Norman Schofield · Gonzalo Caballero · Daniel Kselman *Editors*

Advances in Political Economy

Institutions, Modelling and Empirical Analysis

This book presents latest research in the field of Political Economy, dealing with the integration of economics and politics and the way institutions affect social decisions. The focus is on innovative topics such as an institutional analysis based on case studies; the influence of activists on political decisions; new techniques for analyzing elections, involving game theory and empirical methods.

Social Sciences / Political Science

ISBN 978-3-642-35238-6

9 783642 352386

▸ springer.com

Schofield · Caballero
Kselman *Eds.*

Advances in Political Economy

Norman Schofield
Gonzalo Caballero
Daniel Kselman *Editors*

Advances in Political Economy

Institutions, Modelling and Empirical Analysis

Springer

Zaller JR (1992) The nature and origins of mass opinion. Cambridge University Press, New York
Zaller JR (2004) Floating voters in U.S. presidential elections, 1948–2000. In: Saris WE, Sniderman PM (eds) Studies in public opinion: attitudes, nonattitudes, measurement error, and change. Princeton University Press, Princeton

Inferring Ideological Ambiguity from Survey Data

Arturas Rozenas

Keywords Ideological placement · Ambiguity · Bayesian · Latent variables · Missing data

1 Introduction

It has become conventional wisdom to think of electoral competition in terms of parties taking positions on policy issues and voters choosing their representatives based on those positions. Quite often, however, instead of communicating clear platforms, politicians make contradicting policy statements, remain ambiguous about details or avoid talking about issues altogether. For example, Mitt Romney, a presidential candidate in the U.S. 2012 elections, has been constantly accused of remaining too vague on key policy issues.[1] In the 2008 presidential campaign, Barack Obama promised to withdraw the U.S. troops from Iraq within 16 months whereas John McCain proposed far more ambiguous plan to remain in Iraq for "up to 100 years."[2]

To explain ideological ambiguity, spatial theorists have referred to risk-attitudes of the voters (Shepsle 1972), the desire of politicians to avoid divisive issues (Page 1976), context-dependence of voting decisions (Callander and Wilson 2008), uncertainty of the candidates(Glazer 1990), or strategic benefits of not committing to a certain platform (Alesina and Cukierman 1990; Aldrich 1995). Empirical research, on the other hand, focused mostly on voting behavior finding that accounting for ideological ambiguity improves predictions of the standard spatial voting models (Alvarez 1997; Bartels 1986; Campbell 1983a,b; Tomz and van Houweling 2009).

These examples suggest theoretical and empirical reasons to treat policy platforms not as points but as probability distributions over policy space. Indeed, the

[1] For example, "Where are Mitt Romney's details?", by Scott Lehigh, *Boston Globe*, June 27, 2012.

[2] 'Obama Fuels Pullout Debate With Remarks', *New York Times*, July 4, 2008.

I am grateful to John Aldrich, Scott Desposato, Jeremy Reiter, Fan Li, Mitchell Seligson, and James Stimson for comments and suggestions.

A. Rozenas (✉)
ISM University of Management and Economics, LT-01129, Vilnius, Lithuania
e-mail: arturasro@gmail.com

N. Schofield et al. (eds.), *Advances in Political Economy*,
DOI 10.1007/978-3-642-35239-3_18,

notion of policy platforms as points has a very limited reach. For example, it cannot be applied to study policy positions of decentralized political parties involving a variety of activists with diverse policy preferences (Aldrich 1983; Miller and Schofield 2003). Another case concerns developing democracies, where, for many reasons, parties are known to lack defined ideological positions (Evans and Whitefield 2000; Kitschelt et al. 1999; Mainwaring 1995; Scully 1995). If policy positions are defined as points, it is not clear what it means for a party or a candidate not to have a position. Conceptualizing policy position as a probability distribution provides a more general approach to empirical study of party competition: a "no position" platform can be described by a highly dispersed distribution whereas a platform as a point can be defined as a distribution with a vanishingly small dispersion.

Although there are multiple reasons to study ideological ambiguity, efficient tools to measure this quantity are lacking. The existing scholarship on the measurement of policy positions operates under the assumption that these positions are points, often even referred to as 'ideal points' (Ansolabehere et al. 2001; Clinton et al. 2004; Laver et al. 2003; Martin and Quinn 2002). This paper presents a statistical model to estimate ideological ambiguity from survey data (e.g., opinion polls or expert surveys)—the kind of data that is widely available in terms of temporal depth and geographical width.

