# Analyzing Decision Records from Committees

Moritz Marbach*

December 3, 2018

**Abstract**

In the absence of a complete voting record, decision records are an important data source to analyze committee decision-making in various institutions. Despite the ubiquity of decision records, we know surprisingly little about how to analyze them. This paper highlights the costs in terms of bias, inefficiency, or inestimable effects when using decision instead of voting records and introduces a Bayesian structural model for the analysis of decision-record data. I construct an exact likelihood function that can be tailored to many institutional contexts, discuss identification, and present a Gibbs sampler on the data-augmented posterior density. I illustrate the application of the model using data from US state supreme court abortion decisions and UN Security Council deployment decisions.

*Center for Comparative and International Studies, ETH Zurich, 8092 Zurich, Switzerland. moritz.marbach@gess.ethz.ch. For helpful comments and suggestions, I thank Stephen Chaudoin, Tom Clark, Songying Fang, Thomas Gschwend, Jeff Gill, Carlo Horz, Simon Hug, Thomas König, Matt Loftis, Will Lowe, Lars Mäder, Nikolay Marinov, Christoph Mikulaschek, Simon Munzert, Adam Scharf, Curtis Signorino, Tilko Swalve and Anna Wilke.

At every level of politics, from a city council meeting up to the United Nations Security Council, committees – groups of representatives – make rules, monitor and enforce compliance. While many of these committees adopt decisions by some form of voting, the absence of a complete voting record is an unfortunate but common feature of many of them. A large majority of domestic and international institutions such as courts, central banks, or intergovernmental organizations does not publish voting records consistently.[1]

While the reasons for the lack of a voting record are plentiful, the consequences for quantitative empirical research are the same: making inferences regarding how observables are related to committee members' vote choices is challenging. In a search for a means to make inferences, some studies have turned to committees' decision records. A decision record can generally be defined as a list detailing the adoption or rejection decisions of a committee as a whole. Using these data, such studies estimate the effect of observables on the probability of the committee adopting or rejecting a decision to learn about the effect of these observables on members' vote choice.

A typical example of this strategy is the literature on UN peace operations. The central puzzle in this literature is why the UN Security Council deploys UN peace operations in some conflicts but not others. One major line of inquiry is to evaluate whether UN Security Council permanent members' self-interest – captured by variables such as previous colonial relations, military alliances, and trade relationships – prevents more decisive actions by the Council [2]. Since most votes on UN peace operation deployments are unavailable, studies in this literature can not estimate the effect of (measured) self-interest on permanent members' vote choice but estimate only the effect of *average* (measured) self-interest on the Council's decision to approve or reject UN peace operations.

This paper is about how to analyze decision records and the relative costs of using decision instead of voting records. Typically, as for example in the literature on

[1]For example, acoording to an analysis of data from 93 central banks that responded to a questionnaire from the Bank of England in the 1990s, only six publish voting records (Japan, Korea, Poland, Sweden, the United Kingdom, and the United States) (Fry et al., 2000, Chart 7.3). In a survey conducted by this author of 12 international organizations mentioned in Schermers and Blokker (2011) that have nonplenary organs and operate on a global scale, eight use a show of hands (ITU, UPU, ILO, UNESCO, WHO, IMO, WMO and IAEA), two voting by assent (IMF, IBRD/IDA "World Bank") and one a secret vote (ICAO) as ordinary vote-casting procedure. Except for FAO on some issues, none uses recorded voting as an ordinary voting procedure, but the rules of procedure typically allow committee members to request a recorded vote (exceptions are the IMF, World Bank, and WMO).

[2]For example, Gilligan and Stedman (2003); Hultman (2013); Stojek and Tir (2015).

UN peace operations, the decision record is assumed to be drawn from a convenient stochastic distribution, which allows the analyst to employ a standard model for inference (e.g., a probit model). Deviating from this reduced-form approach, I introduce a Bayesian structural model deriving the exact stochastic distribution of decision record data from the vote-choice distributions that determine a decision. To arrive at the structural likelihood function of the observed data, I model each unobserved vote choice with an ordinary probit model: the choice to vote one way or the other is a function of observable variables and a vector of coefficients. However, since choices are unobserved, I integrate out the actual vote choices to arrive at a likelihood function that is a function of observables, coefficients, and the institutional context but not of the unobserved vote choices. I highlight the intimate connection between the likelihood function and the little-known Poisson's Binomial distribution (Wang, 1993) and its relationship to the bivariate probit with partial observability (Przeworski and Vreeland, 2002; Poirier, 1980) and discuss (classical) parametric identification. I derive a suitable Gibbs sampler to simulate from the exact posterior density. This Gibbs sampler and various other tools used to analyze decision records are implemented in this author's open-source R-package `consilium` that accompanies this paper.

The Bayesian structural model clarifies the main methodological challenge with decision records, incorporates additional information about the structure of the data-generating process, and has practical advantages. First, it makes the costs of (partial) aggregation transparent. As I discuss in detail, these costs include not estimable member-specific effects, an increase in posterior uncertainty, and, in some circumstances, an aggregation bias. These costs can be mitigated by including partially observed votes, which is computationally straightforward within the structural model but infeasible in a reduced-form model. Furthermore, the structural model allows the analyst to calculate vote-choice probabilities, which is also infeasible with a reduced-form model. Perhaps surprisingly, vote-choice probabilities are not linear functions of adoption probabilities. This is because adoption probabilities are conditional probabilities with respect to the institutional context, while vote-choice probabilities are unconditional probabilities. To the extent that the analyst aims to learn how observables are related to members' vote choices or intends to make comparisons across institutional contexts, the structural model is a more suitable way of analyzing decision-record data. Finally, I also show that the correct reduced-form model is not necessarily the one that is typically estimated in practice.

I conduct Monte Carlo experiments to verify that the model works as expected, and I replicate a study by Caldarone et al. (2009) on US state supreme courts to contrast the inference when a voting record is used with the inference when I artificially delete (a subset of) the recorded votes retaining the decision record. To highlight the advantages of the structural model relative to a reduced-form model, I return to the example of the UN Security Council and estimate whether a UN Security Council member is more likely to support the deployment of a UN Blue Helmet operation if it has strong trade relationships with the conflict location. I conclude this paper with a short discussion on the (types of) institutions for which one can successfully compile a decision record in the first place and then apply the partial m-probit.

# 1 Modeling Decision Records

I consider a setting with a committee of $M$ members ($i = 1, ..., M$) and $J$ decisions ($j = 1, ..., J$). A member's vote is a binary random variable, $y_{ij} \in \{0, 1\}$ corresponding to members' binary vote choices to adopt or reject a proposal (*no* or *yes*). Crucially, the votes are not observed. The vote of each member is governed by a vector of $K$ covariates (observables), denoted $\mathbf{x}_{ij}$. While the analyst does not observe the votes, he or she observes the binary outcome of the voting, 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.

Note 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 I address.

My setting is different from that of ecological studies since the covariates are not aggregated but fully observed, the dependent variable is binary instead of continuous or categorical, and the number of vote choices is much smaller. The setting also differs from aggregate studies, where the analyst usually observes only a sample of the members[3]. The setting I consider is one in which the values for all covariates for all members are available to the analyst.

[3]Aggregate data analysis is a growing body of literature in biostatistics (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 respects: they incorporate additional, partially available individual-level data, and they model the aggregate outcome based on models of individual behavior (Wakefield and Salway, 2001).

| | | Observed | | Unobserved |
|---|---|---|---|---|
| Member | Decision | Covariates | Outcome | Vote |
| 1 | 1 | $\mathbf{x}_{11}$ | $b_1$ | $y_{11}$ |
| 2 | 1 | $\mathbf{x}_{21}$ | | $y_{21}$ |
| ⋮ | ⋮ | ⋮ | | ⋮ |
| M | 1 | $\mathbf{x}_{M1}$ | | $y_{M1}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 1 | J | $\mathbf{x}_{1J}$ | $b_J$ | $y_{1J}$ |
| 2 | J | $\mathbf{x}_{2J}$ | | $y_{2J}$ |
| ⋮ | ⋮ | ⋮ | | ⋮ |
| M | J | $\mathbf{x}_{MJ}$ | | $y_{MJ}$ |

**Table 1:** A generic dataset for a committee with $M$ members, having made $J$ decisions. The observed decision outcome ($b_j$) is realized given a voting rule and the (unobserved) votes ($y_{ij}$). For each member-decision combination, there is a vector of covariates ($\mathbf{x}_{ij}$).

## 1.1 Model Statement

Let $\mathbf{X}_j$ be an $M \times K$ matrix that collects all covariates for all $M$ members for each decision $j$, and let $\mathbf{y}_j$ be the vector of length $M$ collecting all votes, $y_{1j}, ..., y_{Mj}$, for the corresponding proposal. I refer to this vector as the vote profile.[4] I define $\mathbf{y}_j^*$ as the vector of latent utilities for $M$ members to support a decision $j$. An element of this vector is the latent utility of member $i$, denoted $y_{ij}^*$. Member $i$ votes *yes* if $y_{ij}^* \geq 0$. For simplicity, I assume that the latent utility is a linear function of the 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 a majority threshold $\mathcal{R}$, such as a simple majority rule or a supermajority rule[5]. If the number of votes, i.e., $\sum_{i=1}^{M} y_{ij}$, is less than $\mathcal{R}$, the rejection decision ($b_j = 0$) is

[4]Throughout the text, I follow the convention denoting vectors and matrices in bold letters.

[5]I use a simple, constant q-rule to reduce notational clutter but, as will become clear, the model and the Gibbs sampler can handle decision-specific simple rules. Simple rules can be characterized by decisive coalitions (the set of all vote profiles that lead to the adoption of a proposal) and encompass many generally used voting rules, including weighted-majority rules (weighted q-rules) and veto-majority rules (collegial rules). However, plurality rule and Borda count are not simple rules. For a formal definition of voting rules, see Austen-Smith and Banks (1999, Chap. 3.1).

realized; otherwise, the decision to adopt is realized ($b_j = 1$). Using this notation, the model can be written as follows:

$$\begin{aligned}
&\mathbf{y}_j^* = \mathbf{X}_j\boldsymbol{\beta} + \boldsymbol{\epsilon}_j \\
&\boldsymbol{\epsilon}_j \overset{iid}{\sim} \boldsymbol{\phi}(0, \mathbf{1}) \\
&y_{ij} = \begin{cases} 0 & \text{if } y_{ij}^* < 0 \\ 1 & \text{otherwise} \end{cases}, \\
&b_j = \begin{cases} 0 & \text{if } \sum_{i=1}^{M} y_{ij} < \mathcal{R} \\ 1 & \text{otherwise} \end{cases},
\end{aligned} \tag{1.1}$$

where $\boldsymbol{\phi}(0, \mathbf{1})$ is the standard multivariate normal density. The model rests on two assumptions: 1) coefficients are shared across all committee members, and 2) vote choices are conditionally independent. The latter assumption corresponds to the familiar sincere-voting assumption made typically in ideal-point models (e.g., Poole and Rosenthal, 1985; Clinton et al., 2004). As will become clear from section 1.3, these two assumptions are necessary for classical identification of the likelihood. However, they could be relaxed if a partially observed voting record is available to the analyst (see section 4.3).

In most applications, vote choices will not be fully independent after conditioning on observables. However, this will not necessarily distort the inference as long as the correlation among vote choices is induced by unobservables that are independent of the covariate for which the analyst wants to estimate the direction of the effect (or calculate marginal effects). In this situation, the unobservables are said to be neglected heterogeneity that will only rescale the coefficient estimates in the same way that neglected heterogeneity affects probit models (e.g. Wooldridge, 2001, p. 470). I provide more details in the supplementary information (SI-C).

I have also made two additional assumptions that could easily be relaxed. First, I assumed that the voting rule by which the committee makes decisions is known with certainty and followed strictly. Second, proposals are conditionally independent. I relax the latter assumption by modeling the unobserved heterogeneity across groups with a random intercept in the supplementary information (SI-D). The former assumption might be relaxed by modeling $\mathcal{R}$ parametrically. I leave this extension to future work.

I refer to the model above as a multivariate probit model with partial observability or, for short, the partial m-probit. 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), bivariate 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). The problem addressed by the partial m-probit is not one of correlated (sequential) choices but of the non-observability of the simultaneous choices.

## 1.2 Likelihood and Prior Density

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

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

$$\begin{aligned} f(\mathbf{y}_j, \mathbf{X}_j|\boldsymbol{\beta}) &= \int_{p_{1j}} \dots \int_{p_{Mj}} \boldsymbol{\phi}(\mathbf{y}_j^*|\mathbf{X}_j\boldsymbol{\beta}) d\mathbf{y}_j^* \\ &= \boldsymbol{\Phi}_{\mathcal{P}(\mathbf{y}_j)}(\mathbf{X}_j\boldsymbol{\beta}), \end{aligned} \tag{1.2}$$

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 $\mathbf{y}_j$, i.e., $p_{ij} = [0, \infty)$ if $y_{ij} = 1$ and $p_{ij} = (-\infty, 0)$ if $y_{ij} = 0$. To write this more compactly, I define $\mathcal{P}(\mathbf{y}_j)$ as the function that generates all $p_{1j}, ..., p_{Mj}$ given $\mathbf{y}_j$ and let $\boldsymbol{\Phi}_{\mathcal{P}(\mathbf{y}_j)}(\cdot)$ be the implied distribution function.

Let $\tilde{\mathbf{y}}$ be a hypothetical vote profile and let $V(1)$ be the set of all hypothetical vote profiles for which $\sum_i \tilde{y}_i \geq \mathcal{R}$ holds. In other words, this set contains all vote profiles that realize an adoption outcome ($b_j = 1$). Let $V(0)$ be the complement set. Both sets are always finite but potentially large. For example, in the case of the UN Security Council, $V(1)$ is of size 848 and $V(0)$ of size 31,920.[6]

Using these two definitions, I can write the probability for $b_j = 1$ (and its complement) as the sum over the probabilities for all hypothetical vote profiles that can realize $b_j = 1$ ($b_j = 0$) and, after additionally relying on the conditional independence assumption across proposals, the likelihood is obtained by taking the product over all decisions. Formally, this is

$$\mathcal{L}(\boldsymbol{\beta}|\mathbf{X}, \mathbf{b}) = \prod_j \sum_{\tilde{\mathbf{y}} \in V(b_j)} \left[ \boldsymbol{\Phi}_{\mathcal{P}(\tilde{\mathbf{y}})}(\mathbf{X}_j \boldsymbol{\beta}) \right]. \tag{1.3}$$

Bayesian inference complements the likelihood with a prior density for the parameters (the coefficients). I follow convention and assume that they are jointly normal with a 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 (1.3) and to the prior density.

