

International Systems and Domestic Politics: Linking Complex
Interactions with Empirical Models in International Relations
(Empirical Appendix)

Stephen Chaudoin University of Illinois at Urbana-Champaign
chaudo@illinois.edu

Helen V. Milner Princeton University
hvmilner@princeton.edu

Xun Pang Tsinghua University
xpang@tsinghua.edu.cn

This appendix describes the supplementary empirical analysis for “International Systems and Domestic Politics: Linking Complex Interactions with Empirical Models in International Relations.” It contains the algorithm for estimating the model with time varying ρ , the algorithm for estimating a multilevel model with a time-varying ρ , a description of the nonconvergence diagnostics run on each model, and a description of the stationarity checks ran on the model with time-invariant ρ .

1 Estimation Algorithm for Time-Varying ρ and W

This section describes the MCMC algorithm for estimating a spatial model with time varying ρ and W , as in Section 2.5 of the paper. The model combining multilevel and spatial analyses is estimated using an algorithm combining the following algorithm and the algorithm proposed in [author information removed].

- A Spatial Model with time-varying ρ and W

$$y_{it} = \rho_t \mathbf{w}_{it} \mathbf{y}_t + \mathbf{x}_{it} \boldsymbol{\beta} + \epsilon_{it}$$

Matrix notation

$$\mathbf{y}_t = \rho_t \mathbf{W}_t \mathbf{y}_t + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\epsilon}_t, \quad t = 1, 2, \dots, T$$

where $\mathbf{y}_t = \{y_{1t}, y_{2t}, \dots, y_{nt}\}$, and ρ_t and \mathbf{W}_t is a $N \times N$ weight matrix with each row summed up to 1 and all diagonal elements equal to 0. Assume that $E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t') = \sigma_e^2 \mathbf{I}_n, \forall t$. This is a simultaneous model and all y 's are determined at the same time. Rearrange the model to get

$$\mathbf{A}(\rho_t) \mathbf{y}_t = \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\epsilon}_t, \quad t = 1, 2, \dots, T$$

where $\mathbf{A}(\rho_t) = (\mathbf{I} - \rho_t \mathbf{W}_t)$.

- Likelihood

$$f(\mathbf{Y}|\boldsymbol{\beta}, \mathbf{W}, \boldsymbol{\rho}, \sigma_e^2) \propto \prod_{t=1}^T |\mathbf{A}(\rho_t)| \sigma_e^{N_t} \exp \left[-\frac{1}{2\sigma_e^2} \{ \mathbf{A}(\rho_t) \mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta} \}' \{ \mathbf{A}(\rho_t) \mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta} \} \right]$$

- Priors

$$\boldsymbol{\beta} \sim \mathcal{N}_k(\boldsymbol{\beta}_0, \mathbf{B}_0), \quad \rho_t \sim \mathcal{U}(-1, 1), \quad \sigma_e^{-2} \sim \mathcal{G}(a_0, b_0)$$

- Posterior

(a) $\boldsymbol{\beta}$

$$\boldsymbol{\beta} | \boldsymbol{\rho}, \sigma_e^{-2} \sim \mathcal{N}(\bar{\boldsymbol{\beta}}, \mathbf{B}_1),$$

$$\text{where, } \mathbf{B}_1 = \left(\mathbf{B}_0 + \sigma_e^{-2} \sum_{t=1}^T \mathbf{X}_t' \mathbf{X}_t \right)^{-1}$$

$$\bar{\boldsymbol{\beta}} = \mathbf{B}_1 \left(\mathbf{B}_0^{-1} \boldsymbol{\beta}_0 + \sigma_e^{-2} \sum_{t=1}^T \mathbf{X}_t' \mathbf{A}(\rho_t) \mathbf{y}_t \right)$$

(b) σ_e^{-2}

$$\sigma_e^{-2} | \boldsymbol{\beta}, \boldsymbol{\rho} \sim \mathcal{G}(\alpha_1, \delta_1),$$

$$\text{where, } \alpha_1 = a_0 + \sum_{t=1}^T N_t$$

$$\delta_1 = b_0 + \sum_{t=1}^T (\mathbf{A}(\rho_t) \mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta})' (\mathbf{A}(\rho_t) \mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta})$$

