Web Appendix for Reed, Clark, Nordstrom and Hwang "War, Power, and Bargaining"

The Spatial/Item Response Theory (IRT) Model and Diagnostics for Estimated Ideal Points

To construct a latent measure of state preferences based on UN roll call votes, we use an item response theory (IRT) model (e.g., Jackman 2001; Clinton, Jackman, and Rivers 2004). The IRT model is identical to the spatial voting model for roll call analysis. The spatial model assumes that countries vote for the location giving them greatest utilities with standard quadratic-loss utilities:

$U_i(\zeta_j) = -\|X_i - \zeta_j\|^2 + \eta_{ij}$ and $U_i(\psi_j) = -\|X_i - \psi_j\|^2 + \nu_{ij}$, where η_{ij} and ν_{ij} are zero-mean normal disturbances and $\|\cdot\|$ is the Euclidean norm. With a one-dimensional model, the roll call voting analysis is expressed as:

$$y_{ij}^* = U_i(\zeta_j) - U_i(\psi_j) = X_i\beta_j - \alpha_j + \varepsilon_{ij} \quad (1)$$

$$y_{ij} = 1 \text{ if } y_{ij}^* > 0$$

$$y_{ij} = 0 \text{ otherwise}$$

, where y_{ij}^* is a choice between a "Yes" position, ζ_j, and a "No" position, ψ_j, for each state i in each vote j. With the assumption of utility maximization, $y_{ij} = 1$ if $y_{ij}^* > 0$, $y_{ij} = 0$ otherwise. In this one-dimensional context, Equation (1) for y_{ij}^* is a linear function of the unobserved ideal points X_i with unknown parameters β_j and α_j. We assume $\varepsilon_{ij} \sim N(0,1)$. X_i is a (n × 1) matrix of ideal points, β_j is a (1 × m) matrix of discrimination parameters, and is α_j an m-vector of intercepts. To identify ideal points, we assign uniform priors (1.0, -1.0) to the density over the ideal points. These uniform priors result in ideal points located within the intervals of 1.0 and –1.0. In the mean time, to identify discrimination parameters, we assign normal priors with mean zeroes and the variance prior of 20. Since the variance prior is diffuse, it cannot dominate our estimators.

To obtain these parameters, we use a hierarchical probit model estimated by a Bayesian method. We utilize truncated normal sampling to operationalize the probit model (negative and positive infinity are operationalized as –10 and +10, respectively). Using Markov chain Monte Carlo (MCMC) methods, we generate a large number of samples from the joint posterior density of the parameters and obtain the summary statistics used for inference. We let the Gibbs sampler run for between 50,000 and 500,000 iterations. The 10,000 samples from every 5th iteration in the last 50,000 iterations (or the last 10,000 samples in some cases) were saved for inference.[1]

To this end, we used the WINBUGS software program, an interactive windows version of the BUGS program for Bayesian analysis. This program (available from http://www.mrc-bsu.cam.ac.uk/bugs/) allows complex statistical models to be estimated using MCMC techniques.

Table A.1 provides classification and discrimination statistics for each year of ideal point estimation based on UNGA roll call votes, for the binary scale ideal points used in Models 1 and 2 of the paper.

Table A.1. Model Fit Diagnostics: Classification and Discrimination by Year

Year	Goodness of fit (%) [a]	Discrimination Rate (%)[b]	Year	Goodness of fit (%) [a]	Discrimination Rate (%) [b]
1946	86.78 [85.77, 87.66]	82.5	1975	96.15 [95.60, 96.67]	90.0
1947	88.65 [87.51, 89.71]	88.9	1976	99.16 [98.61, 99.64]	96.3
1948	90.04 [88.99, 90.97]	87.7	1977	97.59 [97.15, 97.94]	93.3
1949	89.53 [88.73, 90.35]	92.8	1978	95.92 [95.54, 96.34]	86.0

[1] The decision to obtain samples was based on technical considerations. Given the long burin-in period, this method has no effect on the results.

1950	92.00 [91.01, 92.88]	93.3
1951	88.75 [85.81, 91.42]	83.3
1952	93.25 [92.32, 94.06]	91.7
1953	92.16 [90.88, 93.41]	95.0
1954	91.16 [89.74, 92.44]	87.5
1955	87.32 [85.77, 89.08]	95.8
1956	97.16 [96.61, 97.74]	100
1957	94.23 [93.54, 94.90]	97.5
1958	97.14 [96.37, 97.88]	96.0
1959	91.49 [90.78, 92.23]	100
1960	93.12 [92.61, 93.63]	100
1961	92.52 [92.05, 92.98]	97.6
1962	93.36 [92.75, 93.95]	100
1963	96.17 [95.34, 96.87]	100
1965	95.38 [94.58, 96.16]	95.7
1966	92.97 [92.13, 93.82]	97.0
1967	93.61 [92.97, 94.23]	97.3
1968	91.91 [91.22, 92.60]	95.2
1969	91.17 [90.19, 92.10]	92.6
1970	94.87 [94.34, 95.36]	95.5
1971	94.24 [93.82, 94.63]	94.1
1972	91.8 [91.25, 92.32]	80.4
1973	94.24 [93.65, 94.83]	93.2
1974	94.33 [93.88, 94.75]	88.1
1979	96.77 [96.47, 97.07]	98.5
1980	97.32 [97.00, 97.66]	92.0
1981	94.99 [94.61, 95.32]	82.3
1982	95.58 [95.28, 95.87]	92.3
1983	96.50 [96.20, 96.82]	88.9
1984	96.65 [96.47, 96.83]	87.9
1985	96.80 [96.58, 97.03]	88.7
1986	96.74 [96.49, 96.97]	88.9
1987	97.50 [97.18, 97.82]	91.9
1988	98.28 [97.88, 98.46]	94.7
1989	97.60 [97.29, 97.90]	96.3
1990	98.85 [98.55, 99.12]	95.5
1991	98.28 [97.85, 98.70]	100
1992	97.95 [97.45, 98.39]	100
1993	97.33 [96.73, 97.88]	100
1994	97.15 [96.53, 97.69]	100
1995	97.28 [96.92, 97.60]	94.6
1996	97.53 [96.96, 98.04]	100
1997	96.81 [96.25, 97.30]	88.9
1998	97.21 [96.67, 97.73]	100
1999	97.37 [96.85, 97.85]	95
2000	97.5 [97.01, 97.94]	100
Average	94.74	93.72

[a]. The percentage of correct predictions (PCP) of the model is used. The mean and its 95 percentile values are reported.

[b]. The rate of discrimination shows whether a one-dimensional model is a reasonable fit to the data. Given relatively high rates of discrimination over the time period, it is

safe to say that states' voting behavior is well discriminated based on a one-dimensional model.