A Discrete Choice Model for the Analysis of an Aggregate Voting Record

Moritz Marbach*

University of Mannheim
January 5, 2015†

Abstract

For many domestic and international institutions, scholars are unable to observe the individual roll call voting records and only have access to a list of collective adoption or rejection decisions for each proposal. This paper proposes a novel statistical model, the multivariate probit model with partial observability that enables scholars to use an 'aggregate voting record' to study the determinants of individual voting in institutions. The model is built to estimate the same effects and quantities of interest as an ordinary probit model that one might estimate if the roll call voting records for a series of proposal would be available. Taking a Bayesian perspective for inference, I derive a Gibbs sampler with a double augmentation step that makes the model estimation tractable. Using open-source software that accompanies this paper, I demonstrate the model's applicability using simulated and real data from U.S. State Supreme Court abortion decisions in the period 1980-2006 and on U.N. Security Council deployment decisions for U.N. peace operations after the Cold War.

*Ph.D. candidate, Graduate School of Economic and Social Sciences, University of Mannheim, A5,6, 68159 Mannheim, Germany; mmarbach@mail.uni-mannheim.de.

†This is work in progress. Please check for the most recent version at www.moritz-marbach.com or email me.

1 Introduction

In domestic and international institutions, political actors decide by the means of voting. The question of what determines political actors' voting decisions in particular circumstances is widely studied in political science[1]. Unfortunately, in many political institutions excluding the U.S. Congress, only parts, or worse, none of the roll call voting records are available (Zamora, 1980; Saalfeld, 1995; Hug, 2005; Haftel and Thompson, 2006). According to Hug (2005) for example, there are only 20 out of 114 legislatures worldwide where all votes are cast as roll calls and published. The outlook is no better in the case of other international and domestic institutions such as courts, central banks or intergovernmental institutions[2]. Important institutions such as the IMF, the European Court of Justice and the European Central Bank do not publish consistently their roll call voting records.

When roll call voting records are unavailable, scholars resorted to analyzing the aggregate voting record, that is, a list of which proposals passed or failed. Substituting an analysis of how political actors decide with an analysis of how the 'institution decides is problematic for three reasons.

First, if scholars possess independent variables that exhibit variation across actors and proposals, they are unable to systematically include the variation across actors in their statistical model of the aggregate voting record. In the best case, excluding this information will only make inference more uncertain.

Second, if scholars choose to aggregate their independent variables which exhibit variation across actors with some statistic (e.g. their mean), they amplify the aggregation of the data. When using the newly created surrogate variables in a model of the aggregate voting record, scholars' inference will inevitably be subject to the same concerns scholars raised with respect to ecological studies for many years (see e.g. Robinson, 1950; King, 1997; King et al., 1999).

Third, even if scholars know with certainty that actors' choices only depend on the characteristics of the proposal, their model of the aggregate voting record only allows a quantification of how the probability of the proposal passage changes given changes in the independent variables. An actor's probability to vote for a proposal, which can

[1]For American Politics e.g. Levitt (1996); Bailey and Brady (1998); Hiscox (2002); Broz (2005); Comparative Politics e.g. Hibbing and Marsh (1987); Desposato (2001); Schonhardt-Bailey (1998) and International Relations e.g. Boockmann (2003); Aspinwall (2007); Voeten (2004).

[2]In 43 legislatures votes are required to be secret. In the remaining 72 legislatures, only specific votes are roll calls (43) or where representatives can request a roll call (28). To the author's knowledge, a similar systematic study does not exist for other domestic or international institutions. However, for examples see Sabel (2006); Zamora (1980)

be substantially larger or smaller to the probability of the proposal passing, can not be computed. Consequently, the substantive effect of independent variables might be underestimated.

I propose a novel statistical model, the multivariate probit model with partial observability (partial m-probit), that allows scholars to study the determinants of actors' vote choices in the absence of roll call voting records. Using an aggregate voting record, the model is build to estimate the same effects and quantities of interest as an ordinary probit model that one might estimate if the roll call voting records for a series of proposals would be available. The model is implemented in an open-source `R`-package[3].

I construct the likelihood function of the partial m-probit building on the observation that an aggregate outcome (adoption or rejection of a proposal) is a deterministic function of actors' voting choices (a vote profile), which are probabilistic functions of observable covariates. Given a voting rule, there is a finite number of vote profiles that can generate the observed aggregate outcome. Consequently, the probability for an aggregate outcome is the sum over the probabilities of each of the vote profiles. Taking a Bayesian perspective for inference, I derive a double-augmented Gibbs sampler to sample from the posterior density without actually calculating all possible vote profiles.

My approach is different to classical ecological inference in the social science (Robinson, 1950; Goodman, 1953; King, 1997; King et al., 1999) in that I start with the assumption that the analyst observes all independent variables for all actors and that the set of individuals that vote is rather small. My approach is similar to Przeworski and Vreeland's (2002) model of international bilateral cooperation confronting the problem that the observable outcome 'non-cooperation' is the result of the unobservable decision of either both or one country not cooperating.

While the partial m-probit estimates the same effects as an ordinary probit on the roll call voting records, the inference using an aggregate voting record will come with more uncertainty, since aggregation reduces the available information for inference. Using simulated data as well as real data from U.S. State Supreme Court abortion decisions (Caldarone et al., 2009), I illustrate that the inferential uncertainty and computational costs increase primarily as the number of actors increases when analysts have to work with the aggregate voting record instead of the roll call voting records.

To demonstrate that the partial m-probit can also be fruitfully employed even if the analyst assumes that actors' choices only depend on the characteristics of the proposal, I expand an analysis by Hultman (2013) on U.N. peace operation deployment decisions. Using a probit model on the aggregate voting record, she finds that the number of

[3] `consilium` - Estimating multivariate probit models with partial observability.

civilians killed only modestly affect the decision of the U.N. Security Council to deploy peace operations. Using the partial m-probit, I expand the analysis and estimate a strong effect on a Council member's probability to vote for the deployment of a mission.

2 Multivariate Probit with Partial Observability

Let there be M political actors ($i = 1, ..., M$) and J proposals ($j = 1, ..., J$). An actor's vote is a binary random variable, $y_{ij} \in \{0, 1\}$ corresponding to actors' binary vote choice (*no* or *yes*). Crucially, the votes are not observed. Each actor's vote is governed by a vector of K observed covariates, denoted $\mathbf{x}_{ij}$. While the analyst does not observe the votes, he observes the aggregate outcome which I denote with $b_j \in \{0, 1\}$, where b_j is zero if the proposal was rejected. A generic dataset that clarifies the notation appears in table 1.

		Observed		Unobserved
Actor	Proposal	Covariates	Outcome	Vote
1	1	$\mathbf{x}_{11}$		y_{11}
2	1	$\mathbf{x}_{21}$	b_1	y_{21}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
M	1	$\mathbf{x}_{M1}$		y_{M1}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	J	$\mathbf{x}_{1J}$		y_{1J}
2	J	$\mathbf{x}_{2J}$	b_J	y_{2J}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
M	J	$\mathbf{x}_{MJ}$		y_{MJ}

Table 1: A generic dataset for an institution with M political actors, having voted on J proposals. The observed outcome (b_j) is realized given a voting rule and the (unobserved) votes (y_{ij}). For each actor-proposal combination there is a vector of observable covariates ($\mathbf{x}_{ij}$).

Notice, that if the votes had been observed, the data could be analyzed with standard discrete choice models. The aggregation of the voting record complicates matters here and it is this complication that the partial m-probit is constructed to address. My setting is different to ecological studies, since the independent variables are not aggregated and instead fully observed and b_j is binary instead of a continuous or count variable. The setting also differs to aggregate studies, where the analyst usually only observes a sample of the actors[4]. The setting I consider is one in which the values for all independent variables for all actors are available to the analyst.

2.1 Likelihood

I introduce further notation to set up the model. Let $\mathbf{X}_{\mathrm{j}}$ be an $\mathrm{M} \times \mathrm{K}$ matrix that collects all observed covariates for all M actors for each proposal j and let $\mathbf{y}_{\mathrm{j}}$ be the vector of length M collecting all votes for the corresponding proposal. I refer to this vector as vote profile. I define, $\mathbf{y}^*_{\mathrm{j}}$ to be the latent utility for M actors for a decision j. Without loss of generality I assume that actor i votes *yes*, if $\mathrm{y}^*_{\mathrm{ij}} \geq 0$. I follow the literature in assuming that the latent utility is linear function of the observable covariates with the corresponding parameter vector $\boldsymbol{\beta}$.

Let the voting rule that governs the adoption or rejection of a proposal be a q-rule (with threshold $\mathcal{R}$), such as simple majority rule[5]. Using this notation, the multivariate probit with partial observability, abbreviated as the partial m-probit, can be written as follows:

$$
\begin{aligned}
\mathbf{y}^*_{\mathrm{j}} &= \mathbf{X}_{\mathrm{j}}\boldsymbol{\beta} + \boldsymbol{\epsilon}_{\mathrm{j}} \\
\boldsymbol{\epsilon}_{\mathrm{j}} &\overset{\mathrm{iid}}{\sim} \boldsymbol{\phi}(0, \mathbf{1}) \\
\mathrm{y}_{\mathrm{ij}} &= \begin{cases} 0 & \text{if } \mathrm{y}^*_{\mathrm{ij}} < 0 \\ 1 & \text{otherwise} \end{cases}, \\
\mathrm{b}_{\mathrm{j}} &= \begin{cases} 0 & \text{if } \sum_{\mathrm{i}=1}^{\mathrm{M}} \mathrm{y}_{\mathrm{ij}} < \mathcal{R} \\ 1 & \text{otherwise} \end{cases},
\end{aligned}
\tag{1}
$$

where $\boldsymbol{\phi}(0, \mathbf{1})$ is the standard multivariate normal density. I identify the model by setting the covariance matrix to the identity matrix. In other words, for all proposals, the error terms of the actors are assumed to be uncorrelated and actors make their choices independently. I discuss this assumption in the final section.

[4]Aggregate data analysis is growing literature in Biostatistics (Prentice and Sheppard, 1995; Wakefield and Salway, 2001; Hanseuse and Wakefield, 2008), but see Glynn et al. (2008) for a social science application. Aggregate studies differ from ecological studies in two key aspects: They incorporate additional, partially available individual level data and they model the aggregate outcome based on models of individual behavior (Wakefield and Salway, 2001).

[5]I here assume that there is one and only one voting rule. I also assume that this rule is known with certainty and followed strictly. Note, that variation of the voting rule over decisions (e.g. $\mathcal{R}_{\mathrm{j}}$) or more elaborated voting rules can be easily incorporated. I use a simple q-rule to reduce notational clutter.

Multi- or k-variate probit models are usually employed to allow for correlated choices by estimating the correlation matrix from the data. Similar to the selection model for continuous outcomes popularized by Heckman (1976), bi-variate probit models as selection models, for instance, allow for correlated error terms across a sample selection and a structural equation with binary outcomes (Dubin and Rivers, 1989; Sartori, 2003). The problem the partial m-probit address is not one of correlated (sequential) choices, but of non-observability of the simultaneous choices.

The probability for observing an aggregate outcome is the product over the vote profiles' probabilities that could have realized the outcome. The probability for each of these vote profiles is the product over the individual choice probabilities which are – as in a probit model – a linear function of observed covariates and parameters. The product over all aggregate outcome probabilities yields the likelihood of the data. Next, I define the probability for one vote profile and the sets of hypothetical vote profiles that can realize a particular aggregate outcome. Using these two definitions, I state the likelihood of the data.