The structure of the likelihood function is surprisingly general and can accommodate much more specific decision records than those with binary adoption/rejection information. Suppose, for example, that, in addition to knowing that the proposal passed, an analyst also knows that it passed with some vote margin. In this case, the set of permissible vote profiles $V$ in equation (1.3) can be substantially reduced. In fact, if the analyst knows how each and every member voted, that is, if there is a voting record, then $V$ shrinks to a set with a single vote profile. In this case, the equation (1.3) reduces to a multivariate probit model, which is, since the covariance matrix is assumed to be the identity matrix, an ordinary probit model with $J \times M$ observations. There is also nothing in the structure of the likelihood that precludes the amount of information from varying across decisions. This implies that a partially observed voting record can be accommodated within the likelihood function without difficulty or formal extensions.

Finally, it is worth placing this model in the broader context of the (statistical) literature. The bivariate probit with partial observability by Poirier (1980) and the

[6] Since the five permanent members must always agree, the number of adopting coalitions is identical to the number of adopting coalitions among the 10 non-permanent members, which is given by $\sum_{m=4}^{10} \binom{10}{m}$. The number of rejecting coalitions is given by the difference between the total number of coalitions among all members minus the number of adopting coalitions, i.e., $\sum_{m=0}^{15} \binom{15}{m} - \sum_{m=4}^{10} \binom{10}{m}$.

bilateral cooperation model by Przeworski and Vreeland (2002) emerge as special cases of equation (1.3) if $M = 2$ and the voting rule is unanimity. In a recent contribution, Poirier (2014) extended his 1980 model to the case of $M > 2$ but remains focused on the case of unanimity.[7] More importantly, each factor in the likelihood above is the (complementary) cumulative density function of Poisson's Binomial distribution[8] parameterized with a set of probit functions (proofs for both statements appear in SI-A).

## 1.3 Identification

Before I continue with the computation of the posterior distribution, I discuss the (classical) parametric identification of the likelihood. A likelihood is said to be (parametrically) identified if a unique set of estimates exists for the parameters of a model. In the supplementary information (SI-B), I show that the conditional mean for the likelihood in equation (1.3) is always identified. The system of nonlinear equations that maps the structural parameters $\boldsymbol{\beta}$ to the reduced-form conditional means is identified under some conditions. Using a linearization of this system with a first-order Taylor series expansion, I show that it is identified if the aggregate design matrix has full rank. The aggregate design matrix of dimension $J \times K$ results from stacking the $J$ vectors that result from column-averaging all $\mathbf{X}_j$ matrices on top of each other.

The classical parametric identification condition is empirically verifiable by checking if the design matrix of the model ($\boldsymbol{X}$), after averaging all variables for each decision, has linear independent columns. Trivially, this condition will fail whether the design matrix before averaging does not have full rank. However, it will also fail if the design matrix has full rank and a variable exhibits variation within but not across decisions. In that case, the variable will be constant after averaging as well as a linear combination of the intercept. This renders the aggregate design matrix less than full rank, and the effect of the respective variable and the intercept are not separately identifiable. In

[7] Poirier (1980, 2014) allows for member-specific effects and correlated choices among agents. However, his theoretical results suggest that, even under a unanimity voting rule, member-specific effects and the correlation among agents' choices is at best only partially identifiable from the data. See also section 1.3 in this paper.

[8] In the most comprehensive paper on the distribution, Wang (1993) follows a reviewer suggestion and refers to the density as Poisson's binomial density. While this name choice is much less confusing than earlier conventions, an even more descriptive name might be "heterogeneous binomial density" since it emerges from the convolution of heterogeneous Bernoulli densities, and the standard binomial density is a special case if all choice probabilities are identical.

practice, this implies, for example, that, for a committee with constant membership, fixed effects for members or member-specific effects are unidentifiable and consequently cannot be estimated.

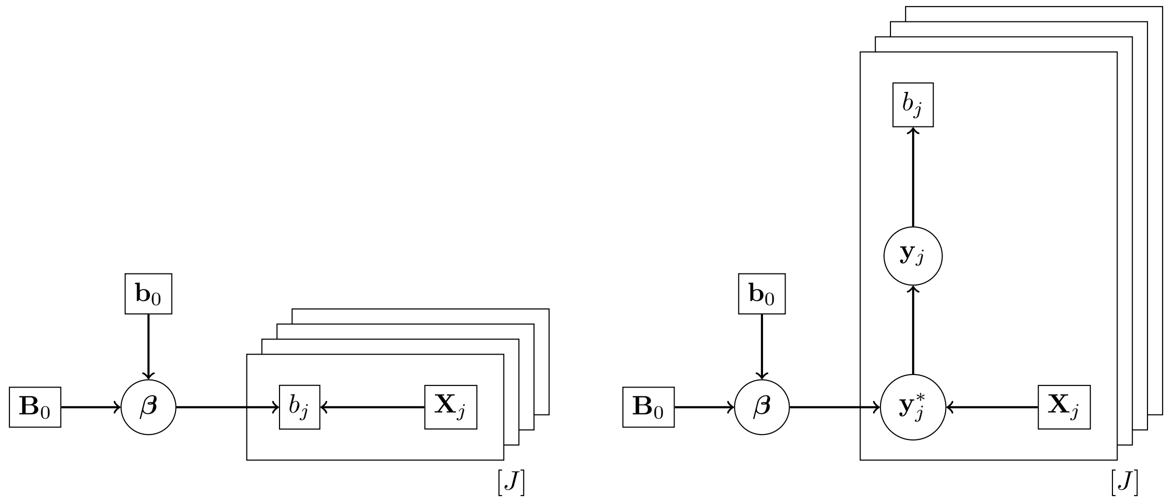
The identification condition is based on the linearization of a system of nonlinear equations. Consequently, there might be instances where the condition of a full-rank aggregate design matrix holds but a unique set of parameters still does not exist. This is problematic for frequentist inference because, for example, the properties of the maximum likelihood estimator are at least inconvenient for unidentified likelihoods. However, in a Bayesian analysis, unidentified likelihoods are of less concern since the posterior density will still be proper if proper priors are used. The only consequence is that the marginal posterior density of the intercept will be, in the worst case, perfectly negatively correlated with the marginal posterior density of the unidentified effect. From a theoretical perspective, this is not a problem, but in practice, it means that the Gibbs sampler presented in the next section will be very slow in exploring the posterior density, which is why an identified likelihood is advantageous for a Bayesian analysis.

# 2 Posterior Computation

As in most Bayesian models, the posterior density cannot be marginalized analytically, which prompts me to construct a Gibbs sampler to simulate from the density and use the samples to characterize the density with a desirable degree of accuracy. A Gibbs sampler requires derivation of the full conditional densities for all unknown quantities in the model. To derive them, I use a theorem by Lauritzen et al. (1990), who show that, if a joint density (such as a posterior density) can be written as a directed acyclic graph (DAG), the full conditionals are given by a simple formula (see SI-D).

A DAG representation of the posterior density appears in figure 1a. Each node in this graph is a random variable. Rectangular nodes indicate observed variables (the data and hyperparameters), while circle nodes represent unobserved variables (parameters). An arrow indicates the dependencies between these variables, and the plates indicate the $J$ replications. The graph is acyclic since it has no cyclic dependency structure.

The conditional for $\boldsymbol{\beta}$ in figure 1a is not a member of a known parametric family from which samples can be easily drawn. To arrive at full conditionals that are easy to sample from, I follow a data augmentation strategy (Tanner and Wong, 1987) and




**(a)** unaugmented **(b)** double augmented

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

explicitly introduce two variables from the derivation of the likelihood. The augmented DAG appears in figure 1b. The first augmentation is identical to the Albert-Chib augmentation in a Bayesian (multivariate) probit model (Albert and Chib, 1993; Chib and Greenberg, 1998), explicitly introducing $\mathbf{y}_j^*$, the latent utility, in the model. The second augmentation augments the latent utility with $\mathbf{y}_j$, the unobserved votes. Because of this sequential augmentation, I refer to the Gibbs sampler as a double-augmented Gibbs sampler.

Applying the result from Lauritzen et al. (1990) cited above yields three full conditionals for the three unobserved variables in the DAG. The conditional for $\boldsymbol{\beta}$ can then be written as follows:

$$\begin{aligned} f(\boldsymbol{\beta}|\mathbf{b}_0, \mathbf{B}_0, \mathbf{y}^*, \mathbf{y}, \mathbf{b}, \mathbf{X}) \propto\ & f(\boldsymbol{\beta}|\mathbf{b}_0, \mathbf{B}_0) \times \prod_j f(\mathbf{y}_j^*|\mathbf{X}_j, \boldsymbol{\beta}) \\ & = \boldsymbol{\phi}\Big((\mathbf{B}_0^{-1} + \mathbf{X}'\mathbf{X})^{-1}(\mathbf{B}_0^{-1}\mathbf{b}_0 + \mathbf{X}'\mathbf{y}^*), (\mathbf{B}_0^{-1} + \mathbf{X}'\mathbf{X})^{-1}\Big). \end{aligned} \tag{2.1}$$

The two other conditionals and their sampling algorithms are given in the supplementary information (SI-D).

It is not a coincidence that the functional form of the conditional for $\boldsymbol{\beta}$ is exactly the same as the conditional for an ordinary probit model and a Bayesian normal regression model when the same prior for $\boldsymbol{\beta}$ is chosen. The primary difference between a probit model, a partial m-probit, and a normal regression is that only in the latter case is the variable $\mathbf{y}^*$ fully observed. In the other two cases, $\mathbf{y}^*$ is observed only in a coerced fashion. However, the precise nature of the coercion is irrelevant once the data are augmented. In fact, the very purpose of the data-augmentation strategy is to render the coefficients conditionally independent of the coerced data.

The Gibbs sampler, which I refer to as the double-augmented Gibbs sampler, is an iterative sampling from the conditionals until convergence (see SI-D for the details). It has a very intuitive sequence: 1) choose some starting value for the coefficients; 2) conditional on these values, the covariates, and the decision record, draw vote profiles for all decisions; 3) conditional on the vote profiles and the covariates, draw the vector of latent utilities for all decisions; 4) conditional on the latent utilities and the covariates, draw the coefficients; and 5) repeat until convergence.

The Gibbs sampler is implemented in an open-source `R`-package `consilium`, which accompanies this paper. I also conducted Monte Carlo experiments to verify that the Gibbs sampler (and its implementation) obtains samples from the posterior density and to provide some insights into the computational costs of the model (see SI-F).

# 3 Aggregation Costs

Whenever data are aggregated, the analyst pays a price in terms of a) effects that can not estimated, b) posterior uncertainty (efficiency), and c) bias for the estimable effects. What are the costs when analyzing a decision record relative to an analysis with a voting record? The discussion on identification has highlighted that member-specific effects in committees with constant membership cannot be estimated with decision-record data. This is in sharp contrast to voting records where member-specific effects can be estimated. If such effects are the object of inquiry, decision records cannot be used. Moreover, even if the effect of interest is assumed to be shared, its inference might be hampered if the analyst suspects relevant, unobserved member-specific heterogeneity. While such heterogeneity could be modeled with varying intercepts in an analysis of voting records, it is infeasible with decision records.

For estimable effects, posterior uncertainty and aggregation bias are further potential costs. Aggregation bias, as discussed in the classical ecological inference literature

(Erbring, 1989; King, 1997), is a form of confounding with the group-assignment variable. Since the number of groups equals the number of observations in the aggregated sample, adjustment strategies, that is, weighting with or conditioning by the group-assignment variable, are not feasible. However, if the group-assignment variable is chosen at random, grouping cannot lead to bias, the classical example for aggregation bias being spatially aggregated data on vote choice and race in mass elections. To the extent that electoral districts are drawn with perfect knowledge about vote choice and race in an election, the effect of race on vote choice in the same election cannot be inferred without bias.

For aggregation bias to be a threat to inference with decision records, the process of assigning members to decisions (the "groups") must be a function of members' vote choices on a proposal and some unmeasured covariate. If that were the case, then the proposal-assignment vector would be a confounder for which we cannot adjust and aggregation bias is unavoidable[9]. While membership in a committee is presumably a function of (expected) vote choices and potentially some unmeasured covariates, the committee's membership is usually constant over a certain period. Within this period of constant membership, aggregation bias cannot occur.

Beyond aggregation bias, there is also the issue of posterior uncertainty since aggregation reduces the effective sample size. While posterior uncertainty might seem secondary, it becomes paramount once the aggregation reduces information to a point where no variation is left to draw inference from. For instance, in an institution where all members have a high (low) average probability of voting one way or the other, there is a chance that the decision record will exhibit no variation and the posterior equals the prior.[10]

## 4 Advantages of the Model

Unsurprisingly, the structural model tends to produce more efficient estimates since the amount of information in the estimation is larger. More importantly, the structural

[9]In SI-E, I illustrate in a proof for the bivariate case with vague priors that the posterior mean of the coefficient is not a function of the vote choices (but only the decision record) if the proposal-assignment vector is orthogonal to the latent utility and the covariates. The proof follows directly from classical results in ecological inference (e.g., Erbring, 1989; Palmquist, 1993; King, 1997) that give extensions to the multivariate case.

[10]The aforementioned Monte Carlo experiments in SI-F also provide some intuition about the increase in posterior uncertainty that comes with aggregation.

model allows one a) to choose the correct reduced-form specification, b) to estimate vote-choice probabilities instead of adoption probabilities, and c) to combine partially observed voting records with decision records.

