point models often reflects strategic behavior and it is unwise to treat it as either missing-at-random (MAR) or missing-conditionally-at-random (MCAR). In this section I define a model for non-ignorable missingness that aims for widespread applicability while taking into account the strategic nature of ideal point actors.

I extend my notation of the outcome Y_{ij} by adding a subscript $r \in \{0, 1\}$ for whether person i chose to vote or answer item j ($r = 1$) or chose not to vote/answer ($r = 0$). We can then add a separate selection model that first estimates $Pr(r = 0)$ and deflates the likelihood $L(Y_{ijr}|\theta)$ accordingly:

$$L(Y_{ijr}|\alpha_i, \gamma_j, \beta_j, \nu_j, \omega_j) = \prod_{i=1}^I \prod_{j=1}^J \begin{cases} \zeta(\alpha'_i \nu_j - \omega_j) & \text{if } r = 0, \text{ and} \\ (1 - \zeta(\alpha'_i \nu_j - \omega_j))L(Y_{ijr}|\alpha_i, \gamma_j, \beta_j) & \text{if } r = 1 \end{cases} \quad (3)$$

I let $\zeta(\cdot)$ stand for the inverse logit function in the equation above. As can be seen, the selection model for absences $\zeta(\alpha'_i \nu_j - \omega_j)$ is very similar to (1) except that I have substituted a new set of discrimination ν_j and difficulty ω_j parameters. The ideal points α_i enter into both the selection model and the main ideal point model $L(Y_{ijr}|\alpha_i, \gamma_j, \beta_j)$. As such, the selection model is able to inflate or deflate the ideal points by taking into account a first

stage process. The inclusion of ν_j and ω_j in the first-stage selection model creates a new “missingness” space where person i first chooses whether she will cast a vote, send a tweet or answer a question, as the case may be. Only if the item is close to person i in this missingness space will person i then also decide to participate and provide information Y_{ij} that is informative of their ideal point. This model can be interpreted as a censoring model where persons may choose to self-censor their ideal points.

The general form of the selection model—which is essentially a first-stage ideal point model—allows it to pick up a range of non-ignorable missingness patterns if missingness correlates with ideal points. The model does not make a priori assumptions about exactly why ideal points may determine missingness, and as such the model cannot provide such an interpretation a posteriori without taking time to interpret what the parameters of the selection model reveal about actors’ intentions. The discrimination parameter ν_j is very helpful for this purpose because it will indicate which set of items show correlation between missingness and one of the poles of the latent variable. For example, if the latent variable is constrained positive for liberal Senators, then high positive discrimination would indicate that more liberal Senators tend to be absent on a particular bill. A naive model that assumed that absences were ignorable would miss this pattern and treat these Senators as more moderate based solely on their observed voting record.

This model is also general enough to allow for the possibility that data is actually missing at random at the item level. If the discrimination parameter ν_j is zero then ideal points do not enter into the equation for missingness and $Pr(r = 0)$ is equal to the item-specific intercept ω_j . In the legislative context, this would be the same as estimating the probability of absence by the proportion of legislators who do not show up for particular bill. Including this parameter will also separate that item-specific random missingness from missingness patterns suspiciously associated with ideal points. For this reason, the selection model will perform at least as well as a standard imputation method where the missing ideal points are MCAR conditional on each item.

While this missingness model is readily applicable to a legislative context where legislators may not want to show up on votes depending on what the vote would reveal about their

ideal points, it is also useful for other non-ignorable missingness patterns. Twitter data is an excellent example. It is well-established that estimating a latent trait like political polarization will be vastly over-stated if the tendency of Twitter users to select who they choose to follow is not taken into account (Barberá et al. 2015). The two-stage selection model can be used to estimate the people’s propensity to retweet ideological content net of their self-selection into whom they follow on Twitter (Kubinec and Owen 2018).

3 Time-varying Inference

While ideal points are usually conceived of as relatively time-invariant constructs, time-varying ideal point models have been available to researchers for quite some time. The most well-known time-varying version is the DW-NOMINATE model (Poole and Rosenthal 2007; Armstrong et al. 2014; Caughey and Schickler 2016), which employs a flexible polynomial function of ideal points to permit them to vary semi-parametrically. The other existing approach is the random walk model of Martin and Quinn (2002) in which ideal points in time t are equal to ideal points in $t - 1$ plus a Normally-distributed random jump. This simple model is nonetheless quite helpful at determining the probable trend in ideal points over time because it imposes virtually no structure on the time-series trends.

Both of these approaches share in common an emphasis measuring over-time trends. By contrast, in this paper I also want to consider effects of hierarchical predictors on ideal points that vary over time. Including such predictors would enable scholars to test much more precise questions about how time-varying covariates influence ideal points, especially as political scientists obtain ever-larger and more fine-grained datasets. However, existing methods cannot easily handle such covariates. The DW-NOMINATE approach is designed to produce accurate and relatively non-parametric estimates of legislator’s ideal points versus inference on those trends, while the random-walk approach cannot include additional covariates without radically changing the model. Any covariate with a constant effect in a random-walk will push the time series in a constant direction over time (i.e., static drift). It is too simple of a model to separate the effect of a covariate from the time series trend.

To produce a wider range of possibilities to time-varying inference, I introduce two new time series models as alternatives to the random-walk model. I begin by defining the random walk by placing a subscript t on the ideal points α_i and stipulating a relationship in the prior between time points:

$$\alpha_{it} \sim N(\delta_i + \alpha_{it-1}, \sigma_i) \quad (4)$$

We give every person i a separate over-time variance parameter σ_i and intercept/offset δ_i . It is possible to identify this model by placing a tight upper bound on the variance parameters and adding a polarity constraint in the intercepts δ_i , though as I discuss later, this approach can still fail. For computational reasons, the implementation of this model in `idealstan` employs a non-centered parameterization (Betancourt and Girolami 2013) that reduces dependence between the variance σ_i and the prior value of α_{it-1} :

$$\begin{aligned} \epsilon_{it} &\sim N(0, 1) \\ \sigma_i &\sim E(1) \\ \alpha_{it} &= \delta_i + \alpha_{it-1} + \sigma_i \epsilon_{it} \end{aligned} \quad (5)$$

The first new time-series model that I consider is an AR(1) or autoregressive parameterization, which can be understood as a generalization of the random-walk model. To do so, we must add a parameter to the model, ψ_i that is constrained to lie in the $(-1, 1)$ interval:

$$\begin{aligned} \epsilon_{it} &\sim N(0, 1) \\ \sigma_i &\sim E(1) \\ \alpha_{it} &= \delta_i + \psi_i \alpha_{it-1} + \sigma_i \epsilon_{it} \end{aligned} \quad (6)$$

If ψ_i lies in the $(-1, 1)$ interval, then the series is stationary and will always return or decay to the long-term mean of the series δ_i . By comparison, the random walk model can be thought of as an AR(1) model where ψ_i is fixed at 1. Importantly, the effect of a hierarchical time-varying covariate $X'_t \phi$ is now well-defined due to the addition of the decay factor ψ_i in

the model. A covariate ϕ now represents the shock of a time-varying covariate X_t on the ideal point series α_t that will decay over time at a rate determined by ψ_i . We can even calculate an impulse-response function showing the over-time decay in the effect of ϕ as shown in the empirical example section.

