APPENDIXES

A.1 Sampling Algorithm

Here we describe the Markov chain Monte Carlo (MCMC) algorithm used to compute marginal posterior distributions of the model parameters. For convenience, we reiterate some definitions.

Let n_i is the number of parties in a unit of analysis $i = 1, ..., N$. Let Z be a $N \times D$ matrix of latent variables obtained by stacking vectors $\mathbf{z}_i$. Similarly, let Y be a $N \times D$ matrix of observed data, where $y_{ij} = z_{ij}$ if $j < n_i$ and $y_{ij} = \text{NA}$ otherwise. Here, $D = \max(n_i) - 1$, is the number of multivariate t regressions estimated in the model. Lastly, let $\mathbf{X}$ be a $N \times K$ design matrix. The sampling algorithm alternates between four blocks and uses Gibbs sampler with a Metropolis step:

$$Z|Y, n, \boldsymbol{\gamma}, \Sigma, \boldsymbol{\tau} \tag{A.1}$$

$$\boldsymbol{\gamma}, \Sigma|Z, m, t, \boldsymbol{\tau} \tag{A.2}$$

$$m, t, s|\boldsymbol{\gamma}, \Sigma \tag{A.3}$$

$$\tau, \nu|Z, \boldsymbol{\gamma}, \Sigma \tag{A.4}$$

Conditional on Z, parameters $\boldsymbol{\gamma}$ and Σ are independent of the incompletely observed Y; this enables us to the Gibbs sampler (**?**). To sample t-distributed random variables, we employ a well-known result that a multivariate t variable can be expressed as a scale-mixture of multivariate normal variables. In particular, if Z^* is a standard multivariate normal and if $\tau \sim \mathcal{G}(\nu/2, \nu/2)$ then, $Z = Z^*/\tau^{1/2}$ is a multivariate t variable with ν degrees of freedom. If one prefers to use the Gaussian latent model, then one should set $\tau_i = 1$ for all i and simply skip the step A.4.

To implement the step A.1, first define

$$\Sigma = \begin{bmatrix} \Sigma_{LL} & \Sigma_{LD} \\ \Sigma_{DL} & \Sigma_{DD} \end{bmatrix} \tag{A.5}$$

where $\Sigma_{LL} = \Sigma_{[1:L,1:L]}$ and $\Sigma_{DD} = \Sigma_{[L+1:D,L+1:D]}$. Let $\boldsymbol{\mu}_i = (\mu_1, ..., \mu_D) = (\mathbf{x}_i'\gamma_1, ..., \mathbf{x}_i'\gamma_D)$ and let $\boldsymbol{\mu}_{l:p} = (\mu_{il}, ..., \mu_{ip})$ for any $l < p$. From the definition of the model in (??) - (??) and the multivariate normal theory, we have for each $i = 1, ..., N$

$$(z_{i1}, ..., z_{ij}, ..., z_{iD})\Big|\mathbf{y}_i, n_i, \boldsymbol{\gamma}, \Sigma \sim \begin{cases} \mathbf{1}_{(\mathbf{Z}_i = \mathbf{y}_i)} & \text{if } j < n_i \\ \mathcal{N}\big(\overline{\boldsymbol{\mu}}_i, \overline{\Sigma}/\tau_i\big) & \text{if } j \geq n_i, \end{cases} \tag{A.6}$$

where $\overline{\boldsymbol{\mu}}_i = \boldsymbol{\mu}_{n_i:D} + \Sigma_{DL}\Sigma_{DD}^{-1}(\mathbf{y}_{1:n_i} - \boldsymbol{\mu}_{1:n_i})$ and $\overline{\Sigma} = \Sigma_{DD} - \Sigma_{DL}\Sigma_{LL}^{-1}\Sigma_{LD}$. Thus, each Z_{ij} is replaced with Y_{ij} whenever the latter is observed, while the remaining Z_{ij}'s are sampled from the corresponding scaled *conditional* multivariate normal distributions.