(c) $\{\rho_t\}$ Metropolis-Hastings algorithm. The proposal density is

$$\rho_t^* \sim \mathcal{N}(\phi, \Phi)$$

$$\text{where, } \Phi = \left(\sigma_e^{-2} \sum_{i=1}^{N_t} (\mathbf{w}_{it} \mathbf{y}_t)' (\mathbf{w}_{it} \mathbf{y}_t) \right)^{-1}$$

$$\phi = \Phi \sigma_e^{-2} \sum_{i=1}^{N_t} (\mathbf{w}_{it} \mathbf{y}_t)' (y_{it} - \mathbf{x}_{it} \boldsymbol{\beta})$$

$$\alpha = \frac{|\mathbf{A}(\rho_t^*)| \exp \left[-\frac{1}{2\sigma_e^2} \{ \mathbf{A}(\rho_t^*) \mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta} \}' \{ \mathbf{A}(\rho_t^*) \mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta} \} \right]}{|\mathbf{A}(\rho_t)| \exp \left[-\frac{1}{2\sigma_e^2} \{ \mathbf{A}(\rho_t) \mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta} \}' \{ \mathbf{A}(\rho_t) \mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta} \} \right]}$$

The determinant $|\mathbf{A}(\rho_t)| = \prod_{i=1}^m (1 - \rho_t \lambda_i)$ where λ_i is the i th eigenvalue of the m eigenvalues of \mathbf{W}_t . Then

$$\log(\alpha) = \sum_{i=1}^m \log(1 - \rho_t^* \lambda_i) - \sum_{i=1}^m \log(1 - \rho_t \lambda_i) - \frac{1}{2\sigma_e^2} (\{ \mathbf{A}(\rho_t^*) \mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta} \}' \{ \mathbf{A}(\rho_t^*) \mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta} \} - \{ \mathbf{A}(\rho_t) \mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta} \}' \{ \mathbf{A}(\rho_t) \mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta} \})$$

if $\log(\alpha) > 0$ accept ρ_t^* ; otherwise, reject the proposal and keep ρ_t as the updated value.

2 Multilevel Spatial Modeling

This section describes the MCMC algorithm for estimating a multilevel spatial model with time varying ρ and \mathbf{W} , as in Section 2.6 of the paper.

- Model specification as in section 2.6 of the paper.

$$Y_{it} = \beta_0 + \beta_{0t} + b_i + \rho_t W_{it} \mathbf{Y}_t + \beta_{Dt} D_{it} + \epsilon_{it} \quad (1)$$

$$\beta_{0t} = \beta_S S_t + c_t \quad (2)$$

$$\beta_{Dt} = \gamma_0 + \gamma_S S_t + \zeta_t, \quad (3)$$

The reduced form of the model (plug the last two equations into the first one, and reagrange the

function)

$$Y_{it} = \beta_0 + \rho_t W_{it} \mathbf{Y}_t + \gamma_0 D_{it} + \beta_S S_t + \gamma_S S_t D_{it} + \zeta_t D_{it} + b_i + c_t + \epsilon_{it} \quad (4)$$

The composite error is $\epsilon_{it} = \zeta_t D_{it} + \beta_{0i} + c_t + \epsilon_{it}$, which is apparently correlated with two terms of the observed for sure. To solve this serious endogeneity problem, the Bayesian approach takes all the three terms $\zeta_t D_{it}, \beta_{0i}, c_t$ out of the error term and treats ζ_t, β_{0i} and c_t as model parameters to be estimated based on data, though those parameters may not be interesting.

If the first column of the matrix of \mathbf{D}_t are all 1, then the matrix expression of the reduced model is as following:

$$\mathbf{Y}_t = \rho_t \mathbf{W}_t \mathbf{Y}_t + \mathbf{D}_t \boldsymbol{\gamma}_0 + S_t \boldsymbol{\beta}_s + S_t \mathbf{D}_t \boldsymbol{\gamma}_s + \mathbf{D}_t \mathbf{c}_t + \mathbf{b}_i + \boldsymbol{\epsilon}_t, \quad (5)$$