In addition to incorporating covariates, the AR(1) time-series model will provide more useful estimates than the random-walk model when there is a prior reason to believe that ideal point trends are stationary over time. This assumption could hold in a setting where shifting ideal points reflect temporary “shocks” as opposed to long-term shifts, such as comparing roll call votes from a single legislative session versus an entire legislator’s career. A priori it can be difficult to know which model is a more accurate description of time trends in ideal points; however, fitting both models and comparing predictions can elucidate how each model performs in a given context.

Based on this comparison of the AR(1) and the random walk model, it would seem advantageous to have a model that estimates time-series trends with relatively little structure while still allowing for time-varying covariates. To include such a model, I further define a Gaussian process ideal point model in which the ideal points α_{it} vary in terms of a Gaussian process with a squared-exponential kernel. A Gaussian process (GP) is chosen because it represents a very flexible framework for semi-parametric inference of either spatial or temporal autocorrelation (Rasmussen and Williams 2006). In addition, while existing models only incorporate discrete time processes where it is assumed that the time points are equally spaced, a Gaussian process is a “continuous” time process where the time points can be irregularly spaced, such as votes occurring at odd intervals throughout a year. Given the limited application of this model in political science research, let alone ideal point models specifically, I briefly review the notation for a GP before combining it with the ideal point model.

A GP is usually defined as a “distribution over functions” (Rasmussen and Williams 2006, 13). We assume that underlying an observed time series x_t is an unobserved function $f(x_t)$. To estimate the most likely values of $f(x_t)$, I can assume $f(x_t)$ is multivariate Normally-distributed where the mean and covariance of this distribution are themselves functions of

the mean and covariance of $f(x_t)$:

$$f(x_t) \sim N(\mu(f(x_t)), \Sigma(f(x_t))) \quad (7)$$

Usually the $\mu(f(x_t))$ function is assumed to be 0 as it does not vary with t . Instead, much of the flexibility of the distribution involves specifying a function for $\Sigma(f(x_t))$, the covariance. The function most often used, and incorporated here, is the squared-exponential kernel $k_t(\cdot)$:

$$k_t(x_t, x_{t'}) = \sigma_f^2 e^{-\frac{(x_t - x_{t'})^2}{2l^2}} + \sigma_{x_t}^2 \quad (8)$$

What this function does is convert all the distances in terms of time between x_t and $x_{t'}$ into a positive semi-definite covariance matrix Σ . We can then sample from this multivariate Normal with covariance Σ to estimate a distribution over the possible time-series functions $f(x_t)$. The GP framework is so flexible that it can fit an *infinite* number of basis functions of x_t (Rasmussen and Williams 2006, 14). Other semi-parametric functions, such as splines and polynomials, are special cases of the GP for certain values of the hyper parameters used in the covariance function (Rasmussen and Williams 2006, 137–40). What is even more appealing is that the GP can handle time-varying covariates $X_t\phi'$ simply by including them in place of $\mu(x_t)$. The multivariate Normal will then sample from the specified covariate matrix while averaging over the possible values of the covariates.

The GP model has another considerable advantage over the random walk and AR(1) models. Because the covariance function accepts the actual values of the time series, it does not have to use consecutive time points. In an AR(1) or random walk model, it is assumed that the time points are discrete and consecutive. For example, in a legislature it would be possible to record the actual day that a bill was voted on rather than simply assign that aggregate the bill to the month-level. In cases in which gaps in the time series are significant, the GP may produce inferences that are more realistic (Tahk 2015).

What gives the GP the ability to fit so many different functions are the three hyper-parameters in (8). $\sigma_{i_f}^2$ can be referred to as the marginal standard deviation and represents

the total amount of variance in x_t explained by the covariance function $k(x_t, x_{t'})$. A higher value for this hyperparameter will result in more bounce in the time series. $\sigma_{ix_t}^2$, on the other hand, represents residual variance in the time series x_t that the covariance function does not fully explain (which could be interpreted as additional measurement error). Finally, the length-scale parameter l_i^2 represents a smoothing factor controlling how much different in time are correlated together. A very low length-scale will allow the time-series to cross the origin at a much higher rate. Conversely, higher length-scales result in smoother functions. The three hyper-parameters both overlap and interact, which makes isolating the effect of any one hyper-parameter difficult.

To employ the GP as a time process for ideal points, we simply replace the observed x_t with the latent ideal points α_{it} :

$$\alpha_i \sim N(0, k_t(\alpha_{it}, \alpha_{it'})) \quad (9)$$

This straightforward parameterization shows some of the power of Bayesian modeling. We can include virtually any time process by simply defining the time series over a set of parameters rather than a set of observed data.

These different time series models can be understood in terms of order of complexity. The random walk is the simplest as it only requires one parameter, σ_i , per person i . The AR(1) model is more complex as it introduces an additional parameter to control the decay rate in the series, ψ_i . The GP is the most complex as it has three separate parameters per person, σ_{if}^2 , l_i^2 and σ_{iY}^2 .

We can also combine these time-series processes with the model of missing data in (10) to permit inference on ideal points that vary over time and may have non-ignorable missingness:

$$L(Y_{ijtr}|\alpha_{it}, \gamma_j, \beta_j, \nu_j, \omega_j) = \prod_{t=1}^T \prod_{i=1}^I \prod_{j=1}^J \begin{cases} \zeta(\alpha'_{it}\nu_j - \omega_j) & \text{if } r = 0, \text{ and} \\ (1 - \zeta(\alpha'_{it}\nu_j - \omega_j))L(Y_{ijtr}|\alpha_{it}, \gamma_j, \beta_j) & \text{if } r = 1 \end{cases} \quad (10)$$

To change from a random-walk to a different time-series model for α_{it} , we can simply change the prior for α_{it} that we include in $Pr(\theta)$ with one of the previously-defined time processes. I do not index the item parameters by t because in many applications of the model, the item parameters only occur at one point in time, such as bills in a legislature. It is possible in some situations, such as students taking a test multiple times, for item parameters to vary over time as well. Inference on these parameters' time-series trends would then involve defining time-series priors over these parameters, whether jointly with α_{it} or separately, which is a possible area for future research.