2 Survey Data and Ambiguity Measurement

The existing literature offers two approaches for measuring ideological ambiguity. The first approach uses direct measures by asking respondents to report their uncertainty about the position of a given candidate (Alvarez 1997) or by asking them to place political actors on a scale in a form of an interval rather than a point (Tomz and van Houweling 2009). Unfortunately, such surveys are rare making it difficult to use these approaches for a systematic study of ideological ambiguity, especially in a cross-national context.

Another approach is to use indirect methods where ambiguity is inferred either from disagreement among the respondents (Campbell 1983a,b) or from the patterns in the missing survey data (Bartels 1986). These indirect methods can be applied to many data sets, which ask citizens or political experts to place political parties on a policy scale. However, a naive application of these approaches is wanting, as I discuss now.

2.1 Interpreting Respondent Disagreement

Every survey where respondents are asked to place political candidates on issue scales generates variation in judgments. It appears intuitive to use the sample standard deviation of the placements $\hat{\sigma}$ as an estimate of a party's ideological ambiguity as suggested by Campbell (1983a,b). However, the intuition is flawed on several levels. First, a high degree of disagreement between the respondents (and hence high $\hat{\sigma}$)

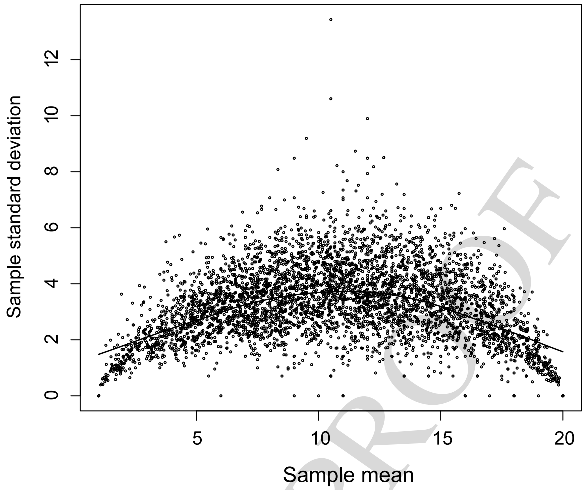

Fig. 1 Sample mean and standard deviation in Benoit and Laver (2006) expert data

may indicate the lack of information on the part of respondents (Marks et al. 2007) or an intrinsic ambiguity of a party's policy position Campbell (1983a,b). Thus, to correctly estimate ideological ambiguity, we have to disentangle the respondent-level and party-level effects on the observed respondent disagreement.

Second, respondent disagreement might occur due to the scale-heterogeneity effect: even if a party is not ambiguous and respondents are well-informed, they might provide conflicting placements due to different interpretation of the measurement scale. Treating disagreement among respondents without proper adjustments for the scaling effects can result in faulty inference about ideological ambiguity.

The third flaw of $\hat{\sigma}$ as the estimator of ideological ambiguity stems from the ordinal nature of placement scales. Since the respondents are almost universally required to place parties on an ordinal scale, the measurement procedure induces dependence between the sample mean, $\hat{\mu}$, and the sample standard deviation, $\hat{\sigma}$. It is easily demonstrated that for an M category measurement scale, $\hat{\sigma} \leq \sqrt{\hat{\mu}(M - \hat{\mu})}$. Therefore, parties with extreme positions will necessarily be evaluated as less ambiguous simply due to the mathematical properties of the estimators $\hat{\mu}$ and $\hat{\sigma}$. Indeed, this pattern is well represented in the real data on party positions in Fig. 1. The quadratic pattern in Fig. 1 could represent the 'true' relationship between positions of candidates and their ambiguity, or it can merely be an artifact of the measurement model; if we use $\hat{\sigma}$ as our estimate of ambiguity, we simply cannot evaluate which is the case.