## 4.1 Choosing Specifications

Decision records are used for empirical inference on a regular basis with convenient models such as a probit. However, perhaps surprisingly, the specification that is usually chosen is not the reduced-form complement to the structural model outlined in the previous section. As an example, consider this simple partial m-probit:

$$p(b_j = 1) = \sum_{\tilde{\mathbf{y}} \in V(1)} \mathbf{\Phi}_{\mathcal{P}(\tilde{\mathbf{y}})}\big(\beta_0 + \mathbf{x}_j \beta_1\big) \tag{4.1}$$

and one reduced-form complement with $z_j = \sum_i x_{ij}$:

$$p(b_j = 1) = \Phi\big(\beta_0' + z_j \beta_1'\big), \tag{4.2}$$

where one might scale $z_j$ by dividing by $M$, which then makes $z_j$ the average of $\mathbf{x}_j$.[11]

However, typically, the sum in equation (4.2) is not taken over all members but only a subset. For example, in studies on the UN Security Council, measures of political or economic closeness between the conflict location and the permanent members are included (e.g., an indicator for a defense alliance), although the Council consists of the five permanent and 10 nonpermanent members (e.g., Gilligan and Stedman, 2003; Mullenbach, 2005; Beardsley and Schmidt, 2012; Hultman, 2013; Stojek and Tir, 2015).

However, leaving out parts of the membership introduces measurement error in $z_j$.[12] As in any other setting with errors in variables, the resulting coefficient estimates will be biased. Moreover, the estimated effect cannot generally be interpreted as a member-specific effect since, as shown in section 1.3, there is no variation in decision-record data that can identify member-specific effects.

[11] Strictly speaking, the probit model (or logit model) has no theoretical basis as a reduced-form complement but is often used in practice. A theoretical justified reduced-form complement is the linear regression model that provides the best linear approximation to the non-linear conditional expectation function of the data.

[12] Only if the means of those members included and those left out coincide for all decisions is no measurement error introduced.

## 4.2 Estimating Vote-Choice Probabilities

Both the structural model and the reduced-form model allow one to estimate the predicted probability of observing the adoption of a proposal (the "adoption probability"). These predicted probabilities can be used to characterize how much a one-unit increase in a covariate changes the adoption probability. In addition to the adoption probability, the structural model also allows one to calculate the predicted probability of a supportive vote choice (the "vote choice probability"). This quantity is typically calculated when one analyzes a voting record and can be used to describe how a one-unit change in a covariate changes the vote-choice probability.

While the adoption probability can be of considerable interest in some situations (e.g., if the analyst intends to predict the adoption of proposals), it must be recognized that it is not only a function of the coefficients and the covariates but also of the institutional structure (the size of the membership and the majority threshold). Consequently, it is a conditional probability whose magnitude, as it turns out, is not a linear function of the vote-choice probability.

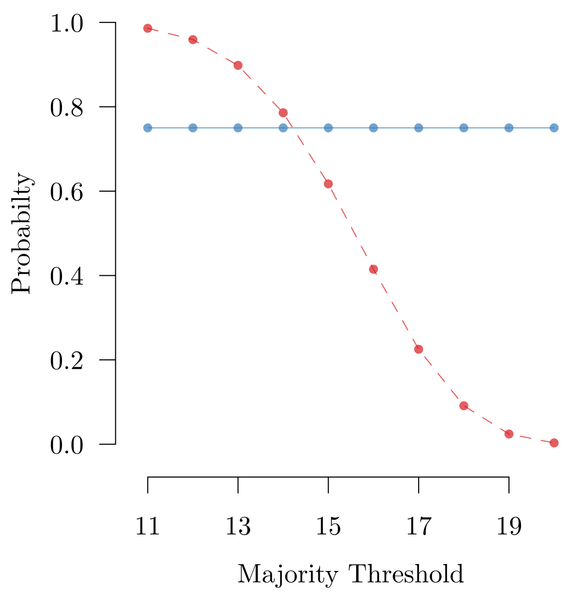



**Figure 2:** Illustration of the relationship between vote-choice and adoption probabilities for a committee of 20 members. The solid line indicates the different vote-choice probabilities and the dashed line the corresponding adoption probabilities.

To illustrate, consider a committee of 20 members with various majority thresholds between 11 (a simple majority) and 20 (unanimity). To simplify matters without loss of generality, suppose also that the vote-choice probability for all members is homogeneous at 0.75. The vote-choice probabilities are shown with a solid line in figure 2. The figure also shows, corresponding to each of these vote-choice probabilities, the implied adoption probabilities conditional on the 10 majority thresholds (dashes). While the vote-choice probabilities are constant across committees of various sizes, the adoption probabilities are a monotone, but nonlinear, function of the vote-choice probabilities.

The monotonicity of the adoption probability with respect to the vote-choice probability is good news because it suggests that the direction of any effect on the vote-choice probability can always be inferred from the direction of the effect on the adoption probability. However, the nonlinearity also suggests that the adoption probability cannot be easily compared across different institutional contexts. Figure 2 illustrates that, even in the absence of differences in vote-choice probabilities in two different institutional contexts, adoption probabilities will vary if the membership or majority threshold differs.

Furthermore, the magnitude of the adoption probability can be a very poor indicator of the magnitude of the vote-choice probability. Figure 2 illustrates that the closer the majority threshold moves toward unanimity, the smaller the adoption probability becomes up to the point where it is minuscule. All the while, the vote-choice probability remains constant. This emphasizes that it is quite important to define what the quantity of interest is when analyzing decision records. If the analyst's interest is in understanding how covariates change the vote-choice probability, the structural model is the more promising approach.

## 4.3 Including a Partially Observed Voting Record

The discussions on the likelihood function and the Gibbs sampler have already highlighted that including a partially observed voting record is very easy when using the structural model but infeasible when using a reduced-form model. Ordering the proposals for which only the decision record is available from $j = 1, ..., K$ and the proposals for which a voting record is available from $j = K + 1, ..., J$, the two-component likelihood function with parameter vector $\dot{\boldsymbol{\beta}}$ takes the following form:

$$\dot{\mathcal{L}}(\dot{\boldsymbol{\beta}}|\mathbf{X}, \mathbf{b}, \mathbf{Y}) = \prod_{j=1}^{K} \sum_{\tilde{\mathbf{y}} \in V(b_j)} \left( \boldsymbol{\Phi}_{\mathcal{P}(\tilde{\mathbf{y}})}(\mathbf{X}_j \dot{\boldsymbol{\beta}}) \right) \cdot \prod_{j=K+1}^{J} \left( \boldsymbol{\Phi}_{\mathcal{P}(\mathbf{y}_j)}(\mathbf{X}_j \dot{\boldsymbol{\beta}}) \right), \tag{4.3}$$

where $\mathbf{Y}$ denotes the stacked matrix of all observed voting profiles. The Gibbs sampler is easy to expand by simply dropping the sampling of the vote profiles for those proposals where a voting record is available. It is fairly intuitive that the posterior inference from this likelihood will be more certain than the posterior inference from the likelihood in equation (1.3).

Including a partially observed voting record can also reduce aggregation bias. In the supplementary information (SI-E), I show that the familiar missing-at-random (MAR) condition from the literature on missing data (Little and Rubin, 2002) is a necessary assumption to reduce aggregation bias. In particular, it is necessary that, conditional on the covariates, the observability of the recorded votes is random. If this assumption is fulfilled, aggregation bias will be removed from the estimates. Complementarily, if the observed voting record is a nonrandom subset, it might cause selection bias if incorporated.

Another benefit of including a partially observed voting record is that one can relax the assumption of shared effects across all committee members. These effects are obviously identifiable from voting records and, as discussed in section 1.3, unidentifiable with decision records. Consequently, if member-specific effects are of interest and included in the model, the identifying variation to estimate these effects will come from the variation in the partially observed voting record. The conditional-independence assumption with respect to vote choices could also be relaxed for the same reasons.

The ability to supplement a decision record with a partially observed voting record can also have advantageous consequences for data collection. Consider, for example, a situation where the analyst wishes to collect another sample of votes from a voting record to decrease the posterior uncertainty but, as it happens, collecting such a sample proves quite expensive. To avoid these costs, the analyst could instead collect a large sample from the decision record. To the extent that collecting a large sample from a decision record is much cheaper than a sample from the voting record, this reduces the costs of data collection.

# 5 Replication: US State Supreme Court Decisions

I replicate a study by Caldarone et al. (2009) to contrast the coefficient estimates when a voting record is used with the coefficient estimates used when I artificially delete (some of) the recorded votes and only use the decision record in the analysis. Caldarone et al.

(2009) test the prediction "that nonpartisan elections increase the incentives of judges to cater to voters' ideological leanings" (p. 563). To test their prediction, the authors assemble a dataset of US state supreme court decisions on abortion for the period from 1980-2006. They collect these data for all state supreme courts for which judges face contested statewide elections. Their dataset contains 19 state supreme courts (which vary in size between five and nine judges) and a total of 85 abortion decisions.

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 toward the state's public opinion) or unpopular. Consequently, the dependent variable takes a 1 if the justice votes "pro-choice" and the state leans "pro-choice" or if he votes "pro-life" and the state leans "pro-life" (Caldarone et al., 2009, p. 565). In the authors' dataset, 261 votes are popular (43%). The authors' independent variable of interest is a binary variable indicating whether a supreme court justice was elected in a nonpartisan election. Of the 85 abortion decisions, 39 were made in a partisan electoral environment (46%).

A replication of the authors' baseline specification (Model 1 in their table) using a Bayesian probit model appears as the lower row (row 5) in the coefficient plot in figure 3. The upper row (row 1) instead shows the results produced when I retained only a binary variable indicating whether the courts passed a popular decision by majority rule and estimated the same specification using the partial m-probit.[13] 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 (N=605).

For the main variable of interest, nonpartisan election, the posterior probability that there is a positive effect of nonpartisan elections is still 0.9 even after dropping all votes and 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 a similar posterior mean but with considerably larger posterior uncertainty. The effect of the justices' party being aligned with public opinion is estimated to be a little larger and to have more posterior uncertainty.

One benefit of the structural model is that it allows one to combine a partially observed voting record with a decision record to decrease the costs of aggregation. To

[13] For both models, I use vague normal priors centered at 0 with a variance of 10. The regression table in SI-H contains detailed information on the Gibbs sampling parameters and convergence.

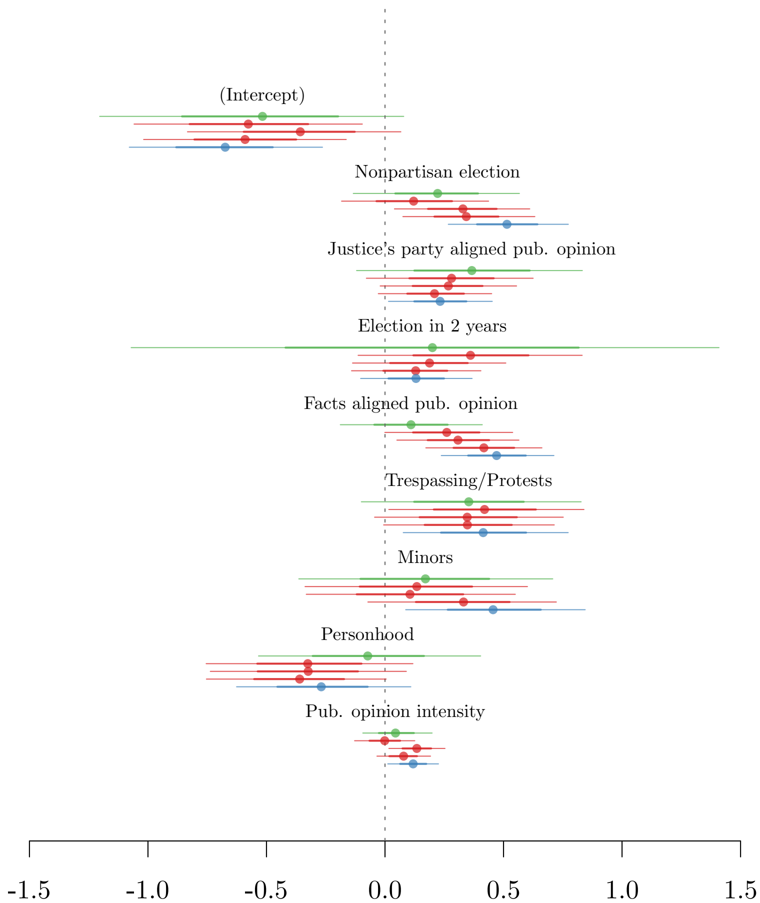



**Figure 3:** Regression results from a Bayesian partial m-probit model with a decision record (row 1) and a partially observed voting records (row 2: 25% observed, row 3: 50%, row 4: 75%) as well as a Bayesian probit model with justices' voting record (row 5). While the dots indicate the posterior mean, the segments represent the 95% and 68% posterior intervals, respectively.

demonstrate this, I re-estimate the partial m-probit with random samples of recorded votes and the same prior. The results appear again in the same coefficient plot (row 2-4). The upper bars (row 2) show the estimates when, in addition to the decision record, 25% of all votes are observed, followed by the estimates for 50% and 75%. As expected, the greater the number of recorded votes included in the analysis is, the higher the similarity will be between the estimates of the partial m-probit and the ordinary probit. For most variables, the trend toward the probit estimates and the decrease in posterior uncertainty appear to be quite linear (e.g., for nonpartisan election or the justices' party alignment). However, for some, there is a significant payoff for observing some votes compared to no votes (e.g., elections in two years). This suggests that, at least in some situations, collecting a few votes to supplement the decision record can greatly improve the quality of the estimates.

## 6 Application: Trade and UN Operations

A major line of inquiry in the literature on the UN Security Council aims to understand Council members' motives in involving themselves in third-party conflicts within the framework of the United Nations (e.g., Gilligan and Stedman, 2003; Hultman, 2013; Stojek and Tir, 2015). Are the members more likely to support a UN Blue Helmet operation in conflicts where they expect economic or political gains from a swift end to the conflict? I reconsider this question by estimating the effect of trade relationships between the members and the territories in conflict, highlighting the advantages of using the partial m-probit.