Identification for models with time-varying α_{it} is more challenging. Martin and Quinn (2002) addressed this issue for random-walk models by restricting the variance σ_i to very low values, generally less than 0.1. I employ this restriction as well for the random walk and AR(1) models, constraining σ_i to the interval $(0, 0.1)$ for the random walk model and $(0, 0.5)$ for the AR(1) model, although I find that even with these very restrictive variances, unidentifiability can still manifest itself for the random-walk model. Time-varying models can have multiple possible rotations of the ideal points at multiple time points, creating many modes and oscillation from one time point to the next. As such, for the random-walk and GP models, I first fit a model with a restricted variance with variational inference, described later in this paper, and then include the differences between the maximum and minimum of two persons' time series as additional prior information. This prevents further oscillation in the ideal points while allowing for slightly more variance in the series. Indeed, fitting an approximate fit of the model first is helpful for the AR(1) model as well as it provides useful starting values for hierarchical parameters.

As a result, the AR(1) model is in fact easier to identify than the random walk. Because of the constraints on the series to return to its over-time mean, all that is usually needed for identification is a polarity constraint on two of the persons' long-term intercepts δ_i . The AR(1) model can generally also permit more variance in σ_i than the random walk model without resulting in multi-modality.

In addition to the polarity constraints, The GP model requires further constraints to have global identification. Its greatest strength, flexibility, is also its greatest weakness in an

ideal point framework. To prevent multi-modality and oscillation in the series, I had to limit the residual standard deviation parameter σ_{iY} to a very small interval of $(0, 0.025)$ as the inclusion of *additional* measurement error caused oscillation in the time series. I then assigned a very high log-normal prior to the length-scale l^2 equal to three times the mean time point $3\bar{T}$. The marginal standard deviation σ_{if} is constrained to the $(0, 0.3)$ interval, permitting some inference on the variability of different persons' trends. These priors and constraints are conservative and are designed to permit identification in a wide array of settings. The priors and constraints can be loosened with larger datasets that also have very high levels of over-time variation. As the hyper-parameters are constant with respect to T , identification will generally improve with longer time series.

4 Diverse Distributions

Up to this point, I have followed convention in assuming that the observed data Y_{ij} is a binary outcome (Bernoulli distribution). Virtually any distribution for Y_{ij} is possible, especially with the flexibility of Bayesian modeling. In this section I incorporate existing distributions used for ideal points while also introducing the reader to distributions, especially ordinal models, that have been rarely employed for this purpose but could have considerable utility. Furthermore, I implement the Kropko (2013) method to combine distributions into one joint posterior distribution, permitting models to use mixed datasets, such as both discrete and continuous variables. This method is especially useful for building time-varying indices incorporating measurement indicators that could be binary, ordinal, continuous or count variables.

There are two possible ordinal models that can be implemented in the ideal point framework. The first is known as a rating-scale model in the IRT framework and is structurally similar to the ordered logit model. Imai and Kim (2016) are the first to implement this model for ideal points, though within an EM algorithm that is unfortunately vulnerable to issues of perfect separation, which occur frequently in legislative data if certain vote categories, such as abstentions, do not occur very often. By contrast, a fully Bayesian approach to ordinal

models permits very sparse distributions of responses across persons.

Following the notation I introduced, a rating-scale model for $k \in K$ outcomes can be modeled as follows for observed data:

$$L(Y_{ijt}| \alpha_{it}, \gamma_j, \beta_j) = \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^T \begin{cases} 1 - \zeta(\gamma_j \alpha_i - \beta_j - c_1) & \text{if } K = 0 \\ \zeta(\gamma_j \alpha_i - \beta_j - c_{k-1}) - \zeta(\gamma_j \alpha_i - \beta_j - c_k) & \text{if } 0 < k < K, \text{ and} \\ \zeta(\gamma_j \alpha_i - \beta_j - c_{K-1}) - 0 & \text{if } k = K \end{cases}$$

where again $\zeta(\cdot)$ represents the logit function. In this version, each ordinal category k less one is assigned a cutpoint c_k . In ideal point terms, these cutpoints divide the ideal point space across categories. A direct application is to use this model to include abstentions as a middle category between yea and nay votes. In this three outcome setting, two cutting planes exist in the ideal point space such that one end of the spectrum would vote No, the opposite end of the spectrum would vote Yes, and those in between the cutpoints would vote Abstain. As the polarity of items switches, Abstain will always remain in the center of the ideal point distribution, but the Yes and No positions can flip sides. This simple but efficient model helps to model situations where abstentions are in fact quite important, such as parliamentary systems (Bräuninger, Müller, and Stecker 2016).

A different parameterization is known as the graded response model in IRT. This model allows the cutpoints c_k to vary by bill. As a result, instead of including separate cutpoints, we can change the model by indexing each item-specific intercept for each category k :

$$L(Y_{ijt}| \alpha_{it}, \gamma_j, \beta_j) = \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^T \begin{cases} 1 - \zeta(\gamma_j \alpha_i - \beta_{jk-1}) & \text{if } K = 0 \\ \zeta(\gamma_j \alpha_i - \beta_{jk-1}) - \zeta(\gamma_j \alpha_i - \beta_{jk-1}) & \text{if } 0 < k < K, \text{ and} \\ \zeta(\gamma_j \alpha_i - \beta_{jk-1}) - 0 & \text{if } k = K \end{cases}$$

Though not employed in political science to any great extent, this alternative formulation to ordered logit permits more flexible inference on the cutpoints. Allowing cutpoints to

vary by item will enable much more accurate prediction of categories with fewer numbers of observations. It also increases information about the items/bills, though of course at a cost of including K additional parameters per item. My intention in including it here is not to suggest that it is a superior model to the standard rating-scale approach, but rather as an alternative that may be useful in some situations. Either model can be profitably used to model rollcall data with middle outcomes like abstentions.

I do not review the distributions for the Poisson count model as that has been widely applied in political science (Slapin and Proksch 2008). However, I do specify two other distributions that have not been widely used in the ideal point literature: the Normal distribution for continuous and the log-normal distribution for positive-continuous outcomes. While Kropko (2013) employed the Normal distribution in a traditional IRT framework, no one has yet used the log-Normal distribution.