2.2 *Interpreting Missing Values*

A different approach to ambiguity measurement is offered by Bartels (1986), who suggests that respondents are more likely not to place a party on a policy scale if they are uncertain about its platform. In Bartels' model, the source of uncertainty is

Table 1 An example of missing data pattern from Benoit and Laver (2006) expert survey

Expert/Party	PAD	PBDNJ	PD	PDr	*PLL*	PR	PS	PSD
1	NA	9	12	13	14	13	7	8
2	18	19	14	17	NA	16	11	10
3	18	13	16	15	NA	19	3	1
4	NA	2	2	2	NA	2	2	2
5	NA	14	14	14	NA	14	14	14
6	NA	NA	12	NA	NA	NA	9	NA
7	NA	10	14	16	NA	13	8	4
8	NA	8	17	13	18	16	3	7
9	7	10	9	12	NA	5	2	NA
10	8	3	14	12	NA	15	5	7
11	NA	8	12	NA	NA	18	5	16
12	NA	12	15	16	NA	16	9	11
13	NA	3	6	10	NA	6	7	4
14	NA	6	15	16	NA	18	3	5
15	NA	5	5	5	NA	7	4	3
16	NA	5	15	15	NA	15	3	3

respondents' personal characteristics like education or exposure to media. One can extend this idea further and argue that the uncertainty about platforms may have to do not only with the respondent-level knowledge but also with the ambiguity of the platform that is being evaluated.

Table 1 shows an excerpt of expert-data on Albanian political parties from the expert survey in Benoit and Laver (2006). Evidently, there are party-specific and expert-specific effects in the non-response rates: PAD and PLL are the two parties with high non-response rates and experts 6 and 11 appear to be the least knowledgeable. It is reasonable to assume that PAD and PLL have such high non-response rates because they are ambiguous about the given policy—the parties either did not make any public statements on the policy or the statements they made varied greatly in their content.

In sum, the discussion suggests that a proper method for estimating ideological ambiguity should (1) adjust for the scale-heterogeneity effects, (2) separate the respondent- and party-level effects on observed respondent disagreement, and (3) exploit the patterns in missing data as an additional source of information on ideological ambiguity.

3 A Model

Suppose that data provided by respondents $i = 1, \dots, N$ on parties $j = 1, \dots, J$ are generated in a two-step process. In the first stage, respondents perceive a 'true' position of each party on some given issue-scale. Since we define party positions to be probability distributions, we treat each such perception as a random draw from that probability distribution.[3]

In the second stage, the respondent has to report his/her observed value on some ordinal measurement scale. The key issue here is that respondents might differ in their interpretation of the measurement scale. To account for that, I follow the framework by Aldrich and Mckelvey (1977) and assume that each respondent's reported placement is an affine transformation of his/her latent perception. Formally, suppose that the measurement scale has M points and let $C = \{c_m : m = 1, \dots, M\}$ be an ordered set of cut-off points with $c_1 = -\infty$ and $c_M = \infty$. Let z^*_{ij} denote unobserved latent perception of party's j platform by respondent i. The latent policy positions are defined as Gaussian probability distribution functions:

$$z^*_{ij} \sim \mathcal{N}\big(\mu_j, \sigma_j^2\big), \tag{1}$$

$$y_{ij} = m \quad \text{iff } c_m < \psi_i z^*_j + \tau_i \le c_{m+1} \tag{2}$$

where τ_i and ψ_i are expert-specific location and scale parameters accounting for scale heterogeneity. A respondent with a low ψ_i tends to place parties closer to each other than a respondent with high ψ_i. Similarly, a respondent with a high τ_i tends to place parties on the right side of the scale relative to a respondent with low τ_i.

Alternatively, one could specify a common location and scale parameter and allow each respondent to have an idiosyncratic cut-off point, similar to Johnson and Albert (1999) and Clinton and Lewis (2007). For an M point scale, this alternative approach introduces $N(M-1)$ respondent-level parameters. In comparison, the model in (5)–(6) has only $2N$ respondent-level parameters. Given that the number of parties in any survey is typically small and M is large, a more parsimonious model is preferred.

The model in (1)–(2) can be seen as an extension of some widely used ordinal data models. It represents cross-classified (rather than nested) hierarchical model (Zaslavsky 2003, p. 341). When $\psi_i = 1$ and $\sigma_j = 1$ for all i and j, we would have the usual random-effects linear model coupled with ordinal data. For $\sigma_j^2 = \sigma$ for all j, the model results in the scaling model by Aldrich and Mckelvey (1977).[4] Finally, when $\sigma_j^2 = 1$ for each j, the model resembles the multiple-rater model as presented in Johnson and Albert (1999, Chap. 5) and applied to expert data by Clinton and Lewis (2007). In contrast to these alternatives, we allow σ_j's and ψ_i's to vary across parties and respondents respectively.