Using the assumption of independent choice making, the probability for observing a vote profile $\mathbf{y}_\mathrm{j}$ is the product over the individual choice probabilities on proposal j or equivalently integrating over the latent utility in each dimension on the interval that corresponds to the observed vote choice. Formally:

$$
\begin{aligned}
\mathrm{f}(\mathbf{y}_\mathrm{j}, \mathbf{X}_\mathrm{j}|\boldsymbol{\beta}) &= \int_{\mathrm{p}_{1\mathrm{j}}} \dots \int_{\mathrm{p}_{\mathrm{Mj}}} \boldsymbol{\phi}(\mathbf{y}_\mathrm{j}^*|\mathbf{X}_\mathrm{j}\boldsymbol{\beta})\mathrm{d}\mathbf{y}_\mathrm{j}^* \\
&= \boldsymbol{\Phi}_{\mathcal{P}(\mathbf{y}_\mathrm{j})}(\mathbf{X}_\mathrm{j}\boldsymbol{\beta}),
\end{aligned}
\tag{2}
$$

where $\boldsymbol{\phi}(.)$ is the M-dimensional multivariate normal density and p_{ij} is the interval that corresponds to the vote choice y_{ij} in the profile $\mathbf{y}_\mathrm{j}$, e.i. $\mathrm{p}_{\mathrm{ij}} = [0, \infty)$ if $\mathrm{y}_{\mathrm{ij}} = 1$ and $\mathrm{p}_{\mathrm{ij}} = (\text{-}\infty, 0)$ if $\mathrm{y}_{\mathrm{ij}} = 0$. To write this more compactly, I define $\mathcal{P}(\mathbf{y}_\mathrm{j})$ to be the function that generates all $\mathrm{p}_{1\mathrm{j}}, ..., \mathrm{p}_{\mathrm{Mj}}$ given $\mathbf{y}_\mathrm{j}$ and let $\boldsymbol{\Phi}_{\mathcal{P}(\mathbf{y}_\mathrm{j})}(\cdot)$ be the implied distribution function.

Let $\tilde{\mathbf{y}}$ be a hypothetical vote profile and let V^+ be the set of all hypothetical vote profiles for which $\sum_\mathrm{i} \tilde{\mathrm{y}}_\mathrm{i} \geq \mathcal{R}$ holds. In other words, this set contains all vote profiles that realize an adoption outcome ($\mathrm{b}_\mathrm{j} = 1$). Let V^- be the complement set. Both sets are always finite but potentially large.

Using these two definitions, I can write the probability for $\mathrm{b}_\mathrm{j} = 1$ and its complement as the sum over the probabilities for all hypothetical vote profiles that can realize $\mathrm{b}_\mathrm{j} = 1$ or $\mathrm{b}_\mathrm{j} = 0$. The likelihood is incidental by taking the product over all decisions and let

$\mathrm{b_j}$ in the exponent select the term that corresponds to the observed outcome. Formally, I write the likelihood of the data as:

$$\mathcal{L}(\boldsymbol{\beta}|\mathbf{X},\mathbf{b}) = \prod_{\mathrm{j}} \Bigg(\overbrace{\sum_{\tilde{\mathbf{y}} \in \mathrm{V}^+} \Big[\boldsymbol{\Phi}_{\mathcal{P}(\tilde{\mathbf{y}})}(\mathbf{X}_{\mathrm{j}}\boldsymbol{\beta}) \Big]^{\mathrm{b_j}}}^{\text{adoption probability}} \Bigg) \times \underbrace{\Bigg(\sum_{\tilde{\mathbf{y}} \in \mathrm{V}^-} \Big[\boldsymbol{\Phi}_{\mathcal{P}(\tilde{\mathbf{y}})}(\mathbf{X}_{\mathrm{j}}\boldsymbol{\beta}) \Big]^{1\text{-}\mathrm{b_j}} \Bigg)}_{\text{rejection probability}}. \tag{3}$$

2.2 Three Special Cases

The likelihood of the partial m-probit embodies three special cases. These special cases provide some intuition behind the model's inferential logic, its relationship to other applications of probit models with partial observability, as well as the circumstances in which the partial m-probit can be fruitfully employed.

First, when the analyst knows the vote profiles, the size of the sets V^+ and V^- is equal to one and contains the observed vote profile for each proposal[6]. Then, the model reduces to the multivariate probit model which is, since the covariance matrix is assumed to be the identity matrix, an ordinary probit model with $\mathrm{J} \times \mathrm{M}$ observations. Two insights follow: First, in principle, analysts can obtain the same estimates from an aggregate voting record as from the roll call voting records. Second, the partial m-probit should only be employed if analysts feel comfortable using an ordinary probit model when they also have access to the roll call voting records.

Since a model with the observed vote profiles is a special case of the partial m-probit, analysts will obtain the same coefficient estimates, provided that the number of observed proposals (J) is sufficiently large. The price analysts pay for not observing the roll call votes is increased uncertainty about the coefficient estimates. The size of this uncertainty-increase is largely a function of the size of the sets $\mathrm{V}^+/\mathrm{V}^-$ which are in turn primarily a function of the numbers of actors (M). I demonstrate this numerically in the Monte Carlo experiments (next section).

The second special case arises when $\mathrm{M} = 2$ and a unanimity voting rule. In this case the partial m-probit reduces to the static case of the bilateral cooperation model (Przeworski

[6] An interesting extension of the model would be to allow for the incorporation of partially observed votes. As already indicated in the introduction, in some international institutions, certain votes are actually observable. However, this extension remains outside of the scope of this paper.

and Vreeland, 2002) which is a special case of the bi-variate probit model with partial observability (Poirier, 1980). To see the equivalence, note that V^{+} contains only a single element, namely the unanimity vote profile. The first product term in equation 3 thus reduces to $\mathbf{\Phi}_{\mathcal{P}(\mathbf{y}_{\mathrm{j}}=\{1,1\})}(\mathbf{X}_{\mathrm{j}}\boldsymbol{\beta})$, which can also be written as $\Phi(\mathbf{x}_{\mathrm{jA}}\boldsymbol{\beta}) \times \Phi(\mathbf{x}_{\mathrm{jB}}\boldsymbol{\beta})$, where A and B are the two actors. The complement probability to this probability gives the second term in equation 3. The resulting likelihood function $\mathcal{L}(\boldsymbol{\beta}|\mathbf{X},\mathbf{b}) = \prod_{\mathrm{j}}(\Phi(\mathbf{x}_{\mathrm{jA}}\boldsymbol{\beta})\Phi(\mathbf{x}_{\mathrm{jB}}\boldsymbol{\beta}))^{\mathrm{b_j}} \times (1 - \Phi(\mathbf{x}_{\mathrm{jA}}\boldsymbol{\beta})\Phi(\mathbf{x}_{\mathrm{jB}}\boldsymbol{\beta}))^{1\text{-}\mathrm{b_j}}$ is identical to the likelihood given in Przeworski and Vreeland (2002, eq. 3).

The third special case arises when the analyst assumes that all actors are interchangeable with respect to their utility function. In this case, the function in equation 1 can be written only with variables that are constant across actors but vary over decisions, that is $\mathbf{x}_{\mathrm{j}} = \mathbf{x}_{\mathrm{1j}} ... = \mathbf{x}_{\mathrm{Mj}}$. Invoking this assumption allows the likelihood function to be rewritten as follows:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}|\mathbf{X},\mathbf{b}) = \prod_{\mathrm{j}}\Bigg[\Bigg(\sum_{\mathrm{k}=\mathcal{R}}^{\mathrm{M}}\binom{\mathrm{M}}{\mathrm{k}}\Phi(\mathbf{x}_{\mathrm{j}}\boldsymbol{\beta})^{\mathrm{k}}(1 - \Phi(\mathbf{x}_{\mathrm{j}}\boldsymbol{\beta}))^{\mathrm{M\text{-}k}}\Bigg)\Bigg]^{\mathrm{b_j}} \times \\ \Bigg[1 - \Bigg(\sum_{\mathrm{k}=\mathcal{R}}^{\mathrm{M}}\binom{\mathrm{M}}{\mathrm{k}}\Phi(\mathbf{x}_{\mathrm{j}}\boldsymbol{\beta})^{\mathrm{k}}(1 - \Phi(\mathbf{x}_{\mathrm{j}}\boldsymbol{\beta}))^{\mathrm{M\text{-}k}}\Bigg)\Bigg]^{\mathrm{b_j\text{-}1}}.\end{aligned} \tag{4}$$

Defining the function $\mathrm{B}(\eta;\mathrm{M},\mathrm{R}) = \sum_{\mathrm{k}=\mathcal{R}}^{\mathrm{M}}\binom{\mathrm{M}}{\mathrm{k}}\Phi(\eta)^{\mathrm{k}}(1 - \Phi(\eta))^{\mathrm{M\text{-}k}}$, the equation simplifies further to:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}|\mathbf{X},\mathbf{b}) = \prod_{\mathrm{j}}\Big[\mathrm{B}(\mathbf{x}_{\mathrm{j}}\boldsymbol{\beta};\mathrm{M},\mathrm{R})\Big]^{\mathrm{b_j}} \times \\ \Big[1 - \mathrm{B}(\mathbf{x}_{\mathrm{j}}\boldsymbol{\beta};\mathrm{M},\mathrm{R})\Big]^{\mathrm{b_j\text{-}1}}.\end{aligned} \tag{5}$$

The function $\mathrm{B}(\eta;\mathrm{M},\mathrm{K})$ can be interpreted as an institution-specific link functiond analog to the familiar logit or probit link functions[7] in the Generalized Linear Model framework (McCullagh and Nelder, 1989). The advantage of this analog is that analysts feeling comfortable making the interchangeability assumption with respect to the actors, only have to switch the link function in their preferred GLM fitting routine in order

[7]Link functions have to be monotonic and differentiable (McCullagh and Nelder, 1989, p. 27). The 'B-link' is a composite function of a binomial complementary cdf and normal cdf (probit). Differentiability and monotonicity with respect to η follow directly from the definition of a cumulative density function and the chain rule of differentiation.

to obtain coefficient estimates (and quantities of interest) with respect to actors' vote choices.

Only in the trivial case where $\mathrm{M} = 1$, coefficient estimates will not change when users apply the B-link instead of a probit link. In all other cases, coefficient estimates will change as they change when analysts switch from logit to probit link for instance. Importantly, predicted probabilities about the aggregate outcome will usually be similar if not equal. Simulations suggest that the differences increase when M and $\mathcal{R}$ increase.

2.3 Quantities of Interest

As with coefficients from all non-linear models, the coefficients from a partial m-probit are not straightforward to interpret since the marginal effects are not only a function of the coefficients but also of the covariates. The common strategy to ease interpretation is to calculate the predicted probabilities for a combination of chosen values for the covariates.

There are two stylized sets of predicted probabilities that are potentially of interest for analysts. First, the predicted probability to observe the aggregate outcome conditional on a combination of chosen values for the covariates for all actors, denoted by $\mathrm{P}(\mathrm{b} = 1|\tilde{\mathbf{X}})$ and second, the predicted probability of a vote conditional on a combination of chosen values for the covariates for one particular actor i, denoted by $\mathrm{P}(\mathrm{y_i} = 1|\tilde{\mathbf{x}}_\mathrm{i})$.