For a normal distribution, the ideal point model is used to predict the mean of Y_{ijt} given residual standard deviation σ_N :

$$L(Y_{ijt}|\alpha_{it}, \gamma_j, \beta_j) = \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma_N^2}} e^{-\frac{(Y_{ijt} - (\gamma_j\alpha_{it} - \beta_j))^2}{2\sigma_N^2}}$$

The log-normal distribution is the same distribution except that the domain of the distribution is on $\log(Y_{ijt})$:

$$L(Y_{ijt}|\alpha_{it}, \gamma_j, \beta_j) = \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^T \frac{1}{\sqrt{2\pi\sigma_N^2}} e^{-\frac{(\log(Y_{ijt}) - (\gamma_j\alpha_{it} - \beta_j))^2}{2\sigma_N^2}}$$

The log-normal parameterization is useful for continuous variables that are only defined on positive real numbers. For variables with substantial skew, such as income, this parameterization has a more useful interpretation than the more general Normal distribution. With both of these distributions, most continuous distributions can be modeled with reasonable accuracy.

The utility of including all of these kinds of distributions, besides granting scholars more options with which data to use for ideal point estimation, is to permit modeling of different

distributions in the same data. It is very common to have panel data sets and survey data with mixed data types, such as binary, ordinal, counts and continuous variables. Other than Kropko (2013) who worked within a traditional IRT framework (i.e., fixed polarity of items), no one has attempted to implement this kind of joint modeling. It promises to offer scholars the ability to easily build time-varying indices such as those employed by V-DEM (Coppedge et al. 2017) and the transparency index (Hollyer, Rosendorff, and Vreeland 2018). In addition, these joint distributions incorporate missing-data inference and time-varying inference defined previously in this paper.

It is straightforward to do so using the Bayesian framework defined in this paper. I simply index Y_{ij} by m for the M models that each relate to a distinct outcome. These outcomes could be binary, ordinal, counts, continuous or positive-continuous per the notation presented:

$$L(Y_{ijtm}|\alpha_{it}, \gamma_j, \beta_j) = \prod_{i=1}^I \prod_{j=1}^J \prod_{t=1}^T \prod_{m=1}^M L_m(\gamma_j \alpha_{it} - \beta_j)$$

where L_m represents the likelihood function for the m th distribution. While I suppressed notation for missing data for the sake of brevity, all of these models can also be estimated marginal of missing data r using the notation presented earlier in addition to incorporating time-series dynamics.

5 Big Data Inference

The Achilles heel of Bayesian inference is generalizing to large data sets. Markov chains, which form the core of the estimation engine, cannot be parallelized as each iteration depends on the value of the previous iteration. As such, I present two methods of scaling this model to the ever-increasing datasets that political scientists employ for analysis.

The first method is variational posterior approximation. A variational approximation converts the posterior distribution into a series of factorized distributions for each variable in the posterior (Grimmer 2011; Blei, Kuculkelbir, and McAuliffe 2017). Though this method is not new, the implementation in Stan that I incorporate as well in `idealstan` is. Variational

algorithmic approximations normally need to be derived separately for each model, which dramatically increases the costs of switching to this form of inference. However, Stan has implemented a form of “black box” variational inference where the gradients necessary to perform variational are calculated using an automatic differentiation library (Kucukelbir et al. 2015). As a result, variational inference is available for *any* of the models presented in this paper (and all future models as well).

Variational inference (VI), though, is not a perfect solution. Because the distribution of each parameter is independent, it ignores inter-parameter correlations, which lowers the uncertainty of estimates. Furthermore, the variational approximation can fail to capture a close summary of the posterior if the algorithm is terminated prematurely (Gelman 2018). This issue can be largely avoided by using a more conservative standard for convergence, and additional statistics can be measured to test for degeneracies in the variational approximation (Gelman 2018). It may even be possible to adjust posterior inference post-hoc by conducting simulation studies on particular models to see how much uncertainty is lost with the VI algorithm.

In addition to these questions about VI’s ability to capture the true posterior, it also has an upper limit on its performance. While replacing MCMC with a much simpler factorized distribution will dramatically lower processing times, for truly enormous datasets it will still slow down. As such, I also examine an emerging method for big-data inference with Bayesian datasets: parallel computation.

Traditionally, parallel computation has had a limited impact on MCMC because Markov chains depend on prior iterations of the algorithm. This dependence between iterations means that the computations cannot be parallelized across iterations. However, recently the Stan team has implemented a way to parallelize the *likelihood* computation within a Markov chain, and in particular the computationally expensive gradients necessary to perform Hamiltonian MCMC. The likelihood can be parallelized so long as parameters are conditionally independent. For a static ideal point model, this means that the likelihood can be parallelized across persons or items. For a dynamic model, parallelization can be performed across persons as ideal points depend on prior values. As this parallelization has

yet to be fully released, it will be implemented in the near future as part of the R package `idealstan`, permitting truly massive computational inference of ideal point models where the upper limit is the number of computer cores available.

There is one another means of speeding up computations, and that is to transfer computation to a graphical processing unit (GPU). This method has recently been implemented successfully to speed up MCMC estimation of a static ideal point model (Lewis and Tausanovitch 2018). Stan has also recently implemented GPU optimization for matrix calculations. This feature will be implemented to speed up GP inference, as GPs require sampling from a multivariate Normal distribution. However, GPU optimization will otherwise be limited as it requires square matrix computation, and the data in `idealstan` is in a “long” form that allows for varying numbers of observations per item and person.

6 Empirical Examples

To show how time-varying estimation can increase our ability to learn about ideal points, I fit a time-varying ideal point model with the most commonly used dataset in ideal point analysis: U.S. Congressional roll-call votes. The research question that I consider is that of the effect of economic disruption on political polarization in Congress. Recent research has proposed that changes in economic conditions, whether in terms of economic inequality and trade, may be affecting increasing polarization in the U.S. Congress. It has long been understood that there is an important interaction in terms of how voters perceive economic recession and their willingness to vote for parties in a polarized environment (Alesina and Rosenthal 1995).

However, we also know that the salience of policy-making varies with economic conditions that correlate with external shocks (Baker, Bloom, and Davis 2016). Most of the studies of trade and other shocks on Congress employ outcomes at the year level, whether the unit of analysis are states or Congressional districts. To show the power of the time-series models I employ, I analyze month-level variation in unemployment by Congressional district and its effect on party-level ideal points in the House to see to what extent localized economic

shocks either polarize or de-polarize parties.

Furthermore, I interact unemployment with the level that these districts experienced adverse shocks due to exposure to Chinese trade to see whether heightened salience due to recession increases polarization among legislators in those districts even further. It has been shown that exposure to Chinese trade has important consequences for manufacturing across the United States (Autor, Dorn, and Hanson 2013), and also has implications for how legislators vote on trade-related legislation (Feigenbaum and Hall 2015) and foreign policy hostility towards China (Kuk, Seligsohn, and Zhang 2017). As such, exposure to trade with China is an important potential mediator of the effect of unemployment on legislative behavior.