Since the policy space is defined only up to an affine transformation, Aldrich and Mckelvey (1977) suggest to constrain the estimates of μ to have zero mean and unit

[3]For example, such interpretation of respondent opinions has been used in the risk analysis literature (Huyse and Thacker 2004).

[4]Palfrey and Poole (1987) analyzed how assumption of heterogeneous variance affects inference about μ but did not address how σ should be estimated.

standard deviation. These constraints turn out to be insufficient to identify the model in (1)–(2). The following restrictions are imposed instead:

$$\mu_j \in [c_1 - \delta, c_{M-1} + \delta] \quad \text{for } j = 1, \dots, J, \tag{3}$$

$$\sum_{i=1}^{N} \tau_i = 0 \quad \text{and} \quad \sum_{i=1}^{N} \psi_i^2 = 1 \quad \text{for } i = 1, \dots, N. \tag{4}$$

Here, δ is a hyper-parameter estimated in the model. Finally, we assume that the cut-off points are fixed at equal intervals between -1 and 1 (any other interval would do as well). Since policy space is defined only up to affine transformation, these constraints do not result in loss of information.

3.1 Model for Missing Data

The model can be extended to exploit the patterns in the missing data (NA responses) as an additional source of information about the ideological ambiguity. In particular, I assume that if a party is perceived to be very ambiguous and/or if a respondent is not knowledgeable, one is more likely to observe an NA answer. Thus, in the terminology of Little and Rubin (1987), we assume that the missing data are non-ignorable. For convenience, let $z_{ij} = \psi_i z_j^* + \tau_i$. Also let $r_{ij} = 1$ if data entry y_{ij} is missing and $r_{ij} = 0$ otherwise. The model for the observed data can be written as

$$z_{ij} \sim \mathcal{N}\big(\psi_i \mu_j + \tau_i, \psi_i^2 \sigma_j^2\big), \tag{5}$$

$$y_{ij} = \begin{cases} m & \text{if } c_m < z_{ij} \le c_{m+1} \text{ and } r_{ij} = 0 \\ \text{NA} & \text{if } r_{ij} = 1, \end{cases} \tag{6}$$

$$\Pr(r_{ij} = 1) = \Phi\big(\alpha_0 + \alpha_1 \sigma_j \psi_i\big). \tag{7}$$

Notice, first, that if a respondent is not highly knowledgeable (high ψ_i) or a party is ambiguous (high σ_j), or both, the answers will exhibit high variation. Second, z_{ij}'s that are drawn from distributions with low standard deviation (low $\psi_i \sigma_j$) are less likely to be reported as NA's, as implied by the missingness model in (7). Here, Φ is a standard normal distribution function, resulting in a probit model. By making missingness dependent both on σ_j and ψ_i we allow for data distributions where some parties and/or some respondents tend to have more missing values than others. Parameter α_1 measures how much missingness in the data depends on the respondent-level scale ψ_i and party-level ambiguity σ_j.

3.2 Prior Distributions

The model is completed by specifying prior distributions. If a cross-national survey is used, one can specify hierarchical priors where some party-level parameters depend upon country-level hyper-parameters. Let $k = 1, \dots, K$ denote a country in which the survey is taken. The mean ideological positions μ_{jk} are assumed *a priori* to follow truncated normal distributions so that

$$\mu_{jk}|\delta \sim \mathcal{N}(0, \eta_\mu)\mathbb{1}\big[\mu_{jk} \in (c_1 - \delta, c_{M-1} + \delta)\big], \tag{8}$$

$$\ln(\delta) \sim \mathcal{N}(c_{i+1} - c_i, v_\delta). \tag{9}$$

We set $\eta_\mu = 100$ resulting in a vague but proper prior distribution. The hyper-parameter δ is *a priori* set to have a log-normal distribution with mean equal to the distance between any two cut-off points. We set $v_\delta = 1$, resulting in identifiable and yet highly flexible model: under this specification, we have 0.44 prior probability that the most extreme party is two units (one-fifth of the scale) away from the smallest or largest cut-off point. For the remaining parameters, we set