The calculation of the predicted probability of a vote is straightforward. The calculation of the predicted probability of the aggregate outcome is more involved, because it depends on the hypothetical vote profiles induced by a voting rule. Since the number of vote profiles is exponentially increasing with the number of actors, the practical computation can be time consuming and memory intensive. Therefore, it is useful to approximate the predicted probability using a Monte Carlo scheme.

The scheme, which I outline in the appendix, builds upon the Gibbs Sampler presented in the next section. I suggest to draw votes for all actors conditional on the corresponding chosen values for the covariates and a simulated draw from the posterior density of the coefficients. Using the simulated votes, one can then calculate the aggregate outcome using the applicable voting rule. Repeating this calculation a few hundred times and for all posterior draws characterizes the posterior density of the predicted probability of the aggregate outcome.

2.4 Double Augmented Gibbs Sampler

Bayesian inference[8] requires one to specify a prior density for the parameters (the coefficients). I follow conventions and assume that they are jointly normal with prior mean $\mathbf{b}_0$ and a diagonal covariance matrix $\mathbf{B}_0$. The posterior density is proportional to the product of the likelihood function in equation 3 and the prior density (see Gill, 2008; Jackman, 2009, for an introduction to Bayesian Statistics).

As in many Bayesian models, the posterior density cannot be marginalized analytically, which prompts me to simulate from the posterior density and then use the simulated samples to characterize the marginal posterior density of the coefficients. I adopt a Gibbs sampler for this simulation (Geman and Geman, 1984). A Gibbs sampler is a Markov Chain Monte Carlo algorithm that can be used to obtain samples from a multivariate probability density such as the posterior density when the full conditional densities are known and easy to sample.

A Gibbs sampler requires deriving the conditional probability density function (pdf) for all unknown quantities in the model. Lauritzen et al. (1990) have shown that if a joint density can be written as a directed acyclic graph (DAG), the conditional pdf of any of the DAG's nodes $(\theta_1, ...\theta_j, ...\theta_J)$ is given by:

$$\mathrm{f}(\theta_j|\theta_{\neg j}) \propto \mathrm{f}(\theta_j|\mathrm{parents}[\theta_j]) \times \prod_{\mathrm{w}\in\mathrm{chidren}[\theta_j]} \mathrm{f}(\mathrm{w}|\mathrm{parents}[\mathrm{w}]), \tag{6}$$

where $\theta_{\neg j}$ denotes all nodes in the DAG other than θ_j. The functions 'parents[q]' collects all nodes that are connected to a node q via an inward edge and the function 'children[p]' collects all nodes that are connected via an outward edge to p.

A DAG representation of the model in equation 3 appears in figure 1(a). Each node is a random variable. Rectangular nodes indicate observed variables (the data), circle nodes represent unobserved variables. An arrow indicates the dependencies between these variables and the plates indicate the J replications. The conditional pdf for $\boldsymbol{\beta}$ in figure 1(a) is not a member of a known parametric family from which samples can be easily drawn.

[8] If one wishes to proceed with frequentist inference, one might use a maximum likelihood estimator. In this case, the log-likelihood needs to be maximized with respect to the parameters. Two difficulties will arise. First, the log-likelihood is not guaranteed to be log-concave. While the multivariate standard normal cdf is log-concave, the sum of log concave functions is not necessarily log-concave (e.g. An, 1998; Boyd and Vandenberghe, 2004). This suggests that for numerical optimization the choice of starting values is important (Altman et al., 2004, p. 84). Second, the number of hypothetical vote profiles is exponentially increasing with the number of actors ($|\mathrm{V}^+ \bigcup \mathrm{V}^-| = 2^{\mathrm{M}}$). For a simple numerical optimization algorithm, any additional member implies that either the requirement memory size or the length of the computation doubles.

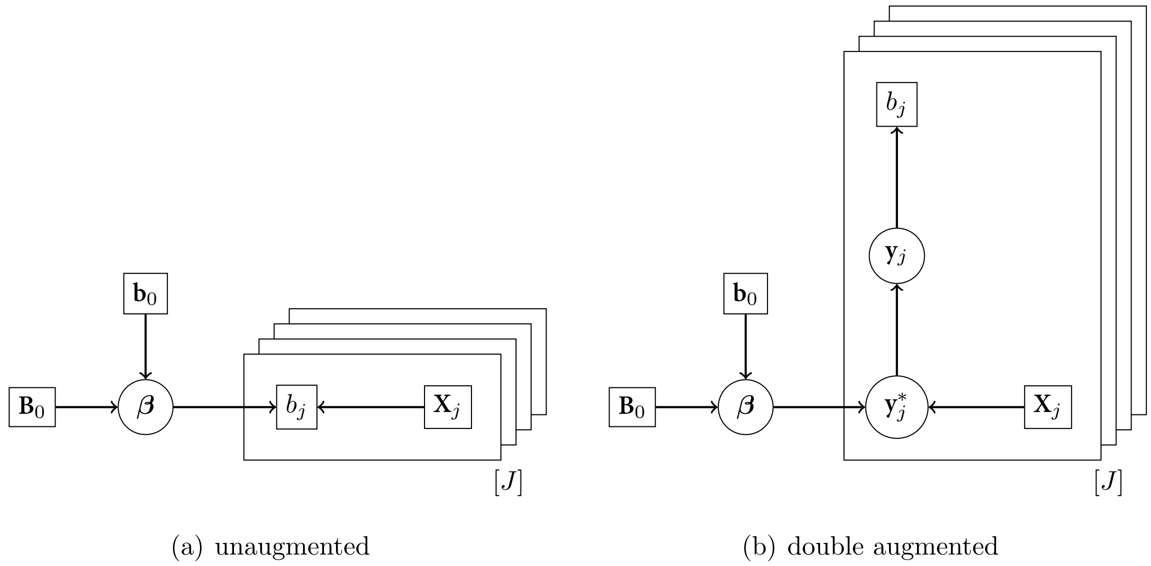

Figure 1: Two directed acyclic graphs of the partial m-probit.

In order to derive easy-to-sample-from full conditionals, I follow a data augmentation strategy (Tanner and Wong, 1987) and introduce two variables from the model setup explicitly. The augmented DAG appears in 1(b). The first augmentation is identical to the Albert-Chib augmentation in a Bayesian (multivariate) probit model (Albert and Chib, 1993; Chib and Greenberg, 1998) introducing $\mathbf{y}_\mathrm{j}^*$, the latent utility, explicitly in the model. The second augmentation augments the latent utility with $\mathbf{y}_\mathrm{j}$, the unobserved votes. Since the second augmentation augments the first, I refer to this Gibbs sampler as double-augmented Gibbs sampler.

Applying the result from Lauritzen et al. (1990) yields three easy-to-sample-from full conditional densities for the three unobserved variables in the DAG (see appendix for their functional form). The Gibbs sampler is an iterative sampling from these densities until convergence. The Gibbs sampler is implemented in an open-source `R`-package `Consilium`[9] which accompanies this paper. The source code is written in `C++` but can be directly called from `R` (R Core Team, 2014) using the `Rcpp` package (Eddelbuettel and François, 2011).

3 Monte Carlo Experiments

I conduct 16 Monte Carlo experiments to verify that the Gibbs sampler (and its implementation) obtains samples from the posterior density and to compare inference from the partial m-probit with inference from an ordinary probit model if the roll call

[9] `consilium` - Estimating multivariate probit models with partial observability.

voting records are available. For both models, I use the same vague priors ($b_0 = 0$, $B_0 = 100$). I rely on pretests to calibrate the Gibbs sampler run-length for both the partial m-probit and the ordinary probit model[10].

For each of the 16 experimental conditions, I run 250 simulations. Across the 16 conditions, I vary the sample size (250, 500), the number of actors (5,10,50,100) and the voting rule (simple majority, $\frac{2}{3}$ super-majority). Each actor's vote choice is governed by two variables: A constant $x_{0ij} = 1$ and uniform distributed variable $x_{1ij} \sim U(-2, 2)$. The coefficients for these variables are also drawn from a uniform density with a range of $[-1, 1]$. I refer to these values informally as 'true coefficient values'. If the aggregate outcomes exhibit less than 5% of either zeros or ones, that is if there is not a minimum amount if variation in the dependent variable, I discard the simulated data and repeat the simulation.

For all conditions I record the Gelman and Rubin (1992) convergence diagnostic, the correlation between the true coefficient values and the posterior means and the coverage rate (the share of true coefficients contained in the 95% posterior interval). If the Gibbs sampler works as expected, the correlation between the true coefficient values and the posterior means should be very close to 1 and the approximate 95% of the true coefficients should be inside the 95% posterior interval.

Table 3 (appendix) summarizes the results of the 16 experiments. In general, the coverage rate and correlation are very high as expected. This suggests that the Gibbs sampler and its implementation work as expected and recover the true coefficient values. Figure 2 illustrates the results from one of the experiments (10 actors, majority rule, 500 proposals). Each of the two scatter plots shows the true coefficient value plotted against the posterior mean estimate along with the 95% posterior interval. The left panel shows the intercept, the right panel the slope coefficient. The red circles indicate the posterior means for which the Gelman and Rubin (1992) convergence diagnostic does not support my choice of run length.

In figure 2, the smallest simulated intercept coefficient is much larger than the bound of the uniform distribution from which the coefficients have been simulated. This difference is a consequence of my choice to only estimate the partial m-probit if the aggregate outcome exhibits a minimum amount of variation. While my 5%-cutoff was arbitrary, the effect reveals a general subtle point: Aggregation reduces information

[10]For the probit model, I use the MCMCpack probit model (Martin et al., 2011). I obtain a posterior sample of 2,000 values by running the Gibbs sampler for 4,500 iterations thinning the chain every 2^{th} draw and discarding the first 500 draws as burn-in. For the *partial m-probit*, I use the `Consilium`-package to obtain a posterior of 2,000 values. I run the Gibbs sampler for 40,500 iterations, discarding the first 500 iterations as burn-in and thinned the chain for every 20^{th} draw. For both models, I run two chains sequentially using distinct seeds and starting values.