The data I collect come from the Bureau of Labor Statistic’s Local Area Unemployment Statistics program, which produces unadjusted estimates of monthly unemployment by local-level units. To aggregate these numbers to the Congressional district level, I performed a spatial join by areal interpolation (i.e., weighted by amount of overlap) between the county-level monthly series and Congressional district boundaries from Lewis et al. (2013). The resulting dataset represents the average monthly unemployment data that would be most relevant to a particular Congressperson’s district. Even with this high level of disaggregation, the dataset has very low missingness (50 district-year-month observations), which I impute non-parametrically using the technique of Stekhoven and Bühlmann (2012).

I then merge this dataset with a decade-level measure of exposure to Chinese imports from Feigenbaum and Hall (2015) who aggregated information from Autor, Dorn, and Hanson (2013) to produce district-level measures of import exposure per worker. I use this measure as a direct way of determining how strongly Chinese trade has affected a particular district. The measure only has two over-time values, unfortunately, but there is substantial variation across districts. Their measure does not have specific units as it is a weighted multiplication of workers in a district in a given industry and that industry’s trade with China, although the measure is scaled to isolate Chinese dependence as opposed to other kinds of trade behavior.

Figure 1 shows the overall distribution of district-level unemployment rates from 1990 to 2018 inclusive, with a distinct mode at 5.0%, the so-called natural rate of unemployment. Figure 2 breaks out the distribution of county-level unemployment rates over time by four

Figure 1: Histogram of District-Level Unemployment

Figure 2: Monthly Unemployment Rates by U.S. County, 1990-2018

states: Alabama, California, Michigan and New York. As can be seen, there is substantial variation both over time and across counties within a state. As such, this series is an excellent data source for showing the utility of time-varying ideal point models. We can also learn empirically from seeing how the effect of these unemployment rates affects legislators in both parties in the House. For the unemployment data itself, we can consider the following hypothesis about the effect of recession on legislators' ideal points:

H1: When monthly unemployment rates rise, legislators in each party become less polarized.

This hypothesis expresses the idea that in times of recession legislators feel more pressure to work across the aisle to produce bipartisan legislation, which we can measure as the spread of ideal points between both parties. We can test this hypothesis by including unemployment rates as an exogenous covariate in the ideal point model while interacting it with an indicator for each legislator's party ID. The effect of the covariate then averages over the ideal points of each individual legislator for each party.

On the other hand, we might think that economic recessions would have the opposite effect. Perhaps polarization should increase as legislators are dealing with higher numbers of unemployed constituents. Moving towards the ideological extremes could be a way of heading off primary challengers who would seize the initiative to take advantage of incumbent weakness. So we also consider a competing alternative hypothesis:

H2: When monthly unemployment rates rise, legislators in each party become more polarized.

With the Chinese import shock data, we might surmise that polarization will increase among legislators in districts where Chinese import exposure per worker is very high. This kind of polarization could arise either by increasing divergence on trade-related legislation (Feigenbaum and Hall 2015) or via increased foreign policy hostility to China (Kuk, Seligsohn, and Zhang 2017).

H2: When monthly unemployment rates rise in districts with a higher exposure to Chinese import competition, legislators in each party become more polarized.

This hypothesis expresses a conditional relationship that runs in the opposite direction of the first hypothesis. It would seem that, as existing literature has shown, exposure to Chinese import shocks shifts legislators' ideal points on trade policy and increases polarization, that there should be a strongly interactive relationship between district-level unemployment and Chinese import exposure. When both are high, we would have a strong prior that legislators will polarize away from each other. This hypothesis can be tested with a three-way interaction of party ID, unemployment rates and Chinese shock exposure at the district level.

To test these hypotheses, I fit an AR(1) time-varying ideal point model to the data. Because the AR(1) model is a discrete-time model, I aggregate the date of votes to months. Each legislator's ideal point is estimated separately, along with legislator-specific autoregressive parameters (ψ_i), while the covariates ϕ average across this time-series variation.

I first examine the overall distributions of the ideal point scores over time. Figure 3 shows the ideal points for two Congressmen who had their ideal points constrained, Joe Linus Barton (a Republican) and Richard Edmund Neal (a Democrat). Unlike with static models, these two were selected not because they are the most ideologically extreme but rather because they served in Congress for nearly the entire sample period. As such, their long vote records help to identify the model over time. Figure 3 shows very stable, though not ideologically extreme, trends for them both. Neal is known as a moderate, so it is not as surprising that his ideal point is modestly positive. Interestingly, the variance in their ideal points declines over time, perhaps suggesting gradual hardening of political positions.

Based on this analysis of the constraints, we can infer that the positive side of the distribution represents more conservative ideology, while the negative side is more liberal. Using over-time models makes this interpretation more nuanced, of course, as it is possible for the meaning of the poles to change over time. We can plot the full combined distribution of all Congresspeople in Figure 4. This figure shows a fascinating evolution of the ideal points over time. At the beginning of Barton and Neal's careers, there were substantial numbers of conservative Democrats and liberal Republicans. By the end of their careers, the two poles

of more liberal Democrats and more conservative Republicans clearly emerge. We can see how this distribution rotates over time by highlighting the two long-serving Congressmen in Figure 5. During the twenty-five year period in the data, Neal went from a more liberal member of the House to a moderate as more liberal memvbers joined the party, while Barton maintained his position in the middle of the Republican distribution by moving rightward over time.

I next turn to the effect of unemployment on these distributions. One challenge with interpreting covariates of ideal point models is providing a clear explanation of the coefficients in terms of the measurement model. Because the IRT parameterization is non-linear and bi-polar, it is possible but not necesssarily useful to solely interpret the coefficients in terms of unit changes on the probability of the outcome. For this reason, I introduce a new statistic which I call *marginal ideal point effects* that average the covariate values over the model’s discrimination scores for each pole of the latent scale. By doing so, we can better capture the effect of the covariate on the latent quantity rather than just on the observed data.