To conduct this analysis, I use a revised version of the cross-sectional panel dataset by Hultman (2013), which combines the UCDP/PRIO Armed Conflict dataset (Gleditsch et al., 2002) with the dataset on third-party interventions by Mullenbach (2005). Focusing on intrastate conflicts that occurred outside the territories of the Council's permanent members, the effective number of observations is 885, nested in 102 conflicts. There are 17 conflicts for which the UN Security Council deployed a UN operation.

I interpret each observation as an instance where each of the 15 Council members[14] must decide to support or oppose the deployment of a UN operation. Consequently, the unit of analysis in my dataset is a UN Security Council member's binary support choice per conflict-year. I supplement these data with information about the size of total trade (export and imports) between a Council member and the conflict location (Barbieri et al., 2009).[15]

There is no complete voting record from the UN Security Council. While some votes from the UN Security Council are on record and could be incorporated, these recorded votes constitute a selected sample from the set of all votes. This is because the 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

[14]The UN Security Council uses a veto-majority rule. Specifically, nine out of 15 member states must approve a proposal, but each of the five permanent members (China, France, Russia, the United Kingdom, and the United States) has a veto.

[15]For a detailed description of the data, see SI-G.

how much support there will be for each" (Luard, 1994, p. 19). Most conflicts are never discussed in the Council or they are discussed but the Council cannot agree on whether to deploy a UN operation. Consequently, recorded votes only occur in very particular circumstances (if the Council agrees to deploy) and incorporating these recorded votes is likely to result in a selection bias.

I condition on a set of common causes to decrease the threat of confounding and also include a varying intercept for the conflict location. In order to account for annual and conflict-period trends, I include two B-splines (with the deployment year and the period of the conflict). Except for the binary independent variables, I center and scale all variables by twice their standard deviation before estimating each model, which aids in the construction of weakly informative, normal priors centered at 0 and a variance of 5.

The estimates appear in table 2 in the row labeled model 1 (see also SI-H, for the full table and details on the Gibbs sampling parameters and convergence). The estimates suggest that an increase in trade between a Council member and the conflict location decreases a member's probability of supporting a UN operation. The posterior probability for this effect to be negative is 0.95.

To illustrate the difference between the inference from the partial m-probit and a reduced-form model, I aggregate the data to a dataset of conflict-years. In the aggregated dataset, the trade variable measures the total trade of all Council members with the conflict location. The estimates from a probit model appear in table 2 in the row labeled model 2. As expected, the sign of the association is identical to model 1. Interestingly, the posterior probability for this association to be negative is only 0.89 – reflecting that the partial m-probit delivers more efficient estimates. Notice, that the magnitude of the coefficient from model 2 provides no information about how trade between a Council member and the conflict location decreases a member's probability of supporting a UN operation. This information is only available from the partial m-probit estimate.

Typical studies on the UN Security Council[16] do not include covariates that measure the variation of a concept across *all* members but, rather, usually focus on the permanent five (the P5). To illustrate that this can lead to a misleading inference in the trade case, I estimate the effect of total trade of the P5 leaving out the contribution from the 10 nonpermanent members (see row labeled model 3 in the table 2)

[16]See, for example, Gilligan and Stedman (2003); Mullenbach (2005); Fortna (2008); Beardsley and Schmidt (2012); Hultman (2013) and Stojek and Tir (2015).

| Model | Variable | Posterior |
|---|---|---|
| Model 1 | log(Trade) | $-1.09$ |
| | | $[-2.61, 0.15]$ |
| Model 2 | log(Trade) of all mbrs. | $-1.02$ |
| | | $[-3.05, 0.66]$ |
| | | |
| Misspecified | | |
| Model 3 | log(Trade) of P5 only | $-1.40$ |
| | | [-3.55,0.26] |
| Model 4 | log(Trade) of US only | $-1.02$ |
| | | [-3.02,0.71] |
| Model 5 | log(Trade) of UK only | $-2.97$ |
| | | [-6.36,-0.50] |
| Model 6 | log(Trade) of FR only | $-0.15$ |
| | | [-1.60,1.63] |
| Model 7 | log(Trade) of RU only | $-2.09$ |
| | | [-4.75,-0.24] |
| Model 8 | log(Trade) of CH only | $-1.31$ |
| | | [-3.34,0.41] |

**Table 2:** Regression results from a Bayesian partial m-probit model (Model 1, $N = 15 \times 885$) and seven Bayesian probit models (Model 2-8, $N = 885$) each with posterior means and 95% posterior intervals in parentheses. All models include covariates, varying intercepts and B-splines ($df. = 3$).

and include each P5 trade share separately (rows labeled models 4-8). As explained in section 4.1, none of these estimates can be interpreted as estimates of the effect of trade on the respective members' vote choices (or the heterogeneous effect of trade on members in general). Instead, the estimates from the models 3-8 can be interpreted as a version of the estimates in model 2 but contaminated by measurement error.

The results here are at odds with the recent analysis by Stojek and Tir (2015). Using data from Fortna (2008) and a logit model on UN peacekeeping deployment, they estimate a positive effect of the P5 total trade volume on the probability of deployment. While their unit of analysis is a ceasefire analysis, the positive effect they estimate is largely driven by conflicts in which permanent members are directly involved (e.g., the Northern Ireland conflict), while the data I use exclude all conflicts that occur in the territory of the permanent member states.

## 7 Discussion

Analyzing a decision record instead of a voting record is not something for which one would hope. The aggregation of vote choices by a voting rule increases the uncertainty of estimable effects and may even bias them. It also prohibits the estimation of member-specific effects. However, confronted with the choice between abstaining from an analysis or relying on decision records, an analyst might still prefer the latter. In this paper, I argue that, if the analyst decides to examine the decision record, his or her analysis can be improved by turning to a structural model instead of opting for a convenient reduced-form model.

In this paper, I highlight several advantages of the structural model; however, the most important might be that it allows one to bring partially observed voting records into the analysis. Inter alia, the replication of the study by Caldarone et al. (2009) highlights that there are large benefits in terms of efficiency when analyzing a decision record jointly with a voting record sample even if the later is small. Beyond efficiency, such a joined analysis opens the route to estimate member-specific effects as well as reduce potential aggregation bias. This suggests that effort should be made to collect a sample of votes from archival documents or committee members' personal notes. Potentially, even if no explicit voting record is provided in existing documents, it might be still feasible to reconstruct a small set of votes with high confidence based on in-depth qualitative research.

Beyond the question of which model to use to analyze an available decision record, one might wonder for which (types of) institutions one can successfully compile a decision record in the first place and then apply the partial m-probit. While a systematic listing is beyond the scope of this paper, a few examples might highlight that decision records are either directly available from particular institutions or can be compiled based on available knowledge about these institutions.

A decision record is typically available from institutions where members vote on a regular basis on issues but decide not to publish these votes. While I have artificially created a decision record in the case of the US state supreme courts in section 5, international courts in particular (e.g., the European Court of Justice or the European Court of Human Rights) typically publish only the decisions on each case but not judges' votes[17]. Another example in this category is central banks other than those of the US and UK where voting records are published.

However, even committees that do not explicitly vote on each decision may adopt proposals by acclamation on a regular basis, which gives rise to a decision record that can be analyzed. The UN Security Council analyzed in section 6 is a case in point: the Council explicitly only votes on the deployment of UN peace operations that are known to pass but implicitly rejects all UN peace operations in ongoing conflicts by never advancing them to the voting stage in the first place. Another example is the IMF Executive Board, which approves loans by acclamation instead of voting and whose decision record has been analyzed previously using reduced-form models (Breen, 2013; Copelovitch, 2010; Broz and Hawes, 2006).

However, not every institution's decision record will be suitable for analysis nor will it always be possible to compile a decision record in the first place. The ability to compile a decision record when it is not directly published by an institution, depends on the availability of a natural agenda that defines the issues under consideration at the respective institution. In the case of the UN Security Council, for example, studies assume that the agenda is defined by the set of ongoing conflicts. Suitable decision records are those where the conflict between committee members across decisions evolves around a binary decision "to do something or not". However, if the conflict across decisions is determined by a conflict over how much to do (and consequently something is always done), the analysis of decisions records will provide little further insights into the institution.

[17] See, e.g., Carrubba et al. (2008) and Helfer and Voeten (2014) for studies using decision records from international courts.

# 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.

Austen-Smith, D. and J. S. Banks (1999). *Positive Political Theory I: Collective Preference*. University of Michigan Press.

Barbieri, K., O. M. G. Keshk, and B. M. Pollins (2009). Trading Data Evaluating Our Assumptions and Coding Rules. *Conflict Management and Peace Science 26*(5), 471–491.

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.

Breen, M. (2013). *The Politics of IMF Lending*. Basingstoke: Palgrave Macmillan.

Broz, L. J. and M. B. Hawes (2006). U.S. Domestic Politics and International Monetary Fund Policy. In D. G. Hawkins, D. A. Lake, D. L. Nielson, and M. J. Tierney (Eds.), *Delegation and Agency in International Organizations*, pp. 77–106. Cambridge University Press.

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.

Carrubba, C. J., M. Gabel, and C. Hankla (2008). Judicial Behavior Under Political Constraints: Evidence from the European Court of Justice. *American Political Science Review 102*(4), 435–452.

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.

Copelovitch, M. S. (2010). *The International Monetary Fund in the Global Economy: Banks, Bonds, and Bailouts*. Cambridge: Cambridge University Press.

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

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

Erbring, L. (1989). Individuals Writ Large: An Epilogue on the "Ecological Fallacy". *Political Analysis 1*(1), 235–269.

Fortna, V. P. (2008). *Does peacekeeping work?* Princeton: Princeton University Press.

Fry, M., D. Julius, L. Mahadeva, S. Roger, and G. Sterne (2000). Key Issues in the Choice of Monetary Policy Framework. In L. Mahadeva and G. Sterne (Eds.), *Monetary Policy Frameworks in a Global Context*. London: Routledge.

Gilligan, M. and S. J. Stedman (2003). Where Do the Peacekeepers Go? *International Studies Review 5*(4), 37–54.

Gleditsch, N. P., P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand (2002). Armed Conflict 1946-2001: A New Dataset. *Journal of Peace Research 39*(5), 615–637.

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.

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.

Helfer, L. R. and E. Voeten (2014). International Courts as Agents of Legal Change: Evidence from LGBT Rights in Europe. *International Organization 68*(1), 77–110.

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

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

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.

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.). London: Maxmillan Press.

Mullenbach, M. J. (2005). Deciding to Keep Peace. An Analysis of International Influences on the Establishment of Third-Party Peacekeeping Missions. *International Studies Quarterly 49*(3), 529–555.

Palmquist, B. L. (1993). *Ecological Inference, Aggregate Data Analysis of U.S. Elections, and the Socialist Party of America*. Ph. D. thesis, University of California, Berkeley.

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

Poirier, D. J. (2014). Identification in Multivariate Partial Observability Probit. *International Journal of Mathematical Modelling and Numerical Optimisation 5*(1-2), 45–63.

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

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

Schermers, H. G. and N. M. Blokker (2011). *International Institutional Law* (5th ed.). Leiden: Nijhoff.

Stojek, S. M. and J. Tir (2015). The Supply Side of United Nations Peacekeeping Operations: Trade Ties and United Nations-led Deployments to Civil War States. *European Journal of International Relations 21*(2), 352–376.

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.

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.

Wang, Y. H. (1993). On the Number of Successes in Independent Trials. *Statistica Sinica 3*(2), 295–312.

Wooldridge, J. M. (2001). *Econometric Analysis of Cross Section and Panel Data.* Cambridge: MIT Press.

# Supplementary Information: Analyzing Decision Records from Committees

Moritz Marbach

December 3, 2018

## Contents

# Appendix A Equivalence

**Proposition 1.** *If $M = \mathcal{R} = 2$ the likelihood function in equation (1.3) is identical to equation 3 in Przeworski and Vreeland (2002).*

*Proof.* If $M = \mathcal{R} = 2$ then $V(1)$ only contains a single element (the unanimous vote profile). The first product term in equation (1.3) thus reduces to $\mathbf{\Phi}_{\mathcal{P}(\mathbf{y}_j=\{1,1\})}(\mathbf{X}_j\boldsymbol{\beta})$, which can also be written as $\Phi(\mathbf{x}_{jA}\boldsymbol{\beta}) \times \Phi(\mathbf{x}_{jB}\boldsymbol{\beta})$, where $A$ and $B$ are the two members. Using the complement probability for the case of rejection of a proposal and using $b_j$ to select the right terms, the likelihood function can be written as $\mathcal{L}(\boldsymbol{\beta}|\mathbf{X}, \mathbf{b}) = \prod_j (\Phi(\mathbf{x}_{jA}\boldsymbol{\beta})\Phi(\mathbf{x}_{jB}\boldsymbol{\beta}))^{b_j} \times (1 - \Phi(\mathbf{x}_{jA}\boldsymbol{\beta})\Phi(\mathbf{x}_{jB}\boldsymbol{\beta}))^{1-b_j}$ which is identical to equation 3 in Przeworski and Vreeland (2002). □

**Definition 1.** (Wang, 1993) For a random variable $Y$ with $y = 0, ..., M$ that follows a Poisson's Binomial density with parameter $\mathbf{p}$, where $\mathbf{p} = (p_1, ..., p_M)$ and $0 < p_i < 1\ \forall i = 1, ..., M$, we write $Y \sim \text{PB}(\mathbf{p})$. The probability of observing exactly $y$ 'hits' is given by the probability mass function (pmf):

$$f_M(Y = y; \mathbf{p}) \equiv \sum_{A \in S_y} \left[ \prod_{i \in A} p_i \prod_{i \in A^c} (1 - p_i) \right]$$
$$\text{s.t. } S_y = \{A : A \subseteq \{1, ..., M\}, |A| = y\} \wedge A^c = S_y \setminus A.$$

**Definition 2.** (Wang, 1993) The probability of observing at most $K$ 'hits' is given by the cumulative distribution function (cdf):

$$F_M(Y \leq K; \mathbf{p}) \equiv \sum_{y=0}^{K} \sum_{A \in S_y} \left[ \prod_{i \in A} p_i \prod_{i \in A^c} (1 - p_i) \right].$$