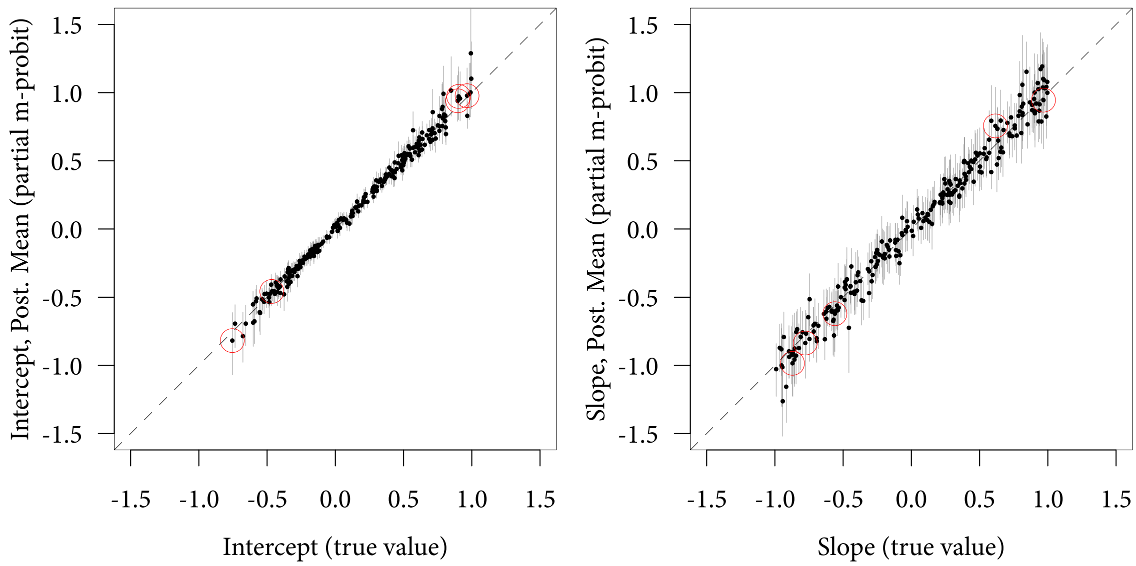

Figure 2: Result from one of the Monte Carlo experiments (10 actors, majority rule, 500 proposals): Scatter plot of posterior means with 95% posterior intervals from the partial m-probit and true coefficient values. The red circles indicate the parameters for which the Gelman and Rubin (1992) convergence diagnostic does not support my choice of run length (U. - PSRF > 1.1) and for which the chain should have been run longer. The dashed line indicates the 45-degree line coinciding with a fitted linear regression line.

potentially up to a point where there is not variation in the aggregate outcome. In other words, even if the aggregate voting record is available, it must exhibit variation in order to apply the partial m-probit. In institutions where all actors have a high average probability to vote one way or the other, there is a chance that the aggregate voting record exhibits no variation and the partial m-probit can not reveal anything to the analyst.

As an illustration, figure 3 shows the trace plot of the first 2500 iterations from one of the simulations. The posterior mean is converging to the true value as expected (upper panel). The mixing of the chains is satisfactory as indicated by decreasing auto-correlation in the ACF plots (lower panel). Table 3 (appendix) shows also the approximate computation time used for one simulation in each of the 16 experimental conditions and the number of converged simulations. Generally, computation time increases with the number of actors and sample size. While all models require more time than an ordinary probit model, even for large institutions (100 actors), computational time is still acceptable (1.30h). From the limited simulations, it appears that convergence speed of the Gibbs sampler depends on the number of actors and the voting rule.

Across the simulations the 95% posterior intervals from the partial m-probit are considerably larger than the probit intervals. Table 4 (appendix) summarizes the

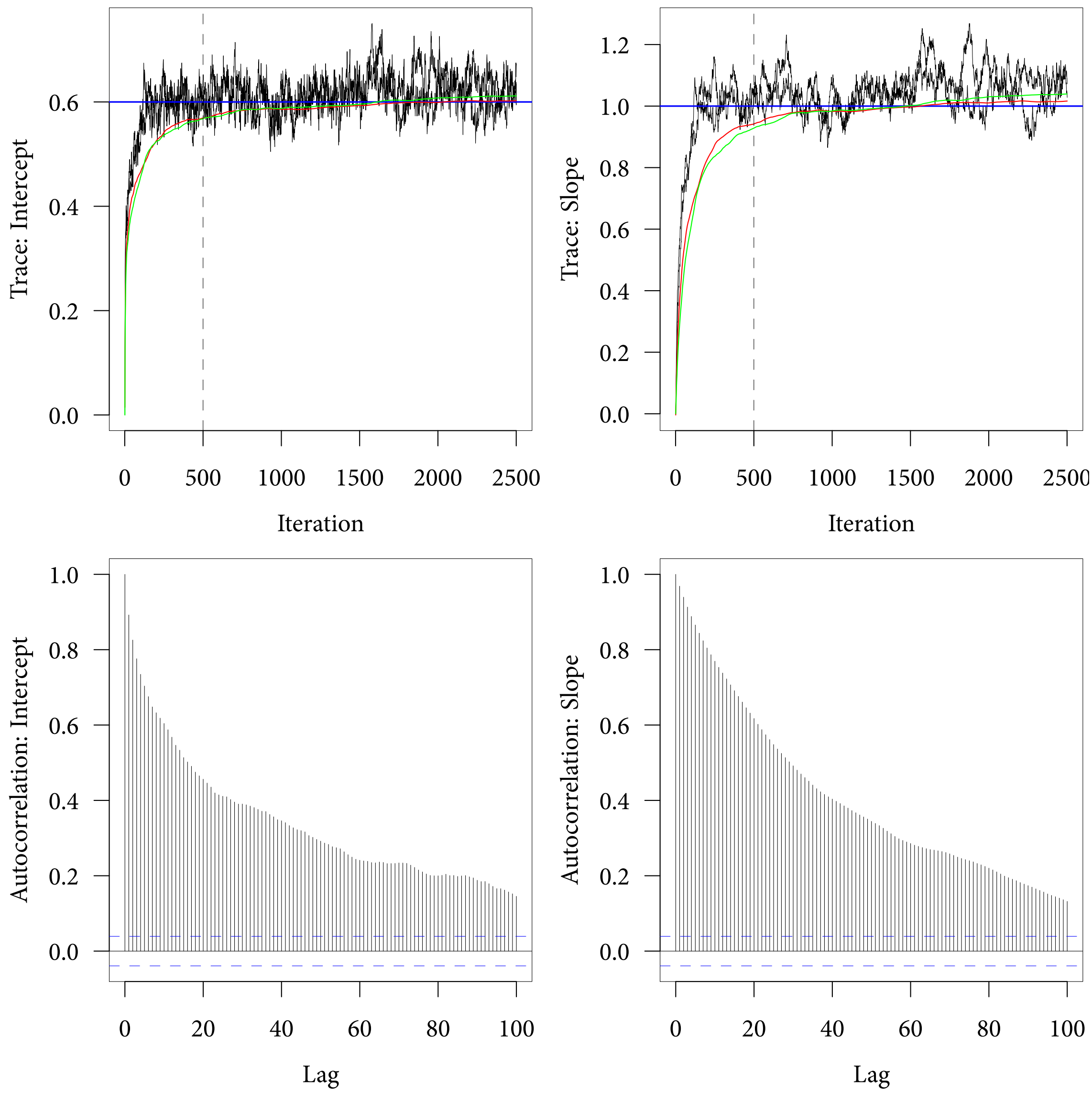

Figure 3: Trace plots (upper level) and ACF plots (lower panel) for one simulated parameter set: Gibbs sampling trace plots from 2 chains for one Monte Carlo simulation (10 members, majority rule, 500 decisions). The true parameters (blue line) are: intercept=0.6 and slope=1. The red and green lines are the running posterior means for the two chains, and the dashed line marks the burn-in period of 500 samples.

median range of the 95% posterior intervals for the partial m-probit and an ordinary probit model in each of the 16 experiments. The relative differences of the slope intervals across the two models are primarily a function of the numbers of actors, where it is also a function of the voting rule for the intercept. For the slope interval, the relative difference for the intervals is decreasing from 34% (for 5 actors) to 7% (for 100 actors). This demonstrates the primary price of aggregation an analyst has to pay in addition to the computational costs: larger inferential uncertainty. This corresponds to what has been observed for the bi-variate probit model with partial observability:

"We would not be surprised to find, in a typical application, t-ratios to be from two to four times as large under full observability as under partial observability" (Meng and Schmidt, 1985, p.83).

4 Applications

In this section, I re-analyze two published studies applying the partial m-probit. In the first part, I re-analyze a study on U.S. State Supreme Court voting. I contrast the coefficient estimates from the authors' probit model with the estimates from a partial m-probit when I artificially delete the roll call votes from the dataset and only retain the aggregate outcome. As in the Monte Carlo experiments, I show that inference comes with greater uncertainty when analysts only posses an aggregate voting record instead of the roll call voting records.

In the second part, I re-analyze a study on U.N. Security Council's decision to deploy U.N. peace operations. Voting records for the U.N. Security Council are not systematically available[11]. Different to the Supreme Court study, this study focuses not on modeling vote choices but rather the aggregate outcome. I then show that applying the partial m-probit (more precisely, the special case of a discrete choice model with a 'B-link') allows one to gain a richer set of empirical insights. In particular, I show the predicted probability to approve a peace operation (which is different to the probability of deployment) and investigate the effect of various voting rules on the probability of deployment.

4.1 U.S. State Supreme Courts

Caldarone et al. (2009) test the prediction "that nonpartisan elections increase the incentives of judges to cater to voters' ideological leanings." (p. 563). The rationale for this argument is informational. In nonpartisan elections (compared to partisan elections) voters have much less a priori information about judges' ideological positions,

[11] To be clear: The decision to deploy a U.N. peace operation usually requires a resolution of the U.N. Security Council which mandates the U.N. Secretary General to conduct this operation (Houck, 1993; Jonah, 1990). Resolutions can be adopted by consensus with or without vote (Bailey and Daws, 1998, p. 259). While most resolutions on peace operations appear to be voted upon, the U.N. Security Council conveys "in public only to adopt resolutions already agreed upon" (Cryer, 1996, p. 518). "By the time the resolutions come to a vote, it is usually known by all how much support there will be for each" (Luard, 1994, p. 19). In this manner, most resolutions are adopted unanimously (Bailey and Daws, 1998, p. 264) and in this sense the available records are not systematically available but rather constitute a selective sample. Due to the space limit, I leave the extension of incorporating such partially available roll call voting records into the partial m-probit to future research.

because judges are not associated with any particular party. Consequently, voters rely more heavily on the information about the judges' ideological positions transmitted by a judges' decisions in a nonpartisan election than a in a partisan election. Anticipating this, judges in nonpartisan elections have higher incentives to use their decisions to signal their ideological positions than their counterparts in partisan elections.

In order to test their prediction, the authors assemble a dataset of State Supreme Court decisions on abortion for the period 1980-2006. They collect these data for all State Supreme Courts whose judges face contested, statewide elections. Their dataset contains 19 State Supreme Courts and a total of 85 abortion decisions. Since State Supreme Court sizes vary (in their dataset between 5 and 9) and the authors use case-wise deletion to handle missing data, the total number of observed vote is 605 in their baseline specification.

The dependent variable in the authors' analysis is a regular justice's vote. Using state-level opinion data, the authors code each justice's vote as either popular (if it leans with the state's public opinion) or unpopular. Consequently, the dependent variable takes a 1 if either the justice votes 'pro-choice' and the state leans also 'pro-choice' or if he votes 'pro-life' and the state also leans 'pro-life' (Caldarone et al., 2009, 565). In the authors' dataset, 261 votes are popular (43%). The author's variable of interest is a binary variable indicating if a Supreme Court Justice is elected in nonpartisan elections. From the 85 abortion decisions, 39 have been made in a partisan electoral environment (46%).

Using various controls, the authors find evidence for their hypothesis. A replication of the authors' baseline specification (Model 1 in their table) using a Bayesian Probit Model with vague normal priors centered at 0 with a variance of 10[12] appears in the coefficient plot in figure 4 (the upper row of each coefficient) and in the regression table 5 (appendix). Each variable's estimated coefficient is displayed as a dot and the 95% and 68% posterior intervals are indicated (in two shades of gray). The simulated posterior probability that there is a positive effect of nonpartisan elections is 1.