$$\sigma^2_{jk}|b_k \sim \text{Inv-Gamma}\big(a, (a-1)b_k\big), \tag{10}$$

$$b_k \sim \text{Gamma}(\epsilon, \epsilon), \tag{11}$$

$$\psi_{ik} \sim \mathcal{U}\big(\tfrac{1}{2}, 2\big), \tag{12}$$

$$\tau_{ik} \sim \mathcal{N}(0, 1). \tag{13}$$

In (10), the shape and scale of the inverse gamma distribution is fixed so that the $\mathbb{E}(\sigma^2_{jk}) = b_k$. Setting $a = 4$ yields a priori variance of $b^2_k/2$. Letting ϵ be a small number (e.g., 0.1), yields a prior on b_k with large variance; thus, the priors end up having a negligible effect on the estimates. The hierarchical priors induce adaptive shrinkage: the estimates of ideological ambiguity in a country k are shrunken towards the common mean b_k. The statistical advantages of the hierarchical shrinkage are well-documented in the literature (Gelman et al. 2003).

Further, under priors in (12), each respondent can expand or shrink the perceptual space at most by a factor of two. Notice that as $\psi_i \to 0$, the distribution of z_{ij} collapses to a degenerate distribution with the mass at τ_i. This implies that, for ψ_i near zero, a respondent would place all parties on the same point. Similarly, if ψ_i is very large, a respondent places all parties on the opposite extremes of the scale. Since both of these alternatives are not common, we constraint ψ_i's to the specified interval.

Relative to the scale of cut-off points, the prior distribution of τ_i in (13) also allows sizable idiosyncratic location shifts. Lastly, the selection model parameters α_0 and α_1 are assumed to follow normal distributions with 0 mean and variance of 100 (a higher variance reduces the speed of convergence without affecting the results).

4 Parameter Estimation

The model is estimated using Markov Chain Monte Carlo (MCMC) methods using Gibbs sampling approach (Gelfand and Smith 1990). Let N_k and J_k denote the number of respondents and number of parties in country k respectively. Let N_{jk} denote the set of respondents in country k who have placed party j on the scale ($N_k - N_{jk}$ is the number of NA answers for party j in country k). Let $\boldsymbol{y} = (\boldsymbol{y}_{obs}, \boldsymbol{y}_{mis})$, where $\boldsymbol{y}_{obs}$ is observed and $\boldsymbol{y}_{mis}$ is missing data respectively and let $\boldsymbol{z}$ and $\boldsymbol{r}$ denote a vector of latent perceptions z_{ijk} and missing data indicators r_{ijk} respectively. For brevity, let θ denote all parameters of the model. The joint distribution of $\boldsymbol{y}$, $\boldsymbol{z}$ and $\boldsymbol{r}$ is

$$\pi(\boldsymbol{y}, \boldsymbol{z}, \boldsymbol{r}|\theta) = \pi(\boldsymbol{y}_{obs}, \boldsymbol{y}_{mis}, \boldsymbol{z}|\boldsymbol{r}, \theta)\pi(\boldsymbol{r}|\theta). \tag{14}$$

This factorizations yields a pattern-mixture model with shared parameters (Little 1993). In this model, there is a set of common parameters affecting both the distribution of data $\boldsymbol{y}$ and missingness pattern in $\boldsymbol{r}$. In the model of missingness given in (7), the distribution of $\boldsymbol{r}$ depends on the vectors of ambiguity and uncertainty parameters $\boldsymbol{\sigma}$ and $\boldsymbol{\psi}$ and the coefficient vector $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)$. Note that this model differs from selection models of missing data where the distribution of $\boldsymbol{r}$ depends on $\boldsymbol{y}_{obs}$ and $\boldsymbol{y}_{mis}$ but not on the data model parameters. The model for the observed data is derived by integrating out the missing data from the complete data model, so that

$$\begin{aligned}\pi(\boldsymbol{y}_{obs}, \boldsymbol{z}, \boldsymbol{r}|\theta) &= \int \pi(\boldsymbol{y}_{obs}, \boldsymbol{y}_{mis}, \boldsymbol{z}|\boldsymbol{r}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\psi})\pi(\boldsymbol{r}|\boldsymbol{\sigma}, \boldsymbol{\psi}, \boldsymbol{\alpha}) d\boldsymbol{y}_{mis} \\ &= \pi(\boldsymbol{y}_{obs}|\boldsymbol{r}, \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\tau}, \boldsymbol{\psi})\pi(\boldsymbol{r}|\boldsymbol{\sigma}, \boldsymbol{\psi}, \boldsymbol{\alpha}).\end{aligned} \tag{15}$$