Note, that $|A|$ is the cardinality of the ordered set $A$ and is at most $M$. $S_y$ is a set of ordered sets with cardinality $\binom{M}{x}$. Some example of $S_y$ might help to clarify the notation. Suppose $M = 3$ then:

$$
\begin{aligned}
S_0 &= \emptyset \\
S_1 &= \{\{1\},\{2\},\{3\}\} \\
S_2 &= \{\{1,2\},\{1,3\},\{2,3\}\} \\
S_3 &= \{\{1,2,3\}\}.
\end{aligned}
$$

**Proposition 2.** *A factor in the likelihood function in (1.3) is equivalent to the (complementary) cdf of a Poisson's Binomial density.*

*Proof.* Using the definition of the cdf we can establish the equivalence. Let $p_i = \Phi(\mathbf{x}_i\boldsymbol{\beta})$. First, notice the equality $\boldsymbol{\Phi}_{\mathcal{P}(\tilde{\mathbf{y}})}(\mathbf{X}\boldsymbol{\beta}) = \prod_{i\in A} p_i \prod_{i\in A^c}(1-p_i)$ since $A = \{i \ : \ \tilde{y}_i = 1 \ \forall i = 1,...,M\}$ as well as $A^c = \{i \ : \ \tilde{y}_i = 0 \ \forall i = 1,...,M\}$. Second, notice the equality $V(0) = \bigcup_{y=0}^{\mathcal{R}} S_y$ as long as we allow only for q-rules with threshold $\mathcal{R}$. By substitution we have:

$$
\begin{aligned}
\sum_{y=0}^{K}\sum_{A\in S_y}\left[\prod_{i\in A} p_i \prod_{i\in A^c}(1-p_i)\right] &= \sum_{y=0}^{K}\sum_{A\in S_y}\left[\prod_{i\in A}\Phi(\mathbf{x}_i\boldsymbol{\beta})\prod_{i\in A^c}(1-\Phi(\mathbf{x}_i\boldsymbol{\beta}))\right] \\
&= \sum_{y=0}^{K}\sum_{A\in S_y}\left[\boldsymbol{\Phi}_{\mathcal{P}(\tilde{\mathbf{y}})}(\mathbf{X}_j\boldsymbol{\beta})\right] \\
&= \sum_{\tilde{\mathbf{y}}\in V(0)}\left[\boldsymbol{\Phi}_{\mathcal{P}(\tilde{\mathbf{y}})}(\mathbf{X}_j\boldsymbol{\beta})\right].
\end{aligned}
$$

The case for $V(1)$ is analog. □

# Appendix B Identification

Define the following likelihood function:

$$.\mathcal{L}(\mathbf{p}|\mathbf{b}) = \prod_j (1 - F_M(y_j \leq (\mathcal{R}-1); \mathbf{p}_j))^{b_j} \cdot F_M(y_j \leq (\mathcal{R}-1); \mathbf{p}_j)^{1-b_j}, \tag{B.1}$$

where $F_M(\cdot)$ is the cdf of Poisson's Binomial defined as in appendix A. If $\mathbf{p}_j = (\Phi(\mathbf{x}_{1j}\boldsymbol{\beta}), ..., \Phi(\mathbf{x}_{Mj}\boldsymbol{\beta}))$ the likelihood is identical to equation (1.3) as shown in appendix A.

Note also, that the expected value for Poisson's Binomial pdf is given by $E(Y) = \theta = \sum_{i=1}^{M} p_i$ (Wang, 1993).

**Lemma 1.** *The maximum-likelihood estimator (MLE) for $\theta$ is unique.*

*Proof.* Since we are interested in the MLE of $\theta$ and not $\mathbf{p}$ we assume without loss of generality that $\mathbf{p} = (p, ..., p)$. Then Poisson's Binomial cdf reduces to

$$F_M(y \leq (\mathcal{R}-1); \mathbf{p}) = \sum_{y=0}^{\mathcal{R}-1} \sum_{A \in S_y} \left[ \prod_{i \in A} p_i \prod_{i \in A^c} (1 - p_i) \right]. \tag{B.2}$$

$$= \sum_{y=0}^{\mathcal{R}-1} \sum_{A \in S_y} \left[ \prod_{i \in A} p \prod_{i \in A^c} (1 - p) \right] \tag{B.3}$$

$$= \sum_{y=0}^{\mathcal{R}-1} \binom{M}{y} p^y (1-p)^{M-y} \tag{B.4}$$

$$= B(M, \mathcal{R}-1; p), \tag{B.5}$$

and the likelihood in B.1 reduces to

$$\mathcal{L}(p|\mathbf{b}) = \prod_j (1 - B(M, \mathcal{R}-1; p))^{b_j} \cdot B(M, \mathcal{R}-1; p)^{1-b_j}, \tag{B.6}$$

and the expected value to

$$E(Y) = \theta = Mp. \tag{B.7}$$

The likelihood in equation (B.6) is a reparameterized Bernoulli likelihood with parameter $\tilde{p} = B(M, \mathcal{R}-1; p)$. The MLE for $\tilde{p}$ exist and is unique since $\tilde{p}$ is a finite moment of the density. By the invariance-property of the MLE (e.g., Casella and Berger, 2002, Theorem 7.2.10), the MLE is invariant under reparameterization. $B(M, \mathcal{R}-1; p)$ is the cdf of a Binomial density which is injective and consequently the MLE for $p$ is unique. For the same reason, the MLE for $E(Y)$ must exist and is unique. □

The expected value is a reduced-form parameter and, as lemma 1 shows, identified. Thus, identifiability of the structural parameters $\boldsymbol{\beta}$ reduces to the question of a unique solution to system of $J$ nonlinear equation with $K$ unknowns:

$$\begin{bmatrix} \theta_1 = \sum_{i=1}^{M} \Phi(\mathbf{x}_{i1}\boldsymbol{\beta}) \\ \theta_2 = \sum_{i=1}^{M} \Phi(\mathbf{x}_{i2}\boldsymbol{\beta}) \\ \vdots \\ \theta_j = \sum_{i=1}^{M} \Phi(\mathbf{x}_{ij}\boldsymbol{\beta}) \\ \vdots \\ \theta_J = \sum_{i=1}^{M} \Phi(\mathbf{x}_{iJ}\boldsymbol{\beta}). \end{bmatrix} \tag{B.8}$$

Using a Taylor series expansions around 0 of the first order for the system, leads to a system where each row can can be written as:

$$\theta_j = \sum_{i=1}^{M} \left( \frac{1}{2} + \frac{1}{\sqrt{2\pi}}(\beta_0 + \beta_1 x_{ij1} + \ldots + \beta_K x_{ijK}) \right) \tag{B.9}$$

$$= \sum_{i=1}^{M} \left( \frac{1}{2} + \beta_0 \frac{1}{\sqrt{2\pi}} + \beta_1 \frac{1}{\sqrt{2\pi}} x_{ij1} + \ldots + \beta_K \frac{1}{\sqrt{2\pi}} x_{ijK} \right) \tag{B.10}$$

$$= \underbrace{M\left( \frac{1}{2} + \beta_0 \frac{1}{\sqrt{2\pi}} \right)}_{\beta_0'} + \underbrace{\beta_1 \frac{M}{\sqrt{2\pi}}}_{\beta_1'} \left( \sum_{i=1}^{M} x_{ij1} \right) + \ldots + \underbrace{\beta_K \frac{M}{\sqrt{2\pi}}}_{\beta_K'} \left( \sum_{i=1}^{M} x_{ijk} \right) \tag{B.11}$$

$$= \beta_0' + \beta_1' \left( \sum_{i=1}^{M} x_{ij1} \right) + \ldots + \beta_K' \left( \sum_{i=1}^{M} x_{ijK} \right). \tag{B.12}$$

**Proposition 3.** *There exist a unique MLE for $\boldsymbol{\beta}'$ to the system approximated by the Taylor series expansion around 0 of the first order in equation* (B.8)*.*

*Proof.* The system is a system of linear equations which is known to have a unique solution (and consequently a unique MLE) iff the matrix (the Jacobian):

$$\begin{bmatrix} 1 & \sum_{i=1}^{M} x_{i11} & \sum_{i=1}^{M} x_{i12} & \dots & \sum_{i=1}^{M} x_{i1K} \\ 1 & \sum_{i=1}^{M} x_{i21} & \sum_{i=1}^{M} x_{i22} & \dots & \sum_{i=1}^{M} x_{i2K} \\ & \vdots & & & \vdots \\ 1 & \sum_{i=1}^{M} x_{iJ1} & \sum_{i=1}^{M} x_{iJ2} & \dots & \sum_{i=1}^{M} x_{iJK} \end{bmatrix}$$

has full rank. □

One might refer to the matrix as the 'aggregate design matrix' since it results from summing (or averaging) over all members for each decision. Note, I eased the exercise by assuming existence. An existence proof might impose a bounded parameter space and then show continuity. The lemma would allow to invoke the extreme-value theorem (Weierstress theorem) that guarantees existence (e.g., Sundaram, 1996, Chapter 3).

# Appendix C Neglected Heterogeneity

Suppose that the correct model for each member's vote is $\mathbf{y}_j^* = \beta_0 + \beta_1 z_j + \boldsymbol{\epsilon}_j$, where $\beta_1 \neq 0$. Notice that, in this model, all members are very likely to cast the same vote since $z_j$ exhibits no member-specific variation. The degree of correlation among members' vote choices is a function of $\beta_1$ and the variance in $\boldsymbol{\epsilon}_j$. Suppose further that the analyst seeks to test whether $x_{ij}$ has an effect on $\mathbf{y}_j^*$ but happens not to observe $z_j$. The analyst estimates the following m-probit model $\mathbf{y}_j^* = \beta_0' + \beta_2' x_{ij} + \boldsymbol{\epsilon}_j'$, where $\boldsymbol{\epsilon}_j' = \beta_1 z_j + \boldsymbol{\epsilon}_j$. In this model, the vote choices among the members are correlated and the assumption of conditional independence violated.

To the extent that $\boldsymbol{\epsilon}_j'$ is normally distributed (which is ensured when $z_j$ is normally distributed) and independent of $x_{ij}$, $\beta_2'$ is only rescaled relative to $\beta_2$ since $\beta_2' = \beta_2/\tau$, where $\tau = \text{var}(\boldsymbol{\epsilon}_j')$. This neglected heterogeneity problem is well known in the context of probit models (e.g., Wooldridge, 2001, p. 470) but presents no problem in practice since the coefficient's scale does not affect i) the direction of the estimated effects, ii) the test statistics, or iii) the marginal effect estimates. The same results apply to the m-probit: the correlation-inducing variable (the neglected heterogeneity) rescales the coefficient estimates but will not affect their directions, the test statistics, or marginal effect estimates. However, if $z_j$ were correlated with $x_{ij}$, the coefficient estimates would be biased. However, this is the case for any model estimated based on a decision record. This confounding bias can be removed only if a voting record is available to the analyst.

| | | | | | Slope | |
|---|---|---|---|---|---|---|
| $M$ | $\mathcal{R}$ | $J$ | Sim. | Conv. | RMSE | Cover. |
| 5 | 3 | 250 | 250 | 249 (1.00) | 0.09 | 0.95 (0.014) |
| 5 | 4 | 250 | 250 | 247 (0.99) | 0.10 | 0.94 (0.015) |
| 10 | 6 | 250 | 250 | 236 (0.94) | 0.09 | 0.97 (0.010) |
| 10 | 7 | 250 | 250 | 240 (0.96) | 0.09 | 0.96 (0.013) |

**Table SI-1:** Results from 4 Monte Carlo experiments. The first four columns report the number of members (column labeled $M$), the voting rule ($\mathcal{R}$), the number of proposals ($J$) and the number of simulations per experiment (Sim.). The latter three columns report the number and percentage shares of simulations for which the convergence diagnostic supports my choice of run length (Conv.), the RMSE for all converged simulations and coverage probabilities (Cover.) of the 95% posterior intervals (with Monte Carlo standard errors in brackets) for all converged simulations.

The theoretical argument above is based on a distributional assumption about the neglected heterogeneity. It is difficult to evaluate how critical this assumption is, but when I replicate parts of the Monte Carlo experiments (section F) including neglected heterogeneity that is non-normally distributed, the coverage probabilities are still very accurate (see table SI-1).

I used the following data-generating process: $x_{ij}, z_j \sim U(-2, 2)$, $\beta_0, \beta_2 \sim U(-1, 1)$, and $\beta_1 = 0$. The estimated models included only the covariate $x_{ij}$, leaving $z_j$ as the neglected heterogeneity that is uniformly distributed.

# Appendix D Full Conditionals and Gibbs Scheme

## D.1 Proposition from Lauritzen et al. (1990)

Let parents($q$) be a function that collects all nodes that are connected to a node $q$ via an inward edge and children($p$) the function that collects all nodes that are connected via an outward edge to $p$.