Applying the partial m-probit, I drop all votes from the authors' dataset and only retained a binary variable indicating if the Court passed a popular decision by majority rule. Dropping all votes leaves me with 36 popular rulings (42%). In essence, dropping all votes reduces the number of observations for the left-hand side of the regression equation to 85, while it leaves the observations on the right-hand side unaffected

[12] A smaller prior variance would be well justified in this case. I used the MCMCpack probit model (Martin et al., 2011) to obtain a posterior sample of 2,000 values by running the Gibbs sampler for 11,000 iterations thinning the chain every 5$^{\text{th}}$ draw and discarding the first 1,000 draws as burn-in. I run two chains sequentially using distinct seeds and starting values. The Gelman-Rubin (Gelman and Rubin, 1992) convergence diagnostic supports the choice of the run length.

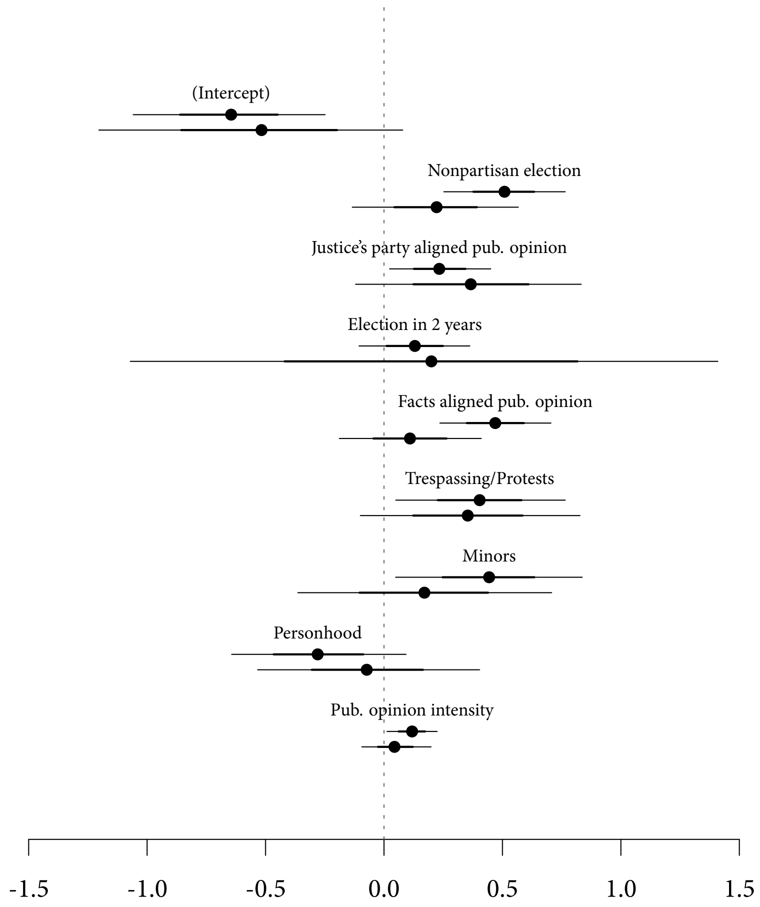

Figure 4: Regression results: Bayesian probit model with justices' voting record, N=605 (upper row) and Bayesian partial m-probit model with an aggregate voting record, N=85 (lower row). The dots indicates the posterior mean, the 95% and 68% posterior intervals are displayed in two shades of gray.

(N=605). In order to estimate the partial m-probit[13], I use the same vague normal priors as above.

The results are displayed in the coefficient plot, figure 4 (lower row). For the main variable of interest, nonpartisan election, the posterior probability that there is a positive effect of nonpartisan elections, is still 0.9 - despite the sharp decrease in available information on the left-hand side of the regression equation. The estimated effects for the two controls, which exhibit within-case variance, are notable. The effect of elections in two years is estimated with the same posterior mean but with considerably larger posterior uncertainty. The effect of justice's party aligned with public opinion is estimated a little larger and also with more posterior uncertainty.

[13]I obtain a posterior sample of 2,000 values by running the Gibbs sampler for 86,000 iterations thinning the chain every 40$^{\text{th}}$ draw and discarding the first 6,000 draws as burn-in. I run two chains sequentially using distinct seeds and starting values. The Gelman-Rubin Gelman and Rubin (1992) convergence diagnostic supports the choice of the run length.

This demonstrates again the primary price analysts pay when using an aggregate voting record: larger posterior uncertainty.

It is also interesting to compare the predictive power of the probit and partial m-probit model. One useful statistic for this is the posterior predicted probability of a popular vote. Using a cutoff of 0.5, I find that the authors' model with the roll call voting records predicts on average 371 votes correctly (61.3%). The 95% posterior interval for this estimate is [59%-64%]. Since the estimated posterior means of the coefficients are very similar, it is not surprising that the partial m-probit can be said to preform equally as the authors' model with the roll call voting records. The partial m-probit predicts on average 379 votes correctly (62.6%). However, the posterior interval for this estimate is larger reflecting the larger variance in the coefficient posteriors [54%-63%].

4.2 United Nations Security Council

Lisa Hultman studies if the U.N. Security Council is more likely to send U.N. peace operations in post-Cold War conflicts between 1989-–2006 "where civilians are deliberately targeted by the warring parties", that is into conflicts with extensive one-sided violence (Hultman, 2013, p. 59). Her dependent variable is the deployment decision in a particular conflict-year. The main independent variable is the logarithm of the number of civilians killed (one-sided violence, OSV). Using a logit model and conditional on several other controls, she finds a positive effect between one-sided violence and the probability for a U.N. peace operation deployment.

The first column (Model 1) in table 2 shows the replicated results from Hultman's 'Model 1' using a Bayesian probit model with vague normal priors (centered at 0 and a variance of 10)[14]. Note, that the results are qualitatively the same as in the original paper. However, due to some improvements of the data[15] and a change in the distributional assumption (Hultman uses a logistic regression), they are not numerically

[14]I obtain the result using the MCMCpack probit model (Martin et al., 2011) with the same priors and running the Gibbs sampler for 11,000 iterations. I discard the first 1000 iterations as burn-in. I run two chains sequentially using distinct seeds and starting values. The Gelman-Rubin (Gelman and Rubin, 1992) convergence diagnostic supports my choice of run length.

[15]In order to make the estimation results numerically more stable, I rescale the variable 'Year' and 'Polity' and used the logarithm for the variables 'Army Size' and 'Mountains'. I also treat observations in the 'Army Size' variable that had a missing value code ('-9') not as actual values, but missing values. This leads to a decrease in sample size by 13 observations. I follow Hultman and used list-wise deletion for handling missing data. Hultman uses robust clustered standard errors to adjust the co-variance matrix post-estimation for unobserved heterogeneity. Multiple imputation (Little and Rubin, 2002; King et al., 2001) and random effects (Gelman and Hill, 2006; King and Roberts, 2012) are preferable strategies to deal with these problems, but I set these issue aside to not further complicate the analysis. In the appendix, table 6, I show the numerically identical results as presented by Hultman (2013) and how they change sequentially when changing the distributional assumption and re-scale the variables.

identical. The second column presents the results from the partial m-probit[16] using the same data and the U.N. Security Council's voting rule, which is not a simple q-rule but a veto majority rule. Specifically, nine out of 15 member states have to approve a proposal, but each of the five permanent members (China, France, Russia, the United Kingdom and the United States) has a veto. I use the same priors as for the Bayesian probit model.

While the signs of the coefficients across model 1 and 2 are identical, their interpretation differs: A positive (negative) signed coefficient in model 1 indicates that the probability of a peace operation increases (decreases) if the corresponding predictor variable increases. In Model 2, a positive (negative) signed coefficient indicates that a member's approval probability increases (decreases) if the corresponding predictor variable increases.

As with other nonlinear models, it is useful to inspect the predicted probabilities for a better understanding of the estimated effects. Figure 5 (left panel) shows the predicted probabilities of mission deployment from model 2 for the typical case in Hultman's data and various levels of one-sided violence against civilians[17]. As expected, the correlation between these predicted probabilities and those from the ordinary probit model equals almost one (not shown).

In addition to the predicted probability of deployment, the partial m-probit permits the calculation of the predicted probability of approving a U.N. peace operation. The right panel shows the predicted probability for various levels of one-sided violence in the typical case. Since none of the variables in model 2 varies across members for any observation, the predicted probabilities to support a peace operation are identical for all Council members. Compared to the effect size on the aggregate level, which is quite small, the size of the effect on the actor-level is substantial. With rising levels of one-sided violence, the probability to approve an operation increases on average from 32% to 45%.

This substantive difference between the two types of predicted probabilities is to some extent a function of the voting rule. To illustrate this, I simulate the predicted probability of deployment given five different veto-majority rules. The lowest line in figure 6 is the same as in figure 5 (left panel). The lines above indicate the same

[16] I obtain the result by running the Gibbs sampler for 140,000 iterations. I discard the first 2000 iterations as burn-in and thinned the chain for every 70^{th} draw. I run two chains sequentially using distinct seeds and starting values. The Gelman-Rubin (Gelman and Rubin, 1992) convergence diagnostic supports my choice of run length.

[17] By typical case, I mean the single observation in the dataset that is most similar (the nearest neighbor) to the average case. In Hultman original data, the Mindanao conflict (Philippines) in 1996 is the typical case. Alternatively, one might use the observed-value approach as advocated recently by Hanmer and Ozan Kalkan (2013) but which is computationally more intense.

	Bay. probit	Bay. partial m-probit
(Intercept)	-1.70	0.24
	[-3.79; 0.42]	[-0.77; 1.25]
$\log(\text{OSV})_{\text{t-1}}$	0.08	0.04
	[-0.02; 0.19]	[-0.01; 0.09]
Year	-0.02	-0.01
	[-0.07; 0.04]	[-0.04; 0.01]
Battle $\text{Deaths}_{\text{t-1}}$	0.08	0.03
	[-0.01; 0.17]	[-0.01; 0.08]
Population	0.06	0.02
	[-0.23; 0.35]	[-0.12; 0.16]
$\text{Polity}_{\text{t-1}}$	-0.18	-0.09
	[-0.64; 0.28]	[-0.30; 0.12]
log(Army Size)	-0.31	-0.14
	[-0.55; -0.06]	[-0.27; -0.02]
log(Mountains)	0.04	0.02
	[-0.14; 0.21]	[-0.07; 0.11]
Non-U.N. Operation	0.74	0.38
	[0.21; 1.27]	[0.07; 0.67]
P5 Colony	-0.46	-0.23
	[-0.98; 0.03]	[-0.48; 0.01]
Num. obs	837	837

Table 2: Regression results: Bayesian probit model (col. 1) and Bayesian partial m-probit model (col. 2), each with posterior means and 95% posterior intervals in parentheses.