This yields the complete data likelihood, which is a product of two likelihoods—one for observed data and one for missing data:

$$\begin{aligned}L(y, z, r; \theta) \propto &\prod_{k=1}^{K} \prod_{i \in N_{jk}} \prod_{j=1}^{J_k} \pi(y_{ijk}, z_{ijk}|\mu_{jk}, \tau_{ik}, \sigma_{jk}, \psi_{ik}) \\ &\times \prod_{k=1}^{K} \prod_{i=1}^{N_k} \prod_{j=1}^{J_k} \pi(r_{ijk}|\sigma_{jk}, \psi_{ik}, \boldsymbol{\alpha}).\end{aligned} \tag{16}$$

Using previously specified prior distributions, the full conditionals for most of the parameters in the model have known distributional form. Specifically, the Gibbs sampler iterates between the following blocks:

1. Sample latent perceptions z_{ijk} conditional on the observed data y_{ijk} and all parameters of the model:

$$z_{ijk}|y_{ijk}, \cdot \sim \mathcal{N}\big(\mu_{jk}\psi_{ik} + \tau_{ik}, \psi_{ik}^2\sigma_{jk}^2\big)\, \mathbb{1}(c_{y_{ijk}} < z_{ijk} \leq c_{y_{ijk}+1}).$$

2. Given the latent variables $\boldsymbol{z}$, the remaining full conditionals do not depend on the ordinal data $\boldsymbol{y}$. The means of the platforms are sampled as follows:

$$\mu_{jk}|z_{jk}, \cdot \sim \mathcal{N}\big(S_{jk}/D_{jk}, \sigma_{jk}^2/D_{jk}\big)\mathbb{1}(c_1 - \delta < \mu_{jk} < c_{M-1} + \delta),$$

Book ID: 306518_1_En, Date: 2013-02-19, Proof No: 1, UNCORRECTED PROOF

where $S_{jk} = \sum_{ik}(z_{ijk} - \tau_{ik})/\psi_{ik}$ and $D_{jk} = N_{jk} + \sigma^2_{jk}/\eta_\mu$. The respondent location parameters are sampled as follows:

$$\tau_{ik}|\cdot \sim \mathcal{N}\left(\frac{\sum_j (z_{ijk} - \psi_{ik}\mu_{jk})/\sigma^2_{jk}}{\sum_{jk}\sigma^{-2}_{jk} + \psi^2_{ik}}, \frac{\psi^2_{ik}}{\sum_{jk}\sigma^{-2}_{jk} + \psi^2_{ik}}\right).$$

3. Full conditionals for $\boldsymbol{\sigma}$ and $\boldsymbol{\psi}$ do not have a recognizable form, thus Metropolis-Hastings algorithm is employed. The log-posterior of σ^2_j is proportional to

$$\ln\pi\left(\sigma^2_j|z,\mu,\tau,\psi,a_k,\alpha\right) \propto -(N_{jk}/2+3)\ln\sigma^2_{jk} - \sigma^{-2}_{jk}\left(S_{jk}/2\psi^2_{ik} + b_k\right) + \sum_{i\in N_k} r_{ijk}\ln p_{ijk} + (1-r_{ijk})\ln(1-p_{ijk}),$$

where $p_{ijk} = \Phi(\alpha_0 + \alpha_1\phi_{jk}\sigma_{jk})$ and $S_{jk} = \sum_{i\in N_{jk}}(z_{ijk} - \psi_{ik}\mu_{jk} - \tau_{ik})^2$. The log-posterior for $\boldsymbol{\psi}^2$ has a similar form:

$$\ln\pi\left(\psi^2_{ik}|z,\mu,\sigma,\tau\right) \propto -(J_{ik}/2+1/2)\ln\psi^2_{ik} + S_1/\psi^2_{ik} - S_2/\psi_{ik} + \sum_{i\in N_k} r_{ijk}\ln p_{ijk} + (1-r_{ijk})\ln(1-p_{ijk}),$$

where $S_1 = .5\sum_{jk}(z_{ijk} - \tau_{ik})^2\sigma^{-2}_{jk}$ and $S_2 = \sum_{jk}(z_{ijk} - \tau_{ik})\mu_{jk}\sigma^{-2}_{jk}$. Proposal values $\sigma^{2(t)}_{jk}$ are sampled from the inverse gamma with shape λ and scale $(\lambda-1)\sigma^{2(t-1)}_{jk}$. Here, λ is the tuning parameter that set to achieve an acceptance rate between 30 and 50 %.
4. $\ln(\delta)$ is sampled from the left-truncated normal distribution with mean $c_{i+1} - c_i$, unit variance and lower bound equal to $\ln(\max\{\mu_{jk}\} - c_{M-1})$. The conditional distribution for the hyper-parameter b_k is gamma with scale $2J_k + \epsilon$ and rate $(a-1)\sum_j \sigma^{-2}_j + 1/\epsilon$.
5. The coefficients in the missing data model, α_0 and α_1, are sampled using the standard data augmentation method by Albert and Chib (1993).