**Proposition 4.** *(Lauritzen et al. (1990)) If a joint density can be written as a directed acyclic graph (DAG), the conditional pdf of any of the DAG's nodes $(\theta_1, \ldots.\theta_j, \ldots\theta_J)$ is given by:*

*$(\theta_1, \ldots.\theta_j, \ldots\theta_J)$ is given by:*

$$f(\theta_j|\theta_{\neg j}) \propto f(\theta_j|\text{parents}(\theta_j)) \times \prod_{w \in \text{chidren}(\theta_j)} f(w|\text{parents}(w)), \tag{D.1}$$

*where $\theta_{\neg j}$ denotes all nodes in the DAG other than $\theta_j$.*

*Proof.* See Lauritzen et al. (1990). □

## D.2 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.

$$\begin{aligned} f(\boldsymbol{\beta}|\mathbf{b}_0, \mathbf{B}_0, \mathbf{y}^*, \mathbf{y}, \mathbf{b}, \mathbf{X}) \propto\ & f(\boldsymbol{\beta}|\mathbf{b}_0, \mathbf{B}_0) \times \prod_j f(\mathbf{y}_j^*|\mathbf{X}_j, \boldsymbol{\beta}) \\ & = \boldsymbol{\phi}\Big((\mathbf{B}_0^{-1} + \mathbf{X}'\mathbf{X})^{-1}(\mathbf{B}_0^{-1}\mathbf{b}_0 + \mathbf{X}'\mathbf{y}^*), (\mathbf{B}_0^{-1} + \mathbf{X}'\mathbf{X})^{-1}\Big). \end{aligned} \tag{D.2}$$

B) The full conditional for $\mathbf{y}_j^*$ 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).

$$
\begin{aligned}
f(\mathbf{y}_j^*|\mathbf{b}_0, \mathbf{B}_0, \boldsymbol{\beta}, \mathbf{y}, \mathbf{b}, \mathbf{X}) &\propto f(\mathbf{y}_j^*|\mathbf{X}_j, \boldsymbol{\beta}) \times f(\mathbf{y}_j|\mathbf{y}_j^*) \\
&\propto f(\mathbf{y}_j^*|\mathbf{X}_j, \boldsymbol{\beta}, \mathbf{y}_j) \\
&\propto \boldsymbol{\phi}(\mathbf{X}_j\boldsymbol{\beta}) \prod_i \Big(\mathcal{I}(y_{ij}^* \geq 0)\mathcal{I}(y_{ij} = 1) + \mathcal{I}(y_{ij}^* < 0)\mathcal{I}(y_{ij} = 0)\Big).
\end{aligned}
\tag{D.3}
$$

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

$$
\begin{aligned}
f(\mathbf{y}_j|\mathbf{b}_0, \mathbf{B}_0, \boldsymbol{\beta}, \mathbf{y}, \mathbf{b}, \mathbf{X}) &\propto f(\mathbf{y}_j|\mathbf{y}_j^*) \times f(b_j|\mathbf{y}_j) \\
&\propto f(\mathbf{y}_i|\mathbf{y}_j^*, b_j) \\
&\propto \prod_i \Big(\Phi(y_{ij}^*)^{y_{ij}} + (1 - \Phi(y_{ij}^*))^{1-y_{ij}}\Big) \times \\
&\quad \Big(\mathcal{I}(\sum_i y_{ij} < \mathcal{R})\mathcal{I}(b_j = 0) + \mathcal{I}(\sum_i y_{ij} \geq \mathcal{R})\mathcal{I}(b_j = 1)\Big).
\end{aligned}
\tag{D.4}
$$

## D.3 Sampling Bernoulli Densities with Constraint

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

**Algorithm 1.** 1. Draw from

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

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:

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

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, in practice 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 $b_j$.

However, the algorithm 1 can be inefficient because there might be thousands of draws which are rejected, because the acceptance ratio is 0. In fact, the ratio $f(b_j|\mathbf{y}_j)$ is zero for many $\mathbf{y}_j$. A more efficient version rescales the proposal density every $i^{th}$ iteration. Define $Q$ and $Q_{max}$ s.t. $1 < Q < Q_{max}$ and let there be some constant $\epsilon$. The algorithm for $(b_j = 1)$ takes the following form:

**Algorithm 2.** 1. Set $Q = 1$ and $Q_{max} = 1/\Phi(max(\mathbf{y}_j^*))$

2. Draw from

$$\mathbf{y}_j \sim \begin{cases} Qf(y_{1j}|y_{1j}^*) \\ ... \\ Qf(y_{Mj}|y_{Mj}^*) \end{cases}$$

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

4. Accept $\mathbf{y}_j$ and stop if $u \leq \frac{f(\mathbf{y}_j|\mathbf{y}_j^*) \times f(b_j|\mathbf{y}_j)}{C \times (1/MQ) \times f(\mathbf{y}_j|\mathbf{y}_j^*)}$

5. If $Q < Q_{max}$ set $Q = Q + g(\epsilon, i)$

6. Repeat.

Notice, that $C$ can be chosen such that it offsets $1/MQ$ and $C/MQ \geq 1$. Consequently, the acceptance ratio is again either 0 or 1 and sampling in practice does not require drawing from a uniform. The version for $b_j = 0$ uses $Q_{max} = 1/(1 - \Phi(max(\mathbf{y}_j^*)))$.

An example helps to clarify the intuition for algorithm 2. Suppose that all elements in $\mathbf{y}_j^*$ are quite small, $b_j = 1$ and $\mathcal{R}$ is high. In this case algorithm 1 takes a long time. Algorithm 2 scales the vector $\mathbf{\Phi}(\mathbf{y}_j^*)$ by some constant $Q$ which increases the probability to sample $\mathbf{y}_j$ that obeys the constrain $b_j = 1$. Since the scaling is uniform across the elements of the vector, the target density is not altered. The scaling constant is increased over the course of iterations given some user-defined $\epsilon$. The implementation of this algorithm in the `consilium`-package adds $\epsilon$ according to fixed schedule (default is every $200^{th}$ iteration) with $\epsilon$ small (default is $\epsilon = 0.05$).

## D.4 Gibbs Sampler

Denote the $s^{th}$ draw with superscript $(s)$ then the Gibbs sampler the following form:

**Algorithm 3.** 1. For all $J$ draw vote profiles $\mathbf{y}_j^{(s)}$ from Bernoulli densities with constraint with algorithm 2.

2. Draw for all $j = 1, ..., J$ and $i = 1, ..., M$ from truncated normal densities:

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

3. Draw from a multivariate normal density:

$$\boldsymbol{\beta}^{(s)} \sim \boldsymbol{\phi}(\mathbf{b}_0, \mathbf{B}_1)$$

$$\begin{aligned} \mathbf{b}_0 &= \mathbf{B}_1(\mathbf{B}_0^{-1}\mathbf{b}_0 + \mathbf{X}'\mathbf{y}^{*(s)}) \\ \mathbf{B}_1 &= (\mathbf{B}_0^{-1} + \mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

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

4. Repeat $S$ times until convergence.

## D.5 Extension with Varying Intercept

Let there be $G$ groups for which the are unobserved effects, $\alpha_1, ..., \alpha_g, ...\alpha_G$. A varying intercept version of the likelihood from equation (1.3) then takes the form:

$$\mathcal{L}(\boldsymbol{\beta}|\mathbf{X}, \mathbf{b}) = \prod_j \sum_{\tilde{\mathbf{y}} \in V(b_j)} \Bigg[ \boldsymbol{\Phi}_{\mathcal{P}(\tilde{\mathbf{y}})}(\mathbf{X}_j\boldsymbol{\beta} + \alpha_{g[j]}) \Bigg]. \tag{D.5}$$

In addition to the prior densities over $\boldsymbol{\beta}$ assume:

$$
\begin{aligned}
\alpha_g &\sim N(0, \omega^2) \\
\omega^2 &\sim \text{invGamma}(e_0/2, h_0/2).
\end{aligned}
\tag{D.6}
$$

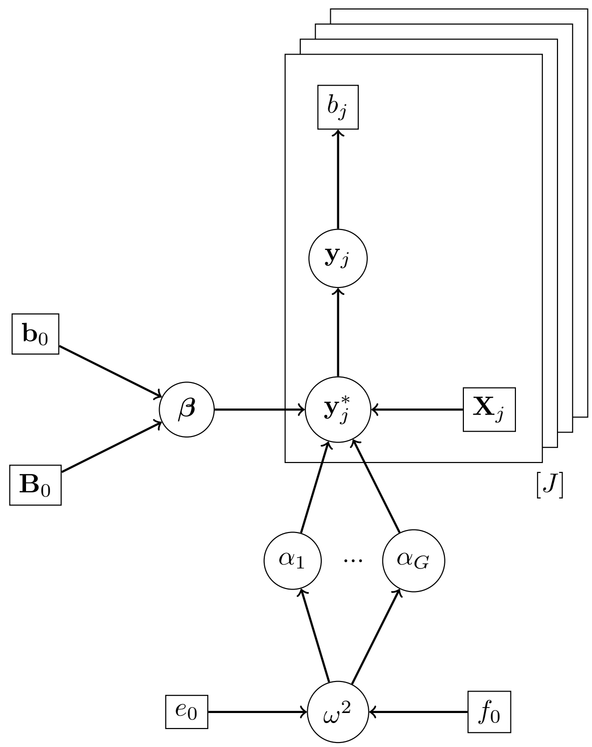



**Figure SI-1:** Directed acyclic graph of the partial m-probit with varying intercept.

An extended version of the graph from figure 1b appears in figure SI-1 suggesting the followig full conditional densities:

$$
\begin{aligned}
&f(\boldsymbol{\beta}|\mathbf{b}_0, \mathbf{B}_0, \mathbf{y}^*, \mathbf{y}, \\
&\mathbf{b}, \mathbf{X}, \boldsymbol{\alpha}, e_0, f_0, \omega^2) \propto \; f(\boldsymbol{\beta}|\mathbf{b}_0, \mathbf{B}_0) \times \prod_j f(\mathbf{y}_j^*|\mathbf{X}_j, \boldsymbol{\beta}, \boldsymbol{\alpha}) \\
&\qquad = \boldsymbol{\phi}\Big((\mathbf{B}_0^{-1} + \mathbf{X}'\mathbf{X})^{-1}(\mathbf{B}_0^{-1}\mathbf{b}_0 + \mathbf{X}'(\mathbf{y}^* - \boldsymbol{\alpha})), (\mathbf{B}_0^{-1} + \mathbf{X}'\mathbf{X})^{-1}\Big),
\end{aligned}
\tag{D.7}
$$

$$
\begin{aligned}
f(\mathbf{y}_j^*|\mathbf{b}_0, \mathbf{B}_0, \boldsymbol{\beta}, \mathbf{y}, \\
\mathbf{b}, \mathbf{X}, \boldsymbol{\alpha}, e_0, f_0, \omega^2) &\propto f(\mathbf{y}_j^*|\mathbf{X}_j, \boldsymbol{\beta}, \boldsymbol{\alpha}) \times f(\mathbf{y}_j|\mathbf{y}_j^*) \\
&\propto f(\mathbf{y}_j^*|\mathbf{X}_j, \boldsymbol{\beta}, \mathbf{y}_j, \boldsymbol{\alpha}) \\
&\propto \boldsymbol{\phi}(\mathbf{X}_j\boldsymbol{\beta} + \boldsymbol{\alpha}) \prod_i \Big(\mathcal{I}(y_{ij}^* \geq 0)\mathcal{I}(y_{ij} = 1) + \mathcal{I}(y_{ij}^* < 0)\mathcal{I}(y_{ij} = 0)\Big),
\end{aligned} \tag{D.8}
$$

$$
\begin{aligned}
f(\mathbf{y}_j|\mathbf{b}_0, \mathbf{B}_0, \boldsymbol{\beta}, \mathbf{y}^*, \\
\mathbf{b}, \mathbf{X}, \boldsymbol{\alpha}, e_0, f_0, \omega^2) &\propto f(\mathbf{y}_j|\mathbf{y}_j^*) \times f(b_j|\mathbf{y}_j) \\
&\propto f(\mathbf{y}_i|\mathbf{y}_j^*, b_j) \\
&\propto \prod_i \Big(\Phi(y_{ij}^*)^{y_{ij}} + (1 - \Phi(y_{ij}^*))^{1-y_{ij}}\Big) \times \\
&\quad \Big(\mathcal{I}(\sum_i y_{ij} < \mathcal{R})\mathcal{I}(b_j = 0) + \mathcal{I}(\sum_i y_{ij} \geq \mathcal{R})\mathcal{I}(b_j = 1)\Big).
\end{aligned} \tag{D.9}
$$

$$
\begin{aligned}
f(\omega^2|\mathbf{b}_0, \mathbf{B}_0, \boldsymbol{\beta}, \mathbf{y}, \mathbf{y}^*, \mathbf{b}, \mathbf{X}, e_0, f_0, \boldsymbol{\alpha}) &\propto f(\omega^2|e_0, h_0) \times \prod_{g=1}^{G} f(\alpha_g) \\
&= \text{invGamma}(e_1/2, h_1/2),
\end{aligned} \tag{D.10}
$$

where $e_1 = e_0 + G$ and $h_1 = h_0 + \sum_{g=1}^{G} \alpha_g^2$.

$$
\begin{aligned}
f(\alpha_g|\mathbf{b}_0, \mathbf{B}_0, \boldsymbol{\beta}, \mathbf{y}, \mathbf{b}, \mathbf{X}, e_0, f_0, \omega^2) &\propto f(\alpha_g|\omega^2) \prod_{i=1}^{N_g} f(y_{gi}^*|\mathbf{x}_{gi}, \boldsymbol{\beta}) \\
&= \boldsymbol{\phi}\big((\bar{\epsilon}_g^* N_g)/(\omega^{-2} + N_g), 1/(\omega^{-2} + N_g)\big),
\end{aligned} \tag{D.11}
$$

where $N_g$ are the number of observations for the $g^{th}$ group, $\mathbf{y}_g^*, \mathbf{X}_g$ are the observations that belong to the $g^{th}$ group and $\bar{\epsilon}_g^* = 1/N_g \sum_{i=1}^{N_g}(y_{gi}^* - \mathbf{x}_{gi}\boldsymbol{\beta})$.