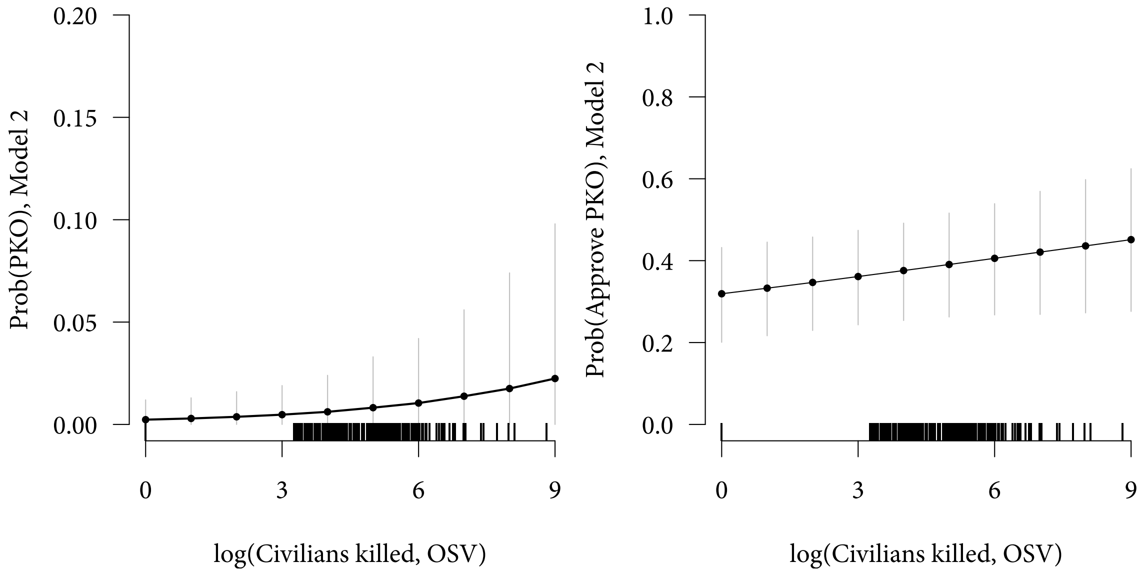

Figure 5: Predicted probability of a U.N. peace operation deployment (left) and approving a U.N. peace operation deployment (right): Posterior means and 95% posterior intervals from model 2 for various levels of one-sided violence.

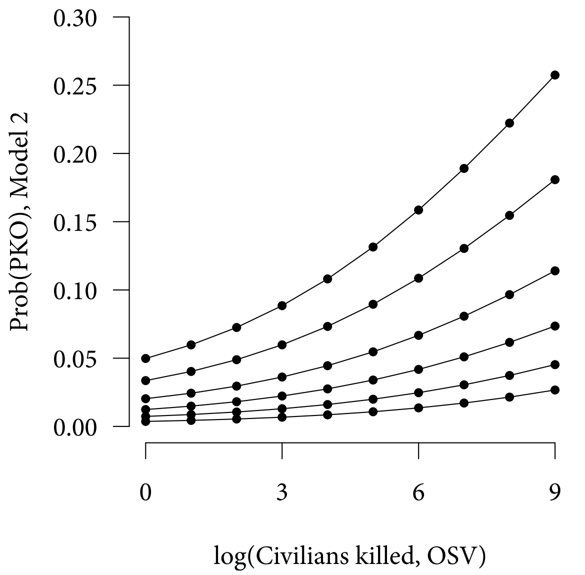

Figure 6: Predicted probability of a U.N. peace operation deployment for various levels of one-sided violence and six different veto-majority rules (lowest line: 5 vetos, highest line: 0 vetos).

probability for four or less vetos. The figure suggests that the probability of a mission is much larger the less vetos there are but also that the increase in probability for greater civilian causalities is steeper the less vetos there are. The ability to statistically quantify the effect of the veto complements formal work, showing the severe impacts of the veto in the U.N. Security Council (e.g. O'Neill, 1996; Winter, 1996).

The literature on U.N. Security Council decision-making and U.N. peacekeeping highlights the importance of diverging interests, especially among the permanent members, as important predictors for U.N. Security Council activity (e.g. Fortna, 2008; Malone, 1998; Beardsley and Schmidt, 2012). However, in the models in table 2, the members are assumed to be interchangeable. Hultman only attempts to relax this assumption by including a covariate 'P5 Colony' which captures whether or not a permanent member state was the last colonizer of the respective country.

I deliberately refrain from expanding the scope of Hultman's analysis and do not include new variables that vary across Council members. Leveraging the partial m-probit, such an analysis would be feasible and, in fact, warranted given arguments in the literature, but it remains outside the scope of this paper.

While it is necessary to assume interchangeable members when using a logit, probit or B-link model, the same assumption must not be made when using a partial m-probit. This suggests that the partial m-probit can also serve scholars by allowing them to better align their theoretical arguments with their statistical modeling while simultaneously obtaining broader empirical insights.

5 Conclusion

In many domestic and international institutions decisions are made by the means of voting. Empirical research on these decision making processes is challenging since scholars often only observe the outcome of the process (adoption or rejection of a proposal) and not the individual voting decisions. When roll call voting records are not available, studying how observable determinants are related to an actor's vote choice is difficult.

This paper introduced a statistical model, the partial m-probit, that allows one to study the determinants of voting even if one can only observe the aggregate voting record. The model is built to estimate the same effects and quantities of interest as an ordinary probit model that one might estimate if the roll call voting records for a series of proposals would be available. While the partial m-probit estimates the same effects as an ordinary probit on the roll call voting records, the inference using an aggregate voting record will come with more uncertainty (and this uncertainty is primarily increasing in the number of actors voting in the institution), since aggregation reduces the available information for inference.

I invoked two important independence assumptions in the model setup: First, I assumed that actors' vote choices are conditionally independent from each other for any proposal.

Second, I assumed conditional independence across proposals. These assumptions have to be invoked as well when analysts use an ordinary probit model to analyze the roll call or the aggregate voting record and correspond to the sincere voting assumption made in various ideal point models (e.g. Poole and Rosenthal, 1985; Clinton et al., 2004). Under what circumstances will they be reasonable?

The first assumption can be a reasonable approximation as long as the actors have not the formal right to make amendments and/or manipulate the sequence of voting on the amendments[18]. Similarly, the second assumption is a good approximation if the actors have not the ability to exchange votes across different proposals. Actors have the ability to trade votes whenever they are able to credibly commit to vote in a particular fashion on future proposals.

The validity of these assumptions could be questioned with respect to the institutions analyzed above. In the U.N. Security Council for example, there is qualitative evidence that suggests that at least the permanent members engage in vote trading across proposals. Malone, for example, reports that the US persuaded Russia to tolerate the Haiti intervention by promising to support a peacekeeping operation in Georgia (Malone, 1998, p. 107). Similar trades might also happen in U.S. State Supreme Courts. However, regular elections and term-limits might decrease the justices' ability to credibly commit.

A fruitful avenue for future research could be to identify ways to relax the independence assumptions I invoked. For example, violations of the second assumption could be mitigated by allowing for unobserved effects across proposals using a random intercept across different groupings of proposals. This would also allow analysts to account for unobservables across conflicts, countries or time periods. It might also be possible to relax the assumption of independent votes and estimate the actors' covariance matrix using an additional Metropolis-Hasting sampling step similar to the approach in Chib and Greenberg (1998).

The analysis of institutional vote choice data is a challenging endeavor in political science. Sometimes these data are only available in form of an aggregate instead of a roll call voting record. The partial m-probit is build to improve the analysis of such aggregated vote choice data but as many other models it has its limitations when it comes to dependencies between observations. Its applicability depends ultimately on the aggregate voting record a research wants to analyze.

[18] In simple voting games, for example, with n players and two alternatives, voting for the most preferred alternative is a weakly dominant strategy (e.g. Osborne, 2004, p. 48).

References

Albert, J. H. and S. Chib (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association 422*(88), 669–679.

Altman, M., J. Gill, and M. P. McDonald (2004). *Numerical Issues in Statistical Computing for the Social Scientist*. New York: Wiley.

An, M. Y. (1998). Logconcavity Versus Logconvexity: A Complete Characterization. *Journal of Economic Theory 80*(2), 350–369.

Aspinwall, M. (2007). Government Preferences on European Integration: An Empirical Test of Five Theories. *British Journal of Political Science 37*(01), 89–114.

Bailey, M. and D. W. Brady (1998). Heterogeneity and Representation: The Senate and Free Trade. *American Journal of Political Science 42*(2), 524–544.

Bailey, S. D. and S. Daws (1998). *The procedure of the UN Security Council* (3rd ed.). Oxford: Clarendon Press.

Beardsley, K. and H. Schmidt (2012). Following the Flag or Following the Charter? Examining the Determinants of UN Involvement in International Crises, 1945-2002. *International Studies Quarterly 56*(1), 33–49.

Boockmann, B. (2003). Mixed Motives: An Empirical Analysis of ILO Roll-Call Voting. *Constitutional Political Economy 14*(4), 263–285.

Boyd, S. and L. Vandenberghe (2004). *Convex Optimization*. Cambridge: Cambridge University Press.

Broz, L. J. (2005). Congressional Politics of International Financial Rescues. *American Journal of Political Science 49*(3), 479–496.

Caldarone, R. P., B. Canes-Wrone, and T. S. Clark (2009). Partisan Labels and Democratic Accountability: An Analysis of State Supreme Court Abortion Decisions. *Journal of Politics 71*(2), 560–573.

Chib, S. and E. Greenberg (1998). Analysis of Multivariate Probit Models. *Biometrika 85*(2), 347–361.

Clinton, J. D., S. Jackman, and D. Rivers (2004). The Statistical Analysis of Roll Call Data. *American Political Science Review 98*(2), 355–370.

Cryer, R. (1996). The Security Council and Article 39: A Threat to Coherence. *Journal of Conflict and Security Law 1*(2), 161–195.

Desposato, S. W. (2001). Legislative Politics in Authoritarian Brazil. *Legislative Studies Quarterly 26*(2), 287–317.

Dubin, J. A. and D. Rivers (1989). Selection Bias in Linear Regression, Logit and Probit Models. *Sociological Methods and Research 18*(2), 360–390.

Eddelbuettel, D. and R. François (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software 40*(8), 1–18.

Fortna, V. P. (2008). *Does Peacekeeping Work?* Princeton: Princeton University Press.

Gelman, A. and J. Hill (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge: Cambridge University Press.

Gelman, A. and D. B. Rubin (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science 7*(4), 457–472.

Geman, S. and D. Geman (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence 6*, 721–741.

Geweke, J. (1991). Efficient Simulation from the Multivariate Normal and Student-T Distributions Subject to Linear Constraints and the Evaluation of Constraint Probabilities. In *Computing science and statistics: proceedings of the 23rd symposium on the interface.*, Fairfax, Virginia, pp. 571–578. Inferface Foundation of North America, Inc.

Gill, J. (2008). *Bayesian Methods: A Social and Behavioral Sciences Approach* (2nd ed.). Boca Raton: Chapman and Hall.

Glynn, A., J. Wakefield, M. S. Handcock, and T. S. Richardson (2008). Alleviating Linear Ecological Bias and Optimal Design with Subsample Data. *Journal of the Royal Statistical Society. Series A 171*(1), 179–202.

Goodman, L. (1953). Ecological Regressions and Behavior of Individuals. *American Sociological Review 18*, 663–666.

Haftel, Y. Z. and A. Thompson (2006). The Independence of International Organizations. Concepts and Applications. *Journal of Conflict Resolution 50*(2), 253–275.

Hanmer, M. J. and K. Ozan Kalkan (2013). Behind the Curve: Clarifying the Best Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent Variable Models. *American Journal of Political Science 57*(1), 263–277.

Hanseuse, S. J.-P. A. and J. Wakefield (2008). The Combination of Ecological and Case–Control Data. *Journal of the Royal Statistical Society. Series B 70*(1), 73–93.

Heckman, J. J. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. *Annals of Economic and Social Measurement 5*(4), 475–492.