To implement the identification constraints in (4), after each block of iterations, each ψ_{ik} is divided by the country average of ψ_{ik}'s; similarly, from each τ_{ik} the country average of τ_{ik}'s is subtracted. This procedure is similar to hierarchical centering by "sweeping" and is known to improve the convergence of MCMC algorithms in weakly identified models (Robert and Casella 2004, p. 397). The convergence can be monitored using the standard tools, as, for example, Geweke (1992) or Heidelberger and Welch (1983) diagnostics.

5 Application: Ideological Ambiguity and Electoral Performance

We apply our model to the expert data by Benoit and Laver (2006). The survey was conducted in 48 countries with about 8 parties and 30 experts per country on average. The analyzed dataset contains 10,603 entries with about 9 % of missing values, 364 parties, and 1493 experts. Our goal is to investigate whether a party's ambiguity on the issue of taxation and provision of public services is related to its ideological extremism and vote-share in the last elections.

In the survey, the experts were asked to place political parties on the 20 point scale with the end-points defined as follows:

[1] Party promotes raising taxes to increase public services.
[20] Party promotes cutting public services to cut taxes.

The posterior estimates of σ from the proposed model are very different from the naive sample standard deviation, with correlation of only 36 percent. The posterior mean of the missing data mechanism parameter α_1 is 0.245 with the standard deviation of 0.014 indicating that the missingness of the data is related to the ambiguity of party positions and the uncertainty of experts. Together this serves as the evidence that (1) the sample standard deviation would yield an incorrect measure of ideological ambiguity *if the assumed data generating model is valid* and that (2) the patterns in missing data do provide additional information about the ideological ambiguity and respondent uncertainty.

Using direct measures of ideological ambiguity and voters' uncertainty, the previous literature has found that ambiguity is related to voting behavior (Alvarez 1997; Tomz and van Houweling 2009). Therefore, ideological ambiguity should also be also related to a party's electoral performance. In case the model provides correct estimates of ideological ambiguity, one should observe *a* relationship between the posterior estimates of ideological ambiguity and vote-shares of political parties. Furthermore, if the sample standard deviation $\hat{\sigma}$ is not a valid measure of ideological ambiguity (as was suggested earlier), the correlation between $\hat{\sigma}$ and the parties' electoral performance should be low.

After computing the posterior distributions of σ_{jk}'s for all parties in the dataset, the following model is estimated:

$$T(v_{jk}) = \beta_0 + \beta_1 |\mu_{jk} - \overline{\mu}| + \beta_2 \frac{1}{1+\sigma_{jk}} + \epsilon_{jk}, \qquad (17)$$

where v_{jk} is a vote-share of party j in country k, $T(\cdot)$ is a Box-Cox transformation, and $\overline{\mu}$ is the estimated empirical center of party platforms. The coefficients β_1 and β_2 represent the effect of ideological extremism and ideological precision (the inverse of the ideological ambiguity) respectively.

The model in (17) is estimated in three settings. In the first setting, I use the sample mean $\hat{\mu}$ and standard deviation $\hat{\sigma}$ in place of μ and σ in (17). In the second setting, the mean posterior estimates $\mathbb{E}(\mu|y)$ and $\mathbb{E}(\sigma|y)$ derived from the latent hierarchical model are used in place of $\hat{\mu}$ and $\hat{\sigma}$. Both of the above models do not take into account the fact that the covariates $(\hat{\mu}, \hat{\sigma})$ and $(\mathbb{E}(\mu|y), \mathbb{E}(\sigma|y))$ are only estimates that are measured with error, not fixed values. Ignoring, the presence of the measurement error in the covariates might lead to invalid inference about the regression parameters in model (17).

Book ID: 306518_1_En, Date: 2013-02-19, Proof No: 1, UNCORRECTED PROOF