# Appendix E Aggregation Bias

**Proposition 5.** *In the bivariate case with vague priors, the posterior mean for* $\beta$ *(the coefficient) is only a function of* $E(\mathbf{Y}^*)$ *if* $(\mathbf{X}, \mathbf{Y}^*) \perp\!\!\!\perp \mathbf{H}$ *where* $\mathbf{H}$ *is a categorical proposal-assignment vector.*

*Proof.* In the bivariate case with vague priors (zero-centered, large variances), the multivariate normal mean of the full conditional reduces to the simple OLS estimator for $\beta$. Consequently, the proof amounts to showing that the OLS $\beta$ is at most a function of $E(\mathbf{Y}^*)$ (but not $\mathbf{Y}^*$). This has been shown elsewhere and I follow a similar approach (Erbring, 1989; Palmquist, 1993; King, 1997). I use the law of total variance and the law of total covariance.

$$
\begin{aligned}
\beta =& \frac{Cov(\mathbf{X}, \mathbf{Y}^*)}{Var(\mathbf{X})} && \text{(E.1)}\\
=& \frac{Cov(E(\mathbf{X}|\mathbf{H}), E(\mathbf{Y}^*|\mathbf{H})) + E(Cov(\mathbf{X}, \mathbf{Y}^*|\mathbf{H}))}{E(Var(\mathbf{X}|\mathbf{H})) + Var(E(\mathbf{X}|\mathbf{H}))} && \text{(E.2)}\\
=& \frac{Cov(E(\mathbf{X}|\mathbf{H}), E(\mathbf{Y}^*|\mathbf{H}))}{Var(E(\mathbf{X}|\mathbf{H})) + E(Var(\mathbf{X}|\mathbf{H}))} + \frac{E(Cov(\mathbf{X}, \mathbf{Y}^*|\mathbf{H}))}{E(Var(\mathbf{X}|\mathbf{H})) + Var(E(\mathbf{X}|\mathbf{H}))} && \text{(E.3)}\\
=& \frac{Cov(E(\mathbf{X}|\mathbf{H}), E(\mathbf{Y}^*|\mathbf{H}))}{Var(E(\mathbf{X}|\mathbf{H})) + E(Var(\mathbf{X}|\mathbf{H}))} \frac{Var(E(\mathbf{X}|\mathbf{H}))}{Var(E(\mathbf{X}|\mathbf{H}))} + && \text{(E.4)}\\
& \frac{E(Cov(\mathbf{X}, \mathbf{Y}^*|\mathbf{H}))}{E(Var(\mathbf{X}|\mathbf{H})) + Var(E(\mathbf{X}|\mathbf{H}))} \frac{E(Var(\mathbf{X}|\mathbf{H}))}{E(Var(\mathbf{X}|\mathbf{H}))} && \text{(E.5)}\\
=& \frac{Cov(E(\mathbf{X}|\mathbf{H}), E(\mathbf{Y}^*|\mathbf{H}))}{Var(E(\mathbf{X}|\mathbf{H}))} \frac{Var(E(\mathbf{X}|\mathbf{H}))}{Var(E(\mathbf{X}|\mathbf{H})) + E(Var(\mathbf{X}|\mathbf{H}))} + && \text{(E.6)}\\
& \frac{E(Cov(\mathbf{X}, \mathbf{Y}^*|\mathbf{H}))}{E(Var(\mathbf{X}|\mathbf{H}))} \frac{E(Var(\mathbf{X}|\mathbf{H}))}{E(Var(\mathbf{X}|\mathbf{H})) + Var(E(\mathbf{X}|\mathbf{H}))} && \text{(E.7)}\\
=& W(\mathbf{X}, \mathbf{H}) \frac{Cov(E(\mathbf{X}|\mathbf{H}), E(\mathbf{Y}^*|\mathbf{H}))}{Var(E(\mathbf{X}|\mathbf{H}))} + (1 - W(\mathbf{X}, \mathbf{H})) \frac{E(Cov(\mathbf{X}, \mathbf{Y}^*|\mathbf{H}))}{E(Var(\mathbf{X}|\mathbf{H}))} && \text{(E.8)}\\
=& W(\mathbf{X}, \mathbf{H}) \frac{Cov(\overline{\mathbf{X}}, \overline{\mathbf{Y}^*})}{Var(\overline{\mathbf{X}})} + (1 - W(\mathbf{X}, \mathbf{H})) \frac{E(Cov(\mathbf{X}, \mathbf{Y}^*|\mathbf{H}))}{E(Var(\mathbf{X}|\mathbf{H}))} && \text{(E.9)}\\
=& W(\mathbf{X}, \mathbf{H}) \beta_{agg} + (1 - W(\mathbf{X}, \mathbf{H})) \frac{E(Cov(\mathbf{X}, \mathbf{Y}^*|\mathbf{H}))}{E(Var(\mathbf{X}|\mathbf{H}))}. && \text{(E.10)}
\end{aligned}
$$

Assuming $(\mathbf{X}, \mathbf{Y}^*) \perp\!\!\!\perp \mathbf{H}$ ('random' grouping) and rearranging we have:

$$=W(\mathbf{X},\mathbf{H})\beta_{agg} + (1 - W(\mathbf{X},\mathbf{H}))\frac{E(Cov(\mathbf{X},\mathbf{Y}^*|\mathbf{H}))}{E(Var(\mathbf{X}|\mathbf{H}))} \tag{E.11}$$

$$=W(\mathbf{X},\mathbf{H})\beta_{agg} + (1 - W(\mathbf{X},\mathbf{H}))\frac{E(Cov(\mathbf{X},\mathbf{Y}^*))}{E(Var(\mathbf{X}))} \tag{E.12}$$

$$=W(\mathbf{X},\mathbf{H})\beta_{agg} + (1 - W(\mathbf{X},\mathbf{H}))\beta \tag{E.13}$$

$$\beta - (1 - W(\mathbf{X},\mathbf{H}))\beta = W(\mathbf{X},\mathbf{H})\beta_{agg} \tag{E.14}$$

$$\beta = \beta_{agg} = \frac{Cov(\overline{\mathbf{X}},\overline{\mathbf{Y}^*})}{Var(\overline{\mathbf{X}})}. \tag{E.15}$$

□

Let $\mathbf{S}$ be a binary vector indicating if the voting profile for a proposal $j$ is observed ($s_j = 1$) or unobserved ($s_j = 0$). Let $\beta$ be the parameter from the likelihood as defined in (1.3) with $(\mathbf{X},\mathbf{Y}^*) \perp\!\!\!\perp \mathbf{H}$ and let $\dot{\beta}$ be the parameter from the likelihood as defined in (4.3) with $(\mathbf{X},\mathbf{Y}^*) \not\perp\!\!\!\perp \mathbf{H}$.

**Proposition 6.** *If* $(\mathbf{X},\mathbf{Y}^*) \perp\!\!\!\perp \mathbf{S}$ *then* $\dot{\beta} = \beta$.

*Proof.* The full conditional for $\dot{\beta}$ is the same as the full conditional for $\beta$. We have already shown in the proof for proposition 5:

$$\dot{\beta} = W(\mathbf{X},\mathbf{H})\dot{\beta}_{agg} + (1 - W(\mathbf{X},\mathbf{H}))\frac{E(Cov(\mathbf{X},\mathbf{Y}^*|\mathbf{H}))}{E(Var(\mathbf{X}|\mathbf{H}))} \tag{E.16}$$

which does not reduce to $\dot{\beta} = \dot{\beta}_{agg}$ since $(\mathbf{X},\mathbf{Y}^*) \not\perp\!\!\!\perp \mathbf{H}$. But,

$$\frac{E(Cov(\mathbf{X},\mathbf{Y}^*|\mathbf{H}))}{E(Var(\mathbf{X}|\mathbf{H}))} = \frac{E(Cov(\mathbf{X},\mathbf{Y}^*|\mathbf{H},\mathbf{S}=1)) + E(Cov(\mathbf{X},\mathbf{Y}^*|\mathbf{H},\mathbf{S}=0))}{E(Var(\mathbf{X}|\mathbf{H},\mathbf{S}=1)) + E(Var(\mathbf{X}|\mathbf{H},\mathbf{S}=0))} \tag{E.17}$$

$$\frac{E(Cov(\mathbf{X},\mathbf{Y}^*|\mathbf{H}))}{E(Var(\mathbf{X}|\mathbf{H}))} = \frac{E(Cov(\mathbf{X},\mathbf{Y}^*|\mathbf{H},\mathbf{S}=1))}{E(Var(\mathbf{X}|\mathbf{H},\mathbf{S}=1))} \tag{E.18}$$

which implies $\dot{\beta} = \beta$. □

# Appendix F Monte Carlo Experiments

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 members (5, 10, 50, 100) and the voting rule (simple majority, $\frac{2}{3}$ supermajority). Each member's vote choice is governed by two variables: a constant $x_{0ij} = 1$ and the 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 decision record exhibits less than 5% of either zeros or ones, that is, if there is not a minimum amount of variation in the dependent variable, I discard the simulated data and repeat the simulation.

I use vague priors ($b_0 = 0$, $B_0 = 100$) and rely on pretests to calibrate the Gibbs sampler's run length[1]. For all conditions, I record the Gelman-Rubin convergence diagnostic (Gelman and Rubin, 1992), the root-mean-square error (RMSE) between the true coefficient values, and the posterior means as well as the coverage rate with the Monte Carlo standard error. If the Gibbs sampler works as expected, the RMSE should be close to zero and approximately 95% of the true coefficient values should be covered by the 95% posterior interval.

Table SI-3 summarizes the results of the 16 experiments. Taking into account the Monte Carlo standard error, the coverage probabilities are accurate and the RMSE is, as expected, very low. This suggests that the Gibbs sampler and its implementation work as expected and recover the true coefficient values. Figure SI-2 illustrates the results from one of the experiments (10 members, 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 and the right panel the slope coefficient. The circles indicate the posterior means for which the Gelman-Rubin convergence diagnostic does not support my choice of run length.

In figure SI-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 estimate only the partial m-probit if the decision record exhibits a minimum amount of variation. While my 5% cutoff was

[1] 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. I run two chains sequentially using distinct seeds and starting values.

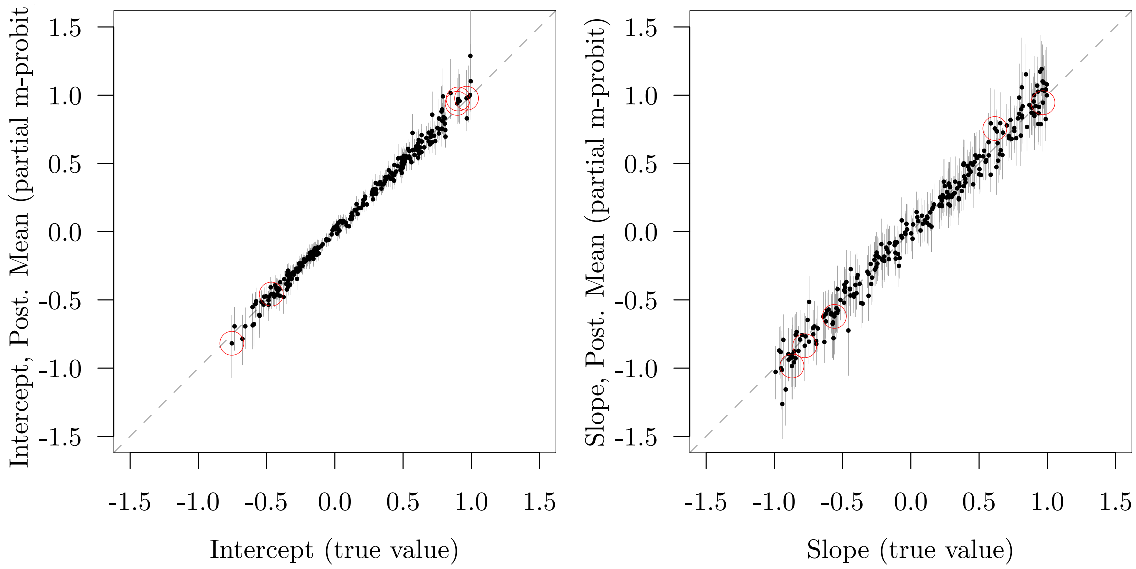



**Figure SI-2:** Results from one of the Monte Carlo experiments (10 members, majority rule, 500 proposals): Scatter plot of posterior means with 95% posterior intervals from the partial m-probit and true coefficient values. The circles indicate the parameters for which the Gelman-Rubin convergence diagnostic does not support my choice of run length 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.

arbitrary, the effect reveals a general subtle point: aggregation reduces information potentially up to a point where no variation is left in the decision record (see also section 3 in the main text).

Table SI-3 shows the approximate computation time used for one simulation in each of the 16 experimental conditions and the number of converged simulations. Generally, the computation time increases with the number of members and the sample size. While all models require more time than an ordinary probit model, even for large committees (100 members), the computational time is still acceptable (1.30h). From the limited simulations, it appears that the convergence speed of the Gibbs sampler depends on the number of members and the voting rule.