Because the latent scale is bi-directional, there are two sets of marginal effects that I need to calculate. For positive values of the latent scale, I average the effect of the covariates on ideal points ϕ across all values of positive discrimination parameters $\gamma_j > 0$ for each posterior draw $s \in S$ and each discrimination parameter $j \in J$:

$$\frac{\partial \phi}{\partial Pr(Y_{ijrt} = 1 | \gamma_j > 0)} = \frac{\sum_{j=1}^J \Theta(X\phi'_s \gamma_{js}) - 0.5}{J} \quad (11)$$

In this equation, $\Theta(\cdot)$ represents an inverse link function that maps the ideal point values back to the observed outcome, such as a logit function for the inverse logit link (binary model) and the natural logarithm for the exponential link (Poisson distribution). For binary models with the inverse logit, I include a constant 0.5 to calculate marginal changes in probability as $Pr(0.5)$ equals 0 on the logit scale. It is important to average this quantity for each posterior draw separately as the parameters are likely correlated in the posterior. I also calculate this statistic for all values of the discrimination parameters less than zero:

Figure 3: Over-time Ideal Points for Constrained Legislators by Month, 1990-2018

$$\frac{\partial \phi}{\partial \Pr(Y_{ijrt} = 1 | \gamma_j < 0)} = \frac{\sum_{j=1}^J \Theta(X \phi'_s \gamma_{js}) - 0.5}{J} \quad (12)$$

As a result we have two sets of marginal effects that represent the effect of the covariate weighted by the model’s estimate as to whether a given item, in this case a rollcall vote, belongs to either side of the dimension. For example, if we assume that the lower end of the spectrum is liberal, then the marginal effect of ϕ for negative discrimination parameters will be weighted by the total amount of “liberalness” in the data. If there are many highly discriminating liberal bills, then the covariate’s marginal effects on liberal votes will be much higher than if there are only a few discriminating liberal bills. Another way of conceptualizing this effect is how much a covariate prompts a given person to take a position that is consistent with a given direction of the latent scale.

The ideal point marginal effects for the unemployment-ideal point interaction model are

Figure 4: Over-time Ideal Points for All U.S. House Legislators by Month, 1990-2018

Figure 5: Overlay of Constrained Legislators on Unconstrained Legislators, 1990-2018

Figure 6: Effect of District-Level Unemployment Rates (AR1 Model) on Legislators' Ideal Points 1990-2018

Figure 7: Interaction of Unemployment Rates and Chinese Import Exposure on Legislative Votes 1990-2010

shown in Figure 6. There are two coefficients that capture the interaction: the first is the Combined Republican effect, which represents the interaction term for Republican times Unemployment plus its constituent term, Unemployment. The effect of unemployment for Democrats is equal to the coefficient Unemployment as Democrats are the baseline category. On the whole, Figure 6 shows that legislators in both parties tend to move to the left during economic recessions. This is likely due to pressures to increase redistributive spending during economic downturns. However, because the left-ward movement is greater for Democrats than for Republicans, hypothesis 2 has more support because on the whole Democrats end up farther to the left than Republicans. That is, even as both groups move to the left, Democrats move farther to the left, resulting in a legislature that is on the whole more polarized. However, it should be noted that this effect is slight, and is likely not a large contributing factor of well-known polarization between Democrats and Republicans in recent years.

It is possible with the ideal point marginal effects to make precise statements about how unemployment affects legislative behavior in terms of the latent dimension. The effect magnitudes in the table appear over-stated as a one-unit change would equal a 100% increase in unemployment. For a more realistic 1% increase in unemployment, the resulting propensity to vote change is much smaller. For Republicans, a 1% increase in unemployment at the district level results in approximately a 0.1% decrease in voting for conservative bills. The effect is stronger for liberal bills: for a 1% increase in unemployment, voting for liberal bills increases by 0.18%. Because the values of discrimination do not need to be identical for both sides of the latent variable, the ideal point marginal effects are also not likely to be identical for both poles.

For Democrats, a 1% increase in unemployment is associated with a 0.2% increase in the probability of voting for liberal bills while voting for conservative bills decreases by about 0.18%. While these effects are modest, it is important to note that the underlying data is at the month level, so the cumulative effects of monthly changes in unemployment over time could still be significant.

The advantage in employing these statistics is because ideal point marginal effects fully

incorporate the utility of the underlying measurement model. These covariate effects average over measurement uncertainty in the ideological content of bills instead of relying on an arbitrary coding. For these reasons, these covariate effects can be given the interpretation of the marginal effect of unemployment on legislators' voting for conservative or liberal bills while taking into account our uncertainty over how liberal or conservative a particular bill in fact is.

The advantage of the AR(1) model relative to the more complicated GP model is that it is fairly straightforward to interpret the coefficients. Because the ideal points in this model are stationary, any over-time covariate equals movement away from the time series' long-term average, or what is known in time-series terminology as a "shock." Furthermore, we know how each ideal point time series responds to these shocks from the AR(1) parameters ψ_i . A higher magnitude of these parameters means that the ideal points display higher autocorrelation, which in terms of the covariate indicates that the effect of the covariate will have a longer lasting effect over time. The sign of the autocorrelation parameter ψ_i indicates in which direction the legislator moves given a time-series shock. A negative autocorrelation parameter will cause the legislator to oscillate in response to a shock, switching from more conservative to more liberal behavior between time points. A positive autocorrelation parameter means that the legislator responds to time-series shocks by slowly decaying to the long term mean without oscillating.

What these magnitudes and sign changes imply for legislative behavior is a fascinating area for future research. To illustrate how these auto-correlation processes matter, I calculate impulse-response functions for each legislators' ideal points time series with respect to the unemployment covariate. An impulse response function measures the marginal effect of a covariate on a time series over time. For a given covariate ϕ and an ideal point time series α_{it} , we can define it as follows for a given time point t where the shock ϕ occurs:

$$\frac{\partial \alpha_{xt+1}}{\partial \phi} = \psi_i \phi \quad (13)$$

We can then iterate over future time periods t by forward-substituting α_{xt} with the values of $\psi_i \phi$ from the prior time period:

Figure 8: Impulse Response Functions for Unemployment on Democrats' Ideal Points

$$\frac{\partial \alpha_{xt+2}}{\partial \phi} = \psi_i(\psi_i \phi) \quad (14)$$

This recursive algorithm can be run for an infinite number of time periods as the effect will never decay truly to zero, although in general ten time periods is usually more than sufficient to capture the impulse-response function dynamics. I calculate these impulse-response dynamics for the unemployment-party ID interaction on latent score marginal effects for each legislator. Figures 8 and 9 show these impulse-response functions for Democrat and Republican legislators respectively. What is interesting to note is that oscillation rather than slow decay is the dominant time-series behavior for legislators in both parties. We also see that there are substantial levels of autocorrelation with effects persisting through the whole 10-month period over which the impulse response functions were calculated.

