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 θ is unique.*

Proof. Since we are interested in the MLE of θ 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 β_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}, β_2' is only rescaled relative to β_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 $(\theta_1, \dots.\theta_j, \dots\theta_J)$ is given by:*

$(\theta_1, \dots.\theta_j, \dots\theta_J)$ 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 θ_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 ϵ. 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 ϵ. The implementation of this algorithm in the `consilium`-package adds ϵ according to fixed schedule (default is every 200^{th} iteration) with ϵ 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:

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

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

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

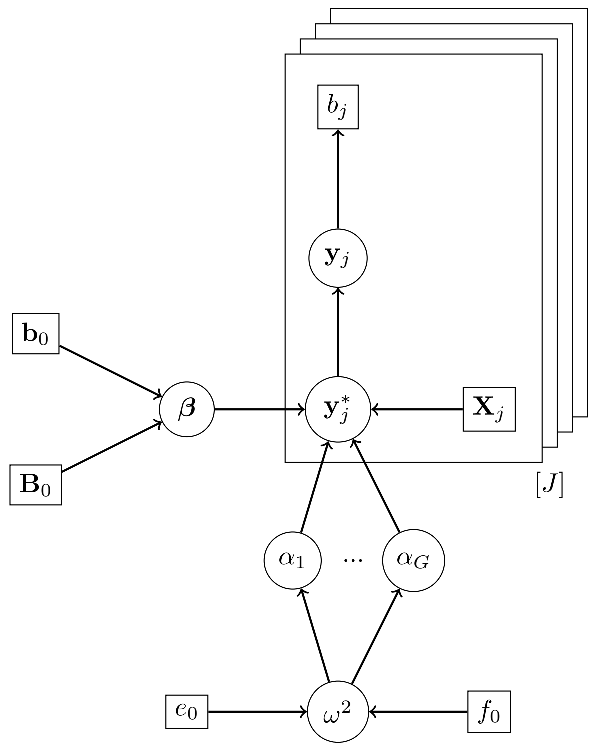

Figure SI-1: Directed acyclic graph of the partial m-probit with varying intercept.

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 β (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 β. Consequently, the proof amounts to showing that the OLS β 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 β 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 β. 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.

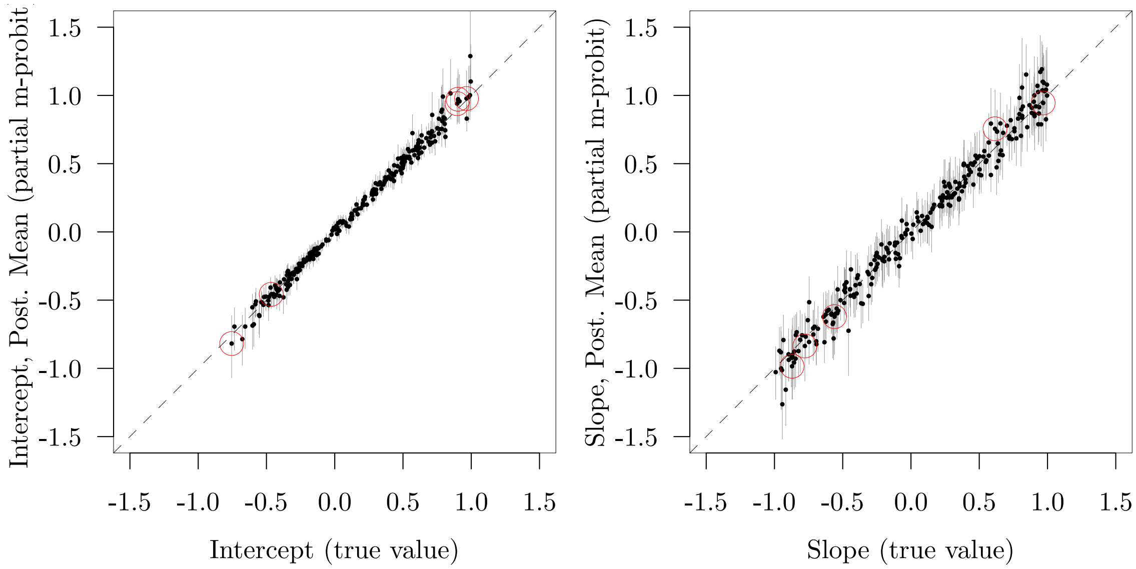

Figure SI-2: Results from one of the Monte Carlo experiments (10 members, majority rule, 500 proposals): Scatter plot of posterior means with 95% posterior intervals from the partial m-probit and true coefficient values. The circles indicate the parameters for which the Gelman-Rubin convergence diagnostic 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.