To provide some intuition about the increase in posterior uncertainty that comes with aggregation, I estimate a series of probit models on the simulated vote-choice data from the Monte Carlo experiments discussed in section F. Across the simulations, the 95% posterior intervals from the partial m-probit are considerably larger than the probit intervals. Table SI-2 summarizes the 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 are primarily a function of the numbers of members. For the slope interval, the relative difference for the intervals decreases from 34% (for five members) to 7% (for 100 members), which highlights the severe increase in posterior uncertainty that comes with aggregation.

| | | | | 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 SI-2:** Results from 16 Monte Carlo experiments. The number of members (column labeled $M$), 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).

| | | | | | Intercept | | | Slope | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $M$ | $\mathcal{R}$ | $J$ | Sim. | Time | Conv. | RMSE | Cover. | Conv. | RMSE | Cover. |
| 5 | 3 | 500 | 250 | 5.5 | 250 (1.00) | 0.05 | 0.94 (0.015) | 250 (1.00) | 0.09 | 0.95 (0.014) |
| 5 | 4 | 500 | 250 | 5.4 | 250 (1.00) | 0.05 | 0.96 (0.012) | 249 (1.00) | 0.08 | 0.94 (0.016) |
| 10 | 6 | 500 | 250 | 10.3 | 245 (0.98) | 0.05 | 0.96 (0.013) | 245 (0.98) | 0.09 | 0.96 (0.013) |
| 10 | 7 | 500 | 250 | 10.2 | 248 (0.99) | 0.04 | 0.95 (0.014) | 246 (0.98) | 0.09 | 0.94 (0.015) |
| 50 | 26 | 500 | 250 | 45.4 | 225 (0.90) | 0.02 | 0.96 (0.012) | 194 (0.78) | 0.09 | 0.93 (0.018) |
| 50 | 33 | 500 | 250 | 46.5 | 202 (0.81) | 0.03 | 0.95 (0.016) | 182 (0.73) | 0.09 | 0.95 (0.016) |
| 100 | 51 | 500 | 250 | 89.5 | 213 (0.85) | 0.01 | 0.93 (0.017) | 145 (0.58) | 0.09 | 0.95 (0.018) |
| 100 | 67 | 500 | 250 | 91.2 | 168 (0.67) | 0.04 | 0.94 (0.018) | 152 (0.61) | 0.10 | 0.93 (0.020) |
| 5 | 3 | 250 | 250 | 3.8 | 250 (1.00) | 0.08 | 0.96 (0.012) | 250 (1.00) | 0.13 | 0.95 (0.014) |
| 5 | 4 | 250 | 250 | 3.2 | 250 (1.00) | 0.07 | 0.93 (0.016) | 250 (1.00) | 0.12 | 0.96 (0.012) |
| 10 | 6 | 250 | 250 | 5.5 | 245 (0.98) | 0.09 | 0.89 (0.020) | 245 (0.98) | 0.16 | 0.91 (0.019) |
| 10 | 7 | 250 | 250 | 5.3 | 249 (1.00) | 0.06 | 0.95 (0.014) | 247 (0.99) | 0.12 | 0.96 (0.012) |
| 50 | 26 | 250 | 250 | 25.0 | 218 (0.87) | 0.03 | 0.97 (0.012) | 186 (0.74) | 0.14 | 0.94 (0.017) |
| 50 | 33 | 250 | 250 | 23.3 | 199 (0.80) | 0.05 | 0.95 (0.016) | 184 (0.74) | 0.13 | 0.97 (0.013) |
| 100 | 51 | 250 | 250 | 43.3 | 207 (0.83) | 0.02 | 0.98 (0.011) | 145 (0.58) | 0.11 | 0.95 (0.018) |
| 100 | 67 | 250 | 250 | 49.1 | 148 (0.59) | 0.06 | 0.93 (0.022) | 127 (0.51) | 0.14 | 0.95 (0.019) |

**Table SI-3:** Results from 16 Monte Carlo experiments. The first five columns report the number of members (column labeled $M$), the voting rule ($\mathcal{R}$), the number of proposals ($J$), the number of simulations per experiment (Sim.), computation time in minutes (Time) for one simulation with an Intel Xeon CPU, 2.8GHz. The latter six columns report the number and percentage shares of simulations for which the convergence diagnostic supports my choice of run length (Conv.), the RMSE for all converged simulations and coverage probabilities (Cover.) of the 95% posterior intervals (with Monte Carlo standard errors in brackets) for all converged simulations.

# Appendix G  Data Description

The original unbalanced panel data contain 1,185 observations (Hultman, 2013). The unit of analysis is a conflict-year between 1989 and 2006. The dataset is based on the UCDP conflict dataset (Gleditsch et al., 2002) which uses a low threshold of 25 annual battle deaths as a criterion to identify conflicts.

I modified the original dataset by dropping a.) all conflict-years coded as located in countries that are not members of the system as defined by the Correlates of War Project (Georgia 1990, Croatia 1991, Bosnia and Herzegovina 1991); b) all conflicts that are located in the territory of a permanent member (seven conflicts), and c) the first period of each conflict since I include lagged independent variables.

My dependent variable is the *initial* onset of a UN operation. This deviates from the onset variable in the original data, which encodes *any* onset of a UN operation. All observations of a conflict after the onset are dropped ($n = 133$). Table SI-4 provides an overview on the variance of deployments per period.

Covariates:

- log(Trade) from Barbieri et al. (2009): Logarithm of the total trade between conflict location and a Council member.
- OSV from Hultman (2013): Total number of victims of one-sided violence.
- log(Battle Deaths) from Hultman (2013): Total number of battle deaths.
- log(Army Size) from the Correlates of War Project (2010): Total government army size (Dataset Version: `4.0`). Categorical versions:
- Polity IV: Categorical Polity IV score from Marshall and Jaggers (2002) (Dataset Version: `polity4v2014`). Categorical versions:
- Peace Treaty from Högbladh (2012): Peace treaty signed by the belligerents.
- Non-UN Ops. from Hultman (2013): Deployment of other non-UN operation.
- Border with Member from the Correlates of War Project (2003): At least one nonpermanent or permanent member of the Council shares a land or river border with the conflict location (Dataset Version: `3.1`).

Note that log(Army Size) and Polity IV exhibit some missing values for 8 (4) countries. I use linear interpolation to fill in missing values.

Auxiliary data used: UN Security Council membership data (Dreher et al., 2009), system-membership and country population data by (Correlates of War Project, 2011, 2010).

| Period | N | Deploy. |
|---|---|---|
| 1 | - | |
| 2 | 109 | |
| 3 | 109 | 5 |
| 4 | 101 | 4 |
| 5 | 88 | 3 |
| 6 | 67 | |
| 7 | 56 | |
| 8 | 51 | |
| 9 | 48 | 2 |
| 10 | 40 | |
| 11 | 37 | 1 |
| 12 | 31 | |
| 13 | 29 | |
| 14 | 28 | |
| 15 | 25 | 1 |
| 16 | 24 | |
| 17 | 23 | 1 |
| 18 | 19 | |

**Table SI-4:** Number of deployments and observations per period.

# Appendix H Estimates

| | Model 1 | Model 2 |
|---|---|---|
| (Intercept) | −0.68 | −0.51 |
| | [−1.08; −0.26] | [−1.19; 0.10] |
| Nonpartisan election | 0.52 | 0.22 |
| | [0.27; 0.77] | [−0.12; 0.57] |
| Justice's party aligned pub. opinion | 0.23 | 0.36 |
| | [0.01; 0.45] | [−0.12; 0.86] |
| Election in 2 years | 0.13 | 0.19 |
| | [−0.10; 0.37] | [−1.15; 1.41] |
| Facts aligned pub. opinion | 0.47 | 0.11 |
| | [0.24; 0.71] | [−0.20; 0.41] |
| Trespassing/Protests | 0.42 | 0.34 |
| | [0.08; 0.77] | [−0.10; 0.81] |
| Minors | 0.46 | 0.15 |
| | [0.09; 0.84] | [−0.38; 0.72] |
| Personhood | −0.27 | −0.08 |
| | [−0.63; 0.11] | [−0.55; 0.38] |
| Pub. opinion intensity | 0.12 | 0.05 |
| | [0.01; 0.23] | [−0.09; 0.19] |
| Num. obs | 605 | 85 |

**Table SI-5:** Regression results for US supreme courts application. Bayesian probit model (model 1) and Bayesian partial m-probit model (model 2), each with posterior means and 95% posterior intervals in parentheses. Model 1 estimated using the Gibbs sampler from the `MCMCpack` package (Martin et al., 2011) and model 2 using the `consilium` package. For both models I run two chains, with 11,000 (model 1) and 86,000 (model 2) iterations. The first 1,000 (model 1) and 6,000 iterations are discarded as burn-in. The Gelman and Rubin (1992) convergence diagnostic supports my choice of run length and visual inspection of the chains show no signs of non-convergence.

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 | Model 8 |
|---|---|---|---|---|---|---|---|---|
| log(Trade) | −1.09 | −1.02 | | | | | | |
| | [−2.61; 0.15] | [−3.05; 0.66] | | | | | | |
| P5: log(Trade) | | | −1.40 | | | | | |
| | | | [−3.55; 0.26] | | | | | |
| US: log(Trade) | | | | −1.02 | | | | |
| | | | | [−3.02; 0.71] | | | | |
| UK: log(Trade) | | | | | −2.97 | | | |
| | | | | | [−6.36; −0.50] | | | |
| FR: log(Trade) | | | | | | −0.15 | | |
| | | | | | | [−1.60; 1.63] | | |
| RU: log(Trade) | | | | | | | −2.09 | |
| | | | | | | | [−4.75; −0.24] | |
| CH: log(Trade) | | | | | | | | −1.31 |
| | | | | | | | | [−3.34; 0.41] |
| OSV | −0.01 | −0.08 | −0.07 | −0.08 | −0.06 | −0.07 | −0.08 | −0.07 |
| | [−0.30; 0.23] | [−0.66; 0.35] | [−0.61; 0.36] | [−0.64; 0.35] | [−0.65; 0.42] | [−0.63; 0.36] | [−0.63; 0.34] | [−0.58; 0.37] |
| log(Battle Deaths) | 0.49 | 1.02 | 1.14 | 1.12 | 1.21 | 1.02 | 1.21 | 1.05 |
| | [−0.01; 1.07] | [0.04; 2.42] | [0.12; 2.64] | [0.15; 2.45] | [0.09; 2.83] | [0.11; 2.40] | [0.16; 2.72] | [0.11; 2.37] |
| log(Army Size) | 0.08 | 0.16 | 0.22 | 0.04 | 1.02 | −0.32 | 0.23 | 0.20 |
| | [−1.08; 1.35] | [−2.09; 2.63] | [−2.00; 3.27] | [−2.10; 2.76] | [−1.71; 4.47] | [−2.45; 2.39] | [−2.27; 3.35] | [−2.00; 3.02] |
| Non-UN Ops. | 1.01 | 1.82 | 1.82 | 1.69 | 1.96 | 1.65 | 2.08 | 1.78 |
| | [0.38; 1.75] | [0.71; 3.13] | [0.69; 3.22] | [0.60; 3.01] | [0.74; 3.57] | [0.63; 2.82] | [0.85; 3.59] | [0.70; 3.03] |
| Border with Member | −0.59 | −1.05 | −1.54 | −1.47 | −2.31 | −1.19 | −1.31 | −1.20 |
| | [−1.40; 0.03] | [−2.57; 0.04] | [−5.78; 1.50] | [−5.38; 1.59] | [−7.31; 1.40] | [−4.95; 1.55] | [−6.38; 2.23] | [−5.01; 1.71] |
| Polity IV | −0.16 | −0.36 | −0.19 | −0.20 | −0.08 | −0.36 | −0.24 | −0.31 |
| | [−1.13; 0.84] | [−2.25; 1.79] | [−2.02; 1.80] | [−2.04; 1.93] | [−2.28; 2.34] | [−2.14; 1.68] | [−2.35; 2.09] | [−2.24; 1.77] |
| Alliance with Member | 0.55 | 0.78 | 0.47 | 0.21 | 0.21 | −0.02 | 0.16 | −0.05 |
| | [−2.19; 4.38] | [−0.98; 2.55] | [−2.21; 3.07] | [−2.47; 2.70] | [−3.19; 3.32] | [−2.89; 2.22] | [−3.05; 3.06] | [−2.76; 2.32] |
| Peace Treaty | 0.55 | 1.04 | 0.84 | 0.91 | 0.66 | 0.88 | 0.78 | 0.75 |
| | [−0.09; 1.28] | [−0.06; 2.28] | [−0.46; 2.11] | [−0.28; 2.18] | [−0.75; 2.01] | [−0.16; 2.00] | [−0.55; 2.07] | [−0.55; 1.98] |
| (Intercept) | −1.93 | −7.55 | −8.28 | −7.91 | −10.14 | −7.34 | −9.79 | −7.92 |
| | [−3.84; −0.38] | [−14.89; −2.79] | [−17.67; −3.45] | [−15.71; −3.19] | [−19.18; −3.80] | [−15.82; −2.82] | [−19.75; −3.44] | [−15.82; −3.16] |
| Varying Intercept | | | | | | | | |
| Country, $n = 62$ | 0.86 | 3.16 | 3.38 | 3.29 | 3.98 | 3.16 | 3.99 | 3.15 |
| B-Splines ($df. = 3$) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Num. obs | (15 × 885) | 885 | 885 | 885 | 885 | 885 | 885 | 885 |

**Table SI-6:** Regression results from Bayesian partial m-probit model (model 1) and Bayesian probit model (model 2-8) each with posterior means and 95% posterior intervals in parentheses. Model 1 estimated using the Gibbs sampler from the `consilium` package and model 2-8 the `rstanarm` package (Stan Development Team, 2015). For both models I run two chains, with 601,000 (model 1) and 2,000 (model 2-8) iterations. The first 1,000 are discarded as burn-in. The Gelman and Rubin (1992) convergence diagnostic supports my choice of run length and visual inspection of the chains show little signs of non-convergence.

# References

Barbieri, K., O. M. G. Keshk, and B. M. Pollins (2009). Trading Data Evaluating Our Assumptions and Coding Rules. *Conflict Management and Peace Science 26*(5), 471–491.

Casella, G. and R. L. Berger (2002). *Statistical Inference* (2nd ed.). Pacific Grove: Duxbury.

Correlates of War Project (2003). Colonial/Dependency Contiguity Data, 1816-2002. Version 3.0.

Correlates of War Project (2010). National Material Capabilities Data Documentation. Version 4.0.

Correlates of War Project (2011). State System Membership List. Version 2011.

Dreher, A., J.-E. Sturm, and J. R. Vreeland (2009). Development Aid and International Politics: Does Membership on the UN Security Council Influence World Bank Decisions? *Journal of Development Economics 88*(1), 1–18.

Erbring, L. (1989). Individuals Writ Large: An Epilogue on the "Ecological Fallacy". *Political Analysis 1*(1), 235–269.

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

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.

Gleditsch, N. P., P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand (2002). Armed Conflict 1946-2001: A New Dataset. *Journal of Peace Research 39*(5), 615–637.

Högbladh, S. (2012). Peace Agreements 1975-2011 - Updating the UCDP Peace Agreement Dataset. Department of Peace and Conflict Research Report 99, Uppsala University.

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

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

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.

Marshall, M. G. and K. Jaggers (2002). Polity IV Project: Political Regime Characteristics and Transitions, 1800-2002. Dataset manual, Center for Systemic Peace.

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.

Palmquist, B. L. (1993). *Ecological Inference, Aggregate Data Analysis of U.S. Elections, and the Socialist Party of America.* Ph. D. thesis, University of California, Berkeley.

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

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

Stan Development Team (2015). STAN: A C++ Library for Probability and Sampling, Version 2.8.

Sundaram, R. (1996). *A First Course in Optimization Theory.* Cambridge: Cambridge University Press.

Wang, Y. H. (1993). On the Number of Successes in Independent Trials. *Statistica Sinica 3*(2), 295–312.

Wooldridge, J. M. (2001). *Econometric Analysis of Cross Section and Panel Data.* Cambridge: MIT Press.