Hibbing, J. R. and D. Marsh (1987). Accounting for the Voting Patterns of British MPs on Free Votes. *Legislative Studies Quarterly 12*(2), 275–297.

Hiscox, M. J. (2002). Commerce, Coalitions, and Factor Mobility: Evidence from Congressional Votes on Trade Legislation. *American Political Science Review 96*(3), 593–608.

Houck, J. W. (1993). The Command and Control of United Nations Forces in the Era of "Peace Enforcement". *Duke Journal of Comparative and International Law 4*(1), 1–70.

Hug, S. (2005). Selection Effects in Roll Call Votes. Typescript, Universität Zürich.

Hultman, L. (2013). UN Peace Operations and Protection of Civilians: Cheap Talk or Norm Implementation? *Journal of Peace Research 50*(1), 59–73.

Jackman, S. (2009). *Bayesian Analysis for the Social Sciences*. New York: Wiley.

Jonah, J. O. C. (1990). The Management of Un Peacekeeping. In *The United Nations and Peacekeeping: Results, Limitations and Prospects : the Lessons of 40 Years of Experience*. Basingstoke: MacMillan.

King, G. (1997). *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton: Princeton University Press.

King, G., J. Honaker, A. Joseph, and K. Scheve (2001). Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation. *American Political Science Review 95*(1), 49–69.

King, G. and M. Roberts (2012). How Robust Standard Errors Expose Methodological Problems They Do Not Fix. *Political Analysis* (forthcoming).

King, G., M. A. Tanner, and O. Rosen (1999). Binomial-Beta Hierarchical Models for Ecological Inference. *Sociological Methods and Research 28*, 61–90.

Lauritzen, S. L., A. P. Dawid, B. N. Larsen, and H.-G. Leimer (1990). Independence Properties of Directed Markov Fields. *Networks 20*(5), 491–505.

Levitt, S. D. (1996). How Do Senators Vote? Disentangling the Role of Voter Preferences, Party Affiliation, and Senator Ideology. *American Economic Review 86*(3), 425–441.

Little, R. J. A. and D. B. Rubin (2002). *Statistical Analysis with Missing Data*. New York: Wiley.

Luard, E. (1994). *The United Nations: How it Works and What it Does* (2nd ed.), Volume 2nd. London: Maxmillan Press.

Malone, D. M. (1998). *Decision-Making in the UN Security Council: The Case of Haiti*. Oxford: Oxford University Press.

Martin, A. D., K. M. Quinn, and J. H. Park (2011). MCMCpack: Markov Chain Monte Carlo in R. *Journal of Statistical Software 42*(9), 1–21.

McCullagh, P. and J. A. Nelder (1989). *Generalized Linear Models* (2nd ed.). New York: Chapman and Hall.

Meng, C.-L. and P. Schmidt (1985). On the Cost of Partial Observability in the Bivariate Probit Model. *International Economic Review 26*(1), 71–85.

O'Neill, B. (1996). Power and Satisfaction in the United Nations Security Council. *Journal of Conflict Resolution 40*(2), 219–237.

Osborne, M. J. (2004). *An Introduction of Game Theory*. Oxford: Oxford University Press.

Poirier, D. J. (1980). Partial Observability in Bivariate Probit Models. *Journal of Econometrics 12*(2), 209–217.

Poole, K. T. and H. Rosenthal (1985). A Spatial Model for Legislative Roll Call Analysis. *American Journal of Political Science 29*(2), 357–384.

Prentice, R. L. and L. Sheppard (1995). Aggregate Data Studies of Disease Risk Factors. *Biometrika 82*(1), 113–25.

Przeworski, A. and J. R. Vreeland (2002). A Statistical Model of Bilateral Cooperation. *Political Analysis 10*(2), 101–112.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Robert, C. P. and G. Casella (2004). *Monte Carlo Statistical Methods* (2nd ed.). New York: Springer.

Robinson, W. S. (1950). Ecological Correlations and the Behavior of Individuals. *American Sociological Review 15*, 351–357.

Saalfeld, T. (1995). On Dogs and Whips: Recorded Votes. In H. Döring (Ed.), *Parliaments and Majority Rule in Western Europe*, pp. 528–65. New York: St Martin's Press.

Sabel, R. (2006). *Procedure at International Conferences: A Study of the Rules of Procedure of Conferences and Assemblies of International Inter-governmental Organisations* (2nd ed.). New York: Cambridge University Press.

Sartori, A. E. (2003). An Estimator for Some Binary-Outcome Selection Models without Exclusion Restrictions. *Political Analysis 11*(2), 111–138.

Schonhardt-Bailey, C. (1998). Parties and interests in the 'Marriage of Iron and Rye'. *British Journal of Political Science 28*, 291–332.

Tanner, M. A. and W. H. Wong (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association 82*(398), 528–540.

Voeten, E. (2004). Resisting the Lonely Superpower: Responses of States in the United Nations to U.S. Dominance. *The Journal of Politics 66*(3), 729–754.

Wakefield, J. and R. Salway (2001). A Statistical Framework for Ecological and Aggregate Studies. *Journal of the Royal Statistical Society. Series A 164*(1), 119–137.

Winter, E. (1996). Voting and Vetoing. *American Political Science Review 90*(4), 813–823.

Zamora, S. (1980). Voting in International Economic Organizations. *American Journal of International Law 74*(3), 566–608.

6 Appendix

Full Conditionals

A) The full conditional for $\boldsymbol{\beta}$ is a product of a normal prior density and the likelihood of J multivariate normal densities. Sampling is standard.

$$
\begin{aligned}
\mathrm{f}(\boldsymbol{\beta}|\mathbf{b}_0, \mathbf{B}_0, \mathbf{y}^*, \mathbf{y}, \mathbf{b}, \mathbf{X}) &\propto \ \mathrm{f}(\boldsymbol{\beta}|\mathbf{b}_0, \mathbf{B}_0) \times \prod_{\mathrm{j}} \mathrm{f}(\mathbf{y}^*_{\mathrm{j}}|\mathbf{X}_{\mathrm{j}}, \boldsymbol{\beta}) \\
&= \boldsymbol{\phi}\Big((\mathbf{B}_0^{\text{-}1} + \mathbf{X}'\mathbf{X})^{\text{-}1}(\mathbf{B}_0^{\text{-}1}\mathbf{b}_0 + \mathbf{X}'\mathbf{y}^*), (\mathbf{B}_0^{\text{-}1} + \mathbf{X}'\mathbf{X})^{\text{-}1}\Big).
\end{aligned} \tag{7}
$$

B) The full conditional for $\mathbf{y}^*_{\mathrm{j}}$ is a truncated multivariate normal. Since the components are uncorrelated (the covariance matrix is the identity matrix by assumption), sampling can be conducted component-wise using the standard algorithm from Geweke (1991).

$$
\begin{aligned}
\mathrm{f}(\mathbf{y}^*_{\mathrm{j}}|\mathbf{b}_0, \mathbf{B}_0, \boldsymbol{\beta}, \mathbf{y}, \mathbf{b}, \mathbf{X}) &\propto \ \mathrm{f}(\mathbf{y}^*_{\mathrm{j}}|\mathbf{X}_{\mathrm{j}}, \boldsymbol{\beta}) \times \mathrm{f}(\mathbf{y}_{\mathrm{j}}|\mathbf{y}^*_{\mathrm{j}}) \\
&\propto \ \mathrm{f}(\mathbf{y}^*_{\mathrm{j}}|\mathbf{X}_{\mathrm{j}}, \boldsymbol{\beta}, \mathbf{y}_{\mathrm{j}}) \\
&\propto \ \boldsymbol{\phi}(\mathbf{X}_{\mathrm{j}}\boldsymbol{\beta}) \prod_{\mathrm{i}} \Big(\mathcal{I}(\mathrm{y}^*_{\mathrm{ij}} \geq 0)\mathcal{I}(\mathrm{y}_{\mathrm{ij}} = 1) + \mathcal{I}(\mathrm{y}^*_{\mathrm{ij}} < 0)\mathcal{I}(\mathrm{y}_{\mathrm{ij}} = 0)\Big).
\end{aligned} \tag{8}
$$

C) The conditional density for $\mathbf{y}_{\mathrm{j}}$ is a set of Bernoulli densities with the constraint that their sum is consistent with the observed b_{j}.

$$
\begin{aligned}
\mathrm{f}(\mathbf{y}_{\mathrm{j}}|\mathbf{b}_0, \mathbf{B}_0, \boldsymbol{\beta}, \mathbf{y}, \mathbf{b}, \mathbf{X}) &\propto \mathrm{f}(\mathbf{y}_{\mathrm{j}}|\mathbf{y}^*_{\mathrm{j}}) \times \mathrm{f}(\mathrm{b}_{\mathrm{j}}|\mathbf{y}_{\mathrm{j}}) \\
&\propto \mathrm{f}(\mathbf{y}_{\mathrm{i}}|\mathbf{y}^*_{\mathrm{j}}, \mathrm{b}_{\mathrm{j}}) \\
&\propto \prod_{\mathrm{i}} \Big(\Phi(\mathrm{y}^*_{\mathrm{ij}})^{\mathrm{y}_{\mathrm{ij}}} + (1 \text{ - } \Phi(\mathrm{y}^*_{\mathrm{ij}}))^{1\text{-}\mathrm{y}_{\mathrm{ij}}}\Big) \times \\
&\quad \Big(\mathcal{I}(\sum_{\mathrm{i}} \mathrm{y}_{\mathrm{ij}} < \mathcal{R})\mathcal{I}(\mathrm{b}_{\mathrm{j}} = 0) + \mathcal{I}(\sum_{\mathrm{i}} \mathrm{y}_{\mathrm{ij}} \geq \mathcal{R})\mathcal{I}(\mathrm{b}_{\mathrm{j}} = 1)\Big).
\end{aligned} \tag{9}
$$

To sample from this target density, it is useful to use accept-reject sampling (e.g. Robert and Casella, 2004, p. 51). A simple version of an algorithm takes $\mathrm{f}(\mathbf{y}_{\mathrm{j}}|\mathbf{y}^*_{\mathrm{j}})$ as the proposal density:

1. Draw from

$$\mathbf{y}_j \sim \begin{cases} f(y_{1j}|y^*_{1j}) \\ \vdots \\ f(y_{Mj}|y^*_{Mj}) \end{cases} \tag{10}$$

2. Draw $u \sim U(0,1)$

3. Accept $\mathbf{y}_j$ if $u \leq \frac{f(\mathbf{y}_j|\mathbf{y}^*_j)\times f(b_j|\mathbf{y}_j)}{C\times f(\mathbf{y}_j|\mathbf{y}^*_j)}$ otherwise repeat.

where C is a choosen constant s.t. $C \geq 1$ absorbing the normalizing constant of the target density and $U(0,1)$ is the uniform density on the interval $[0,1]$. Note that:

$$\frac{f(\mathbf{y}_j|\mathbf{y}^*_j) \times f(b_j|\mathbf{y}_j)}{C \times f(\mathbf{y}_j|\mathbf{y}^*_j)} = f(b_j|\mathbf{y}_j) \tag{11}$$

if C is set to 1. Since $f(b_j|\mathbf{y}_j)$ is either 0 or 1, the acceptance ratio in the second step is either 0 or 1. Hence, partically sampling does not require to draw from a uniform but only requires to check if a proposed $\mathbf{y}_j$ obeys the constrain implied by b_j.