To obtain the effects in the covariate plots in Figure 6, the covariate effect would need to be

Figure 9: Impulse Response Functions for Unemployment on Republicans' Ideal Points

averaged over the time-series trends shown in these figures. In other words, while it might seem at first counter-intuitive that legislators would flip from side to side as a result of an unemployment shock, that is because the model estimates that in general these legislators tend to respond to such shocks through oscillation rather than slow decay. A positive effect from Figure 6 will result in higher positive oscillations versus negative oscillations over time for these legislators. Again, this difference in time series behavior is substantively interesting, but would require further research to understand the reasons for why legislators respond to shocks in varying ways.

Overall, the results of these models do show that legislative behavior changes during times of higher local unemployment. Furthermore, we can measure these movements in ideal points down to the monthly-level, permitting very precise statements about how legislative behavior changes. In general, the effects of 1% change in district-level unemployment on legislators are modest, though we would expect modest effects as this 1% change represents one-month's worth of unemployment data. Aggregated over time, these effects could grow to be substantial.

Finally I turn to the results of the Chinese import exposure model. The ideal point marginal effects for this model are shown in Figure 7. Unfortunately, it is not possible to read off the interactive relationship as the main effect is a continuous by continuous by binary variable interaction. The only way to interpret this relationship is by calculating the effect of the interaction for different values of Chinese import exposure per worker and party ID. This plot for the mean posterior values is shown in Figure 10. There is a strongly non-linear relationship between Chinese import exposure and unemployment. At low rates of Chinese import exposure, unemployment appears to have a weak effect on legislators in either party. At higher levels of import exposure, however, there is a much stronger relationship. In this case, it would appear that the relationship is strongly de-polarizing. Democrats with in districts with high unemployment and high import exposure tend to vote more conservative, while Republicans in similar districts tend to vote more liberal than other members of their party. This suggests that Chinese import exposure is indeed a strong moderating influence on the relationship between unemployment and voting behavior. Unfortunately, it is not

Figure 10: Effect of Unemployment Conditional on Chinese Import Exposure per Worker, 1990-2010

possible to do a straightforward nested model comparison between the two-way and three-way interaction models, though further probing of the relationship could yield further insights into why these findings seem to run in opposite directions.

7 Discussion

The aim of this paper was to put forward a general framework for ideal point estimation that would both integrate existing approaches while providing new areas for analysis, particularly in missing data and over-time inference. The intention is not to suggest, however, that this framework represents all that can be done with ideal points. This paper's approach is general and designed to fit a wide variety of applications, but investigating the intricacies of social

situations will always require more custom modeling. The utility of this study is rather to promote the use of ideal point models as a core part of empirical research in political science, not just for legislative studies or rollcall voting data.

The reason that ideal point models have wider applicability is because social choice processes where individuals decide between competing alternatives happen very often. If these alternatives are polarizing, or represent competing poles, than an ideal point model can help locate people’s positions along the axis of competition while also determine the relative weight of the items they are choosing. The ultimate payoff is that this kind of modeling will help political scientists focus on inference on the constructs they care about, which are usually unobservable, such as partisanship, ethnic identity, corruption, and political ideology. Instead of being forced into arbitrary measurement decisions over the available indicators, political scientists can maximize the information in their data by collapsing multiple indicators to a single dimension.

As a result, this paper largely sidesteps the debate about how to determine when ideal points represent political ideology. Defining a the latent construct can only be done in the context of a specific empirical application. The political ideology interpretation is most useful for American politics where legislators have considerable latitude in voting behavior, but this does not mean that ideal point representations of other polities or social situations are not as useful. So long as a social choice process is involved, ideal point models can extract useful information from noisy data.

8 Conclusion

In this paper, I presented a generalization of the increasingly popular ideal point model in the Bayesian IRT framework. I extended the model with new modes of missing-data, time-varying and joint distribution inference. Rather than focus on a single empirical application, the contribution of this paper is an analytical tool that can extend the domain and applicability of ideal point models across the discipline as all of these models are available in a single R package, `idealstan`. Crucially, all of these methods build on each other so that

missing data can be incorporated in time-varying models or in joint distributions and vice versa.

For missing-data inference, I implement a two-stage selection model in which social actors first decide whether to signal their ideal points by providing observed data or to censor their ideal points by choosing not to provide any observable data. This model is flexible enough to adjust ideal points whenever missing data is correlated with either pole of the ideal point spectrum, and has wider applicability than just legislative contexts.

I implemented two new time-varying models for ideal points, the AR(1) stationary model and the Gaussian process model. The AR(1) model imposes more structure on the time-series trends in ideal points, but also has more interpretable coefficients than the Gaussian process model. The Gaussian process model, on the other hand, is able to fit a wide variety of time-series trends. Both models can fit time-varying covariates to predict ideal points, permitting new forms of precise inference on what changes ideal points.

All of these models are available in an R package called `idealstan`. This package is designed to automate the sometimes arduous process of preparing data for ideal point modeling, including identifying parameters. Furthermore, the use of a single R package for all models enables researchers to compare different models on the same data with ease. Big data alternatives to standard Bayesian inference are also incorporated in the package to enable estimation of the ever larger data sets available to political scientists.

References

- Alesina, Alberto, and Howard Rosenthal. 1995. *Partisan Politics, Divided Government and the Economy*. Cambridge University Press.
- Armstrong, David A., Ryan Bakker, Royce Carroll, Christopher Hare, Keith T. Poole, and Howard Rosenthal. 2014. *Analyzing Spatial Models of Choice and Judgment with R*. CRC Press.
- Autor, David H., David Dorn, and Gordon H. Hanson. 2013. “The China Syndrome: Local

- Labor Market Effects of Import Competition in the United States.” *The American Economic Review* 103 (6): 2121–68.
- 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.
- Baker, Scott R., Nicholas Bloom, and Steven J. Davis. 2016. “Measuring Economic Policy Uncertainty.” *The Quarterly Journal of Economics* 131 (4): 1593–1636.
- Barberá, Pablo. 2015. “Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data.” *Political Analysis* 23: 76–91.
- Barberá, Pablo, John T. Tost, Jonathan Nagler, Joshua A. Tucker, and Richard Bonneau. 2015. “Tweeting from Left to Right: Is Online Political Communication More Than an Echo Chamber?” *Association for Psychological Science* 26 (10): 1531–42.
- Betancourt, Michael J., and Mark Girolami. 2013. “Hamiltonian Monte Carlo for Hierarchical Models.” *arXiv*, December.
- Blei, David M., Alp Kucukelbir, and Jon D McAuliffe. 2017. “Variational Inference: A Review for Statisticians.” *Journal of the American Statistical Association* 112 (518).
- Bräuninger, Thomas, Jochen Müller, and Christian Stecker. 2016. “Modeling Preferences Using Roll Call Votes in Parliamentary Systems.” *Political Analysis* 24 (2): 189–210.
- Carpenter, Bob, Andrew Gelman, Matthew D. 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).
- Carroll, Royce, Jeffrey B. Lewis, James Lo, Keith T. Poole, and Howard Rosenthal. 2013. “The Structure of Utility in Spatial Models of Voting.” *American Journal of Political Science* 57 (4): 1008–28.
- Carroll, Royce, Jeffrey B. Lewis, James Lo, Keith T. Poole, and Howard L. Rosenthal. 2009. “Comparing Nominate and Ideal: Points of Difference and Monte Carlo Tests.” *Legislative Studies Quarterly* 34 (4): 555–91.