Second, the parameters $\boldsymbol{\gamma}$ are sampled from the multivariate Gaussian distribution:

$$\boldsymbol{\gamma}|\Sigma, \mathbf{m}, \mathbf{t}, \boldsymbol{\tau} \sim \mathcal{N}\Big(\mathbf{C}_0^{-1}\big(\mathbf{X}'(\Sigma^{-1} \otimes diag(\boldsymbol{\tau})\big)Z + \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\gamma}_0\big), \mathbf{C}_0^{-1}\Big) \tag{A.7}$$

where $\mathbf{C}_0 = \mathbf{X}'\Big(\Sigma^{-1} \otimes diag(\boldsymbol{\tau})\Big)\mathbf{X} + \boldsymbol{\Gamma}_0^{-1}$ and $\otimes$ denotes the Kronecker product. Here, $\boldsymbol{\gamma}_0 = (m_1\mathbf{1}_L', ..., m_K\mathbf{1}_L')$, a priori means of each γ_{jk} stacked into one vector. The joint prior variance of $\boldsymbol{\gamma}$ is defined in a similar manner: $\boldsymbol{\Gamma}_0 = diag(t_1\mathbf{1}_L', ..., t_K\mathbf{1}_L')$. Each m_k is sampled from the normal distribution with mean $\frac{1}{D}\sum_{j=1}^{D}\gamma_{jk}$ and standard deviation t_k^{-1}/D. Shrinkage parameters t_k are sampled from the gamma distribution with shape $D/2 + a$ and rate $\frac{1}{2}\sum_{j=1}^{D}(\gamma_{jk} - m_k)^2 + b$.

Third, Σ is sampled from the inverse Wishart distribution:

$$\Sigma | \boldsymbol{\gamma}, \mathbf{s}, \boldsymbol{\tau} \sim \mathcal{IW}\Big(N + w_1, w_2\mathbf{I} + \sum_{i=1}^{N} \tau_i \mathbf{u}_i' \mathbf{u}_i\Big) \tag{A.8}$$

where $\mathbf{u}_i = \mathbf{z}_i - \mathbf{x}_i'\boldsymbol{\gamma}$. We let $w_1 = h + D + 1$ and $w_2 = hs$ and then sample parameters (h, s). The conditional density $\pi(h, s|\Sigma)$ does not have a closed form; thus, we employ the Metropolis algorithm with random walk proposals (see **?**).

The step A.4 is implemented by first sampling

$$\tau | \nu, Z, \Sigma, \boldsymbol{\gamma} \sim \mathcal{G}\Big(\nu/2 + D/2, (Z - X\boldsymbol{\gamma})'\Sigma^{-1}(Z - X\boldsymbol{\gamma})/2 + \nu/2\Big) \tag{A.9}$$

and then sampling $\nu|\tau$ using the Metropolis algorithm with independence sampling. Alternatively, if one prefers to use the t model with known degrees of freedom (e.g., for sensitivity analysis), then the step A.4 is reduced to sampling $\tau|\cdot$ while ν is fixed at a chosen value.

Finally, the Metropolis algorithm is also used to sample the count model parameters $\boldsymbol{\beta}$ and w. The convergence of the algorithm is monitored using **?** diagnostic as well as Heidelberg and Welch diagnostic (**?**).

A.2 Average Predicted Party System

A predicted party party system $\hat{\mathbf{v}}^{(t)}$, given a vector of covariates $\mathbf{x}'$, can be simulated by the following algorithm (here, the superscript t indexes an iteration of the MCMC algorithm):

1. Draw a sample $\big(\boldsymbol{\beta}^{(t)}, w^{(t)}, \boldsymbol{\gamma}^{(t)}\big)$;

2. Calculate the median predicted number of parties $\hat{n}$ from the truncated negative binomial distribution with the mean $\mathbf{x}'\boldsymbol{\beta}^{(t)}$ and the dispersion $w^{(t)}$. The median is calculated using Monte Carlo method. We use the rejection sampling (**?**) to generate samples from the truncated negative binomial distribution;

3. Calculate $\hat{\boldsymbol{\mu}}^{(t)} = \mathbf{x}'\boldsymbol{\gamma}^{(t)}$;

4. Apply the transformation in (2) to the first $\hat{n} - 1$ elements of $\hat{\boldsymbol{\mu}}^{(t)}$, which will yield $\hat{\mathbf{v}}^{(t)}$.