Gibbs Sampler

Denote the s^{th} draw with superscript (s). The Gibbs sampler appears below. Note, that the second and third step are a special case of the Gibbs sampler in Chib and Greenberg (1998) where the covariance matrix is known to be the identity matrix and does not need to be inferred from the data.

1. For all J draw element-wise a proposal for vector $\mathbf{y}_j^{(s)}$ from Bernoulli:

$$y_{ij}^{(s)} \sim \mathrm{Bern}(\Phi(\mathbf{x}_{ij}\boldsymbol{\beta}^{(s\text{-}1)}))$$

   and accept the proposed vector if $\sum_i y_{ij}^{(s)} < \mathcal{R}$ and $b_j = 0$ or $\sum_i y_{ij}^{(s)} > \mathcal{R}$ if $b_j = 1$. Otherwise repeat until acceptance.

2. Draw for all $j = 1, ..., J$ and $i = 1, ..., M$ from Truncated Normals:

$$y_{ij}^{*(s)} \sim \begin{cases} \phi(\mathbf{x}_{ij}\boldsymbol{\beta}^{(s\text{-}1)})\mathcal{I}(y_{ij}^{*(s)} \geq 0) & \text{if } y_{ij}^{(s)} = 1 \\ \phi(\mathbf{x}_{ij}\boldsymbol{\beta}^{(s\text{-}1)})\mathcal{I}(y_{ij}^{*(s)} < 0) & \text{if } y_{ij}^{(s)} = 0 \end{cases}$$

3. Draw from a Multivariate Normal:

$$\boldsymbol{\beta}^{(s)} \sim \boldsymbol{\phi}(\mathbf{b}_0, \mathbf{B}_0)$$

$$\begin{aligned} \mathbf{b}_0 &= (\mathbf{B}_0^{-1} + \mathbf{X}'\mathbf{X})^{-1}(\mathbf{B}_0^{-1}\mathbf{b}_0 + \mathbf{X}'\mathbf{y}^{*(s)}) \\ \mathbf{B}_0 &= (\mathbf{B}_0^{-1} + \mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

   with $\mathbf{X}$ and $\mathbf{y}^{*(s)}$ ordered correspondingly.

4. Repeat S times until convergence.

Predicted Probability Calculation

The following Monte Carlo scheme approximates the predicted probability for the aggregate outcome $\mathrm{P}(\mathrm{b}=1|\tilde{\mathbf{X}})$ for one draw of coefficients from the posterior density ($\boldsymbol{\beta}^{(\mathrm{s})}$):

1. For each actor $\mathrm{i}, \mathrm{i}=1,...,\mathrm{M}$ and posterior draw s, generate a value $\mathrm{y}_\mathrm{i}^{(\mathrm{t})}$ from:

$$\mathrm{y}_\mathrm{i}^{(\mathrm{t})} = \mathrm{Bern}(\Phi(\tilde{\mathbf{x}}_\mathrm{i}\boldsymbol{\beta}^{(\mathrm{s})}))$$

2. Compute $\mathrm{b}^{(\mathrm{t})}$ using:

$$\mathrm{b}^{(\mathrm{t})} = \begin{cases} 0 & \text{if } \sum_{\mathrm{i}=1}^{\mathrm{M}} \mathrm{y}_\mathrm{i}^{(\mathrm{t})} < \mathcal{R} \\ 1 & \text{otherwise} \end{cases}$$

3. Repeat T times.

After obtaining T values for b, averages them to obtain an estimate for the predicted probability for posterior draw $\boldsymbol{\beta}^{(\mathrm{s})}$. Repeating the algorithm above for each posterior draw $\mathrm{s}, \mathrm{s}=1,...,\mathrm{S}$ yields the posterior density of the predicted probability that can then be summarized in any desired way.

					Intercept			Slope		
M	$\mathcal{R}$	J	Sim.	Time	Conv.	Corr.	Cover.	Conv.	Corr.	Cover.
5	3	500	250	5.5	250 (1.00)	1.00	236 (0.94)	250 (1.00)	0.99	238 (0.95)
5	4	500	250	5.4	250 (1.00)	0.99	241 (0.96)	249 (1.00)	0.99	233 (0.94)
10	6	500	250	10.3	245 (0.98)	0.99	235 (0.96)	245 (0.98)	0.99	234 (0.96)
10	7	500	250	10.2	248 (0.99)	0.99	235 (0.95)	246 (0.98)	0.99	231 (0.94)
50	26	500	250	45.4	225 (0.90)	1.00	217 (0.96)	194 (0.78)	0.99	181 (0.93)
50	33	500	250	46.5	202 (0.81)	0.99	191 (0.95)	182 (0.73)	0.99	173 (0.95)
100	51	500	250	89.5	213 (0.85)	1.00	199 (0.93)	145 (0.58)	0.99	138 (0.95)
100	67	500	250	91.2	168 (0.67)	0.98	158 (0.94)	152 (0.61)	0.99	142 (0.93)
5	3	250	250	3.8	250 (1.00)	0.99	241 (0.96)	250 (1.00)	0.98	238 (0.95)
5	4	250	250	3.2	250 (1.00)	0.99	233 (0.93)	250 (1.00)	0.98	241 (0.96)
10	6	250	250	5.5	245 (0.98)	0.98	219 (0.89)	245 (0.98)	0.97	222 (0.91)
10	7	250	250	5.3	249 (1.00)	0.99	236 (0.95)	247 (0.99)	0.98	238 (0.96)
50	26	250	250	25.0	218 (0.87)	0.99	211 (0.97)	186 (0.74)	0.97	175 (0.94)
50	33	250	250	23.3	199 (0.80)	0.98	189 (0.95)	184 (0.74)	0.98	178 (0.97)
100	51	250	250	43.3	207 (0.83)	0.99	202 (0.98)	145 (0.58)	0.98	138 (0.95)
100	67	250	250	49.1	148 (0.59)	0.95	137 (0.93)	127 (0.51)	0.97	121 (0.95)

Table 3: Results from 16 Monte Carlo Experiments: The number of political actors (M), the voting rule ($\mathcal{R}$), the number of proposals (J), the number of simulations per experiment (Sim.), approx. computation time in minutes (time) for one simulation with an Intel Xeon CPU, 2.8GHz, the number and percentage shares of simulations for which the Gelman and Rubin (1992) convergence diagnostic supports my choice of run length (U. - PSRF > 1.1), for all converged simulations the correlations between true and estimated coefficients and the number as well as shares of simulations for which the true coefficients were inside the 95% posterior intervals.

				Posterior Interval Range					
				Intercept			Slope		
M	$\mathcal{R}$	J	Sim.	PMP	Probit	%	PMP	Probit	%
5	3	500	250	0.16	0.11	68.2	0.30	0.10	33.8
5	4	500	250	0.17	0.11	64.7	0.30	0.10	33.6
10	6	500	250	0.12	0.08	62.5	0.31	0.07	22.3
10	7	500	250	0.13	0.08	58.9	0.32	0.07	22.2
50	26	500	250	0.06	0.03	58.9	0.32	0.03	09.6
50	33	500	250	0.09	0.03	39.0	0.33	0.03	09.7
100	51	500	250	0.04	0.02	64.0	0.31	0.02	07.1
100	67	500	250	0.08	0.03	31.0	0.32	0.02	07.0
5	3	250	250	0.24	0.16	67.8	0.43	0.15	33.8
5	4	250	250	0.25	0.16	63.6	0.44	0.14	33.0
10	6	250	250	0.18	0.11	61.6	0.47	0.10	22.3
10	7	250	250	0.19	0.11	58.7	0.44	0.10	22.9
50	26	250	250	0.08	0.05	61.6	0.44	0.04	09.9
50	33	250	250	0.12	0.05	41.3	0.47	0.05	09.7
100	51	250	250	0.05	0.03	64.0	0.44	0.03	07.1
100	67	250	250	0.12	0.04	30.8	0.49	0.03	06.8

Table 4: Results from 16 Monte Carlo Experiments: The number of political actors (M), the voting rule ($\mathcal{R}$), the number of proposals (J), the number of simulations per experiment (Sim.), for all converged simulations the median range of the 95% posterior interval from the partial m-probit (PMP), the median range of the 95% posterior interval from an ordinary probit model (Probit) and the differences of the probit model posterior interval compared to the partial m-probit interval (in percent).

	Model 1	Model 2
(Intercept)	-0.65	-0.53
	[-1.06; -0.25]	[-1.20; 0.08]
Nonpartisan election	0.51	0.22
	[0.25; 0.76]	[-0.13; 0.57]
Justice's party aligned pub. opinion	0.23	0.36
	[0.02; 0.45]	[-0.12; 0.83]
Election in 2 years	0.13	0.20
	[-0.11; 0.36]	[-1.07; 1.41]
Facts aligned pub. opinion	0.47	0.11
	[0.24; 0.70]	[-0.19; 0.41]
Trespassing/Protests	0.40	0.36
	[0.05; 0.76]	[-0.10; 0.83]
Minors	0.44	0.17
	[0.05; 0.84]	[-0.36; 0.71]
Personhood	-0.28	-0.07
	[-0.64; 0.09]	[-0.53; 0.40]
Pub. opinion intensity	0.12	0.05
	[0.01; 0.22]	[-0.09; 0.20]
Num. obs	605	85

Table 5: Regression results for U.S. Supreme Court application: Bayesian probit model (col. 1) and Bayesian partial m-probit model (col. 2), each with posterior means and 95% posterior intervals in parentheses.

	Model 1	Model 2	Model 3
(Intercept)	35.39	35.27	-1.67
	(117.21)	(52.29)	(0.98)
$\text{OSV}_{\text{t-1}}$	0.16	0.07	0.08
	(0.06)	(0.03)	(0.03)
Year	-0.02	-0.02	
	(0.06)	(0.03)	
Year (scaled)			-0.01
			(0.03)
$\text{Battle Deaths}_{\text{t-1}}$	0.16	0.06	0.07
	(0.10)	(0.05)	(0.05)
Population	0.08	0.02	0.06
	(0.24)	(0.11)	(0.14)
$\text{Polity}_{\text{t-1}}$	-0.04	-0.02	
	(0.05)	(0.02)	
$\text{Polity}_{\text{t-1}}$ (scaled)			-0.17
			(0.24)
Army Size	-0.01	0.00	
	(0.00)	(0.00)	
log(Army Size)			-0.29
			(0.13)
Mountains	0.00	0.00	
	(0.01)	(0.00)	
log(Mountains)			0.03
			(0.07)
Non-U.N. Operation	1.78	0.79	0.73
	(0.52)	(0.23)	(0.25)
P5 Colony	-1.10	-0.47	-0.44
	(0.58)	(0.29)	(0.28)
Num. obs.	850	850	837

Table 6: Replication Results: Maximum likelihood estimation, logistic regression coefficients with clustered, robust standard errors based on conflict location in parentheses (Model 1), probit regression coefficients (Model 2-3). Model 1 is the model presented in Hultman (2013). Variables with a subscript $_{\text{t-1}}$ indicate lagged variables. Model 2 is the same as Model 1 but with a probit link function. Model 3 uses various transformed variables. A '(scaled)' indicates scaling based on either the smallest observed value (Year) or to the interval $[0, 2]$ ($\text{Polity}_{\text{t-1}}$). Variable names wrapped by 'log( )' are log-transformed. The sample size is reduced since values in the variable 'Army Size' that had a missing value code, are treated as missing values.