- Caughey, Devin, and Eric Schickler. 2016. "Substance and Change in Congressional Ideology: NOMINATE and Its Alternatives." *Studies in American Political Development* 30 (2): 128–46.
- Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Rollcall Data." *American Political Science Review* 98 (2): 355–70.
- Coppedge, Michael, John Gerring, Staffan I. Lindberg, Svend-Erik Skaaning, Jan Teorell, David Altman, Michael Bernhard, et al. 2017. "V-Dem Dataset V7." Working Paper. Social Science Research Network. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2968289.
- Durbin, J. 1954. "Errors in Variables." *Review of the International Statistical Institute* 1 (22): 23–32.
- Enelow, James M., and Melvin J. Hinich. 1984. *The Spatial Theory of Voting: An Introduction*. Cambridge University Press.
- Feigenbaum, James J., and Andrew B. Hall. 2015. "How Legislators Respond to Localized Economic Shocks: Evidence from Chinese Import Competition." *Journal of Politics* 77 (4): 1012–30.
- Gelman, Andrew. 2018. "Benefits and Limitations of Randomized Control Trials: A Commentary on Deaton and Cartwright." *Social Science & Medicine* 210: 48–49.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2013. *Bayesian Data Analysis*. CRC Press.
- Greenacre, Michael. 2017. *Correspondence Analysis in Practice*. 3rd ed. Chapman; Hall/CRC.
- Grimmer, Justin. 2011. "An Introduction to Bayesian Inference via Variational Approximation." *Political Analysis* 19 (1): 32–47.
- Harman, Harry H. 1960. *Modern Factor Analysis*. University of Chicago Press.
- Heckman, James J., and Guilherme Sedlacek. 1985. "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market." *Journal of Political Economy* 93 (6): 1077–25.

- Hollyer, James R., B. Peter Rosendorff, and James Raymond Vreeland. 2018. *Information, Democracy and Autocracy: Economic Transparency and Political (in)Stability*. Cambridge University Press.
- Hotelling, H. 1933. "Analysis of a Complex of Statistical Variables into Principal Components." *Journal of Educational Psychology* 24 (6): 417–41.
- Imai, Kosuke, and In Song Kim. 2016. "When Should We Use Linear Fixed Effects Regression Models for Causal Inference with Longitudinal Data?" Online.
- Krehbiel, Keith, and Zachary Peskowitz. 2014. "Legislative Organization and Ideal Point Bias." *Journal of Theoretical Politics* 27 (4): 673–703.
- Kropko, Jonathan. 2013. "Measuring the Dynamics of Political Power: A Time-Series Irt Model." Presentation. Political Methodology Society.
- Kubinec, Robert, and John Owen. 2018. "When Groups Fall Apart: Measuring Transnational Polarization with Twitter from the Arab Uprisings." *SocArchiv*, August. <https://doi.org/10.31235/osf.io/wykmj>.
- Kucukelbir, Alp, Rajesh Ranganath, Andrew Gelman, and David Blei. 2015. "Automatic Variational Inference in Stan." In *Advances in Neural Information Processing Systems 28*, edited by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, 568–76. Curran Associates, Inc. <http://papers.nips.cc/paper/5758-automatic-variational-inference-in-stan.pdf>.
- Kuk, John Seungmin, Deborah Seligsohn, and Jiakun Jack Zhang. 2017. "From Tiananmen to Outsourcing: The Effect of Rising Import Competition on Congressional Voting Towards China." *Journal of Contemporary China* 27 (109): 103–19.
- Lewis, Jeffrey B., Brandon DeVine, Lincoln Pitcher, and Kenneth C. Martis. 2013. "Digital Boundary Definitions of United States Congressional Districts, 1789-2012." <http://cdmaps.polisci.ucla.edu/>.
- Lewis, Jeffrey B., and Chris Tausanovitch. 2018. "Gpuideal: Fast Fully Bayesian Estimation of Ideal Points with Massive Data." *Working Paper*, August. <http://www.ctausanovitch.com/gpuideal.pdf>.

- Little, Roderick J. A., and Donald B. Rubin. 2002. *Statistical Analysis with Missing Data*. Hoboken, NJ: J. Wiley.
- 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.
- Poole, Keith T., Jeffrey B. Lewis, James Lo, and Royce Carroll. 2008. “Scaling Roll Call Votes with W-Nominate in R.” Working Paper. Social Science Research Network. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1276082.
- Poole, Keith T., and Howard L. Rosenthal. 2007. *Ideology & Congress*. Transaction Publishers.
- Rabiner, L. R. 1989. “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition.” *Proceedings of the IEEE* 77 (2): 257–86.
- Rasmussen, Carl Edward, and Christopher K. I. Williams. 2006. *Gaussian Processes for Machine Learning*. MIT Press.
- Roesser, R. 1975. “A Discrete State-Space Model for Linear Image Processing.” *IEEE Transactions on Automatic Control* 20 (1): 1–10.
- Rosas, Guillermo, Yael Shomer, and Stephen R. Haptonstahl. 2015. “No News Is News: Nonignorable Nonresponse in Roll-Call Data Analysis.” *American Journal of Political Science* 59 (2): 511–28.
- Slapin, Jonathan B., and Sven-Oliver Proksch. 2008. “A Scaling Model for Estimating Time-Series Party Positions from Texts.” *American Journal of Political Science* 52 (3): 705–22.
- Stekhoven, Daniel J., and Peter Bühlmann. 2012. “MissForest-Non-Parametric Missing Value Imputation for Mixed-Type Data.” *Bioinformatics* 28 (1): 112–18.
- Tahk, Alex M. 2015. “A Continuous-Time, Latent-Variable Model of Time-Series Data.” *Political Analysis* 26 (2): 131–46.
- . 2018. “Nonparametric Ideal-Point Estimation and Inference.” *Political Analysis* 26

(2): 131–46.

Ullman, Jodie B., and Peter M. Bentler. 2012. “Structural Equation Modeling.” In *Research Methods in Psychology*, edited by Irving B. Weiner, John A. Schinka, and Wayne F. Velicer. Vol. 2. Wiley.