$$= \rho_t \mathbf{W}_t \mathbf{Y}_t + \mathbf{X}_t \boldsymbol{\beta} + \mathbf{D}_t \mathbf{c}_t + \mathbf{b}_i + \boldsymbol{\epsilon}_t \quad (6)$$

where $\mathbf{Y}_t = \{Y_{1t}, Y_{2t}, \dots, Y_{nt}\}$, $\mathbf{b}_i = \{b_1, b_2, \dots, b_n\}$, $\mathbf{c} = \{\zeta_t, c_t\}$, $\boldsymbol{\beta} = \{\gamma_0, \beta_s, \gamma_s\}$, $\mathbf{X}_t = \{\mathbf{D}_t, S_t, S_t \mathbf{D}_t\}$ and \mathbf{W}_t is a $N \times N$ weight matrix with each row summed up to 1 and all diagonal elements equal to 0. Assume that $E(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t') = \sigma^2 \mathbf{I}_n$, $\forall t$. This is a simultaneous model and all y 's are determined at the same time. Rearrange the model to get

$$\mathbf{A}(\rho_t) \mathbf{Y}_t = \mathbf{X}_t \boldsymbol{\beta} + \mathbf{D}_t \mathbf{c}_t + \mathbf{b}_i + \boldsymbol{\epsilon}_t, \quad (7)$$

where $\mathbf{A}(\rho_t) = (\mathbf{I} - \rho_t \mathbf{W}_t)$.

- Priors: priors are required for a Bayesian model. The parameters are assigned with priors which assume the following distributive forms:

$$\begin{aligned} \boldsymbol{\beta} &\sim \mathcal{N}_{K_1}(\boldsymbol{\beta}_0, \mathbf{B}_0), & \{\mathbf{b}_i\} &\sim \mathcal{N}_{K_2}(\mathbf{0}, \mathbf{D}), & \mathbf{D}^{-1} &\sim \mathcal{W}(\nu_0, \mathbf{D}_0), \\ \{\mathbf{c}_t\} &\sim \mathcal{N}(\mathbf{0}, \mathbf{E}), & \mathbf{E}^{-1} &\sim \mathcal{W}(\boldsymbol{\eta}_0, \mathbf{E}_0), & \rho_t &\sim \mathcal{U}(\rho : \rho \in S_\rho), \\ \epsilon_t &\sim \mathcal{N}(0, \sigma^2), & \sigma^{-2} &\sim \mathcal{G}(a_0, b_0), \end{aligned}$$

where \mathcal{N}_k denotes a k-dimensional multivariate normal distribution, \mathcal{W} denotes a Wishart distribution, \mathcal{U} denotes a uniform distribution, and \mathcal{G} denotes a Gamma distribution.

- The posterior is as follows:

$$\pi(\Theta|\mathbf{Y}) \propto f(\mathbf{Y}|\Theta)\pi(\Theta) \quad (8)$$

$$\begin{aligned} \pi(\boldsymbol{\beta}, \{\rho_t\}, \{\mathbf{b}_i\}, \{\mathbf{c}_t\}, \mathbf{D}, \mathbf{E}, \sigma^2|\mathbf{Y}) &\propto \prod_{t=1}^T |\mathbf{A}(\rho_t)| \sigma^{N_t} \\ &\times \exp \left[-\frac{1}{2\sigma^2} \{ \mathbf{A}(\rho_t) \mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta} - \mathbf{D}_t \mathbf{c}_t - \mathbf{b}_i \}' \{ \mathbf{A}(\rho_t) \mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta} - \mathbf{D}_t \mathbf{c}_t - \mathbf{b}_i \} \right] \\ &\times \pi(\boldsymbol{\beta}) \pi(\{\rho_t\}) \pi(\{\mathbf{b}_i\}) \pi(\{\mathbf{c}_t\}) \pi(\mathbf{D}) \pi(\mathbf{E}) \pi(\sigma^2) \end{aligned} \quad (9)$$

2.1 MCMC Algorithm

Besides the spatial autoregressive coefficient, all model parameters can be directly sampled from their full conditional posteriors using the Gibbs Sampler. For $\{\rho_t\}$, a Metropolis-Hastings algorithm with a normal proposal distribution is applied to update the parameters. The iterative simulation scheme is as follows:

- Sample $\boldsymbol{\beta}$ from the multivariate distribution below conditional on the most updated values of other parameters and the data:

$$\boldsymbol{\beta} \sim \mathcal{N}(\bar{\boldsymbol{\beta}}, \mathbf{B}_1), \quad (10)$$

$$\text{where, } \mathbf{B}_1 = \left(\mathbf{B}_0 + \sigma^{-2} \sum_{t=1}^T \mathbf{X}_t' \mathbf{X}_t \right)^{-1} \quad (11)$$

$$\bar{\boldsymbol{\beta}} = \mathbf{B}_1 \left(\mathbf{B}_0^{-1} \boldsymbol{\beta}_0 + \sigma^{-2} \sum_{t=1}^T \mathbf{X}_t' \mathbf{A}(\rho_t) (\mathbf{Y}_t - \mathbf{D}_t \mathbf{c}_t - \mathbf{b}_i) \right) \quad (12)$$

- Sample σ^{-2} from the gamma distribution below conditional on the most updated values of other parameters and the data

$$\sigma^{-2} \sim \mathcal{G}(\alpha_1, \delta_1), \quad (13)$$

where, $\alpha_1 = a_0 + \sum_{t=1}^T N_t$ (14)

$$\delta_1 = b_0 + \sum_{t=1}^T (\mathbf{A}(\rho_t)\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta} - \mathbf{D}_t\mathbf{c}_t - \mathbf{b}_i)' (\mathbf{A}(\rho_t)\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta} - \mathbf{D}_t\mathbf{c}_t - \mathbf{b}_i) \quad (15)$$

(c) Update \mathbf{b}_i one by one based on the following conditional posterior distribution:

$$\mathbf{b}_i \sim \mathcal{N}(\bar{\mathbf{b}}_i, \mathbf{D}_{1i}), \quad (16)$$

where

$$\mathbf{D}_{1i} = (\mathbf{D}^{-1} + \sigma^{-2}\mathbf{I})^{-1} \quad (17)$$

$$\bar{\mathbf{b}}_i = \mathbf{D}_{1i}\sigma^{-2} \left(\sum_{t=1}^T (Y_{it} - \rho_t W_{it} \mathbf{Y}_t - X_{it}\boldsymbol{\beta} - D_{it}\mathbf{c}_t) \right) \quad (18)$$

(d) Draw \mathbf{c}_t one by one from their respective conditional posterior distribution conditional on the most updated values of other parameters and the data:

$$\mathbf{c}_t \sim \mathcal{N}(\bar{\mathbf{c}}_t, \mathbf{E}_{1i}), \quad (19)$$

where

$$\mathbf{E}_{1i} = (\mathbf{E}^{-1} + \sigma^{-2}\mathbf{D}'_t\mathbf{D}_t)^{-1} \quad (20)$$

$$\bar{\mathbf{c}}_t = \mathbf{E}_{1i}\sigma^{-2}\mathbf{D}'_t(\mathbf{A}(\rho_t)\mathbf{y}_t - \mathbf{X}_t\boldsymbol{\beta} - \mathbf{b}_i) \quad (21)$$

(e) Update the variance-covariance matrices of \mathbf{D} in the following way:

$$\mathbf{D}^{-1} \sim \mathcal{W}(\nu_1, \mathbf{D}_1), \quad (22)$$

where $\nu_1 = \nu_0 + N$, and $\mathbf{D}_1 = (\mathbf{D}_0^{-1} + \sum_{i=1}^N \boldsymbol{\beta}_i\boldsymbol{\beta}'_i)^{-1}$,

(f) Similarly, \mathbf{E} is updated using its conditional posterior:

$$\mathbf{E}^{-1}|\{\mathbf{c}_t\} \sim \mathcal{W}(\boldsymbol{\eta}_1, \mathbf{E}_1), \quad (23)$$

where, $\boldsymbol{\eta}_1 = \boldsymbol{\eta}_0 + T$, and $\mathbf{E}_1 = (\mathbf{E}_0^{-1} + \sum_{i=1}^T \mathbf{c}_i \mathbf{c}_i')^{-1}$

(g) $\{\rho_t\}$ MH algorithm. The proposal density is

$$\rho_t^* \sim \mathcal{N}(\phi, \Phi) \quad (24)$$

$$\text{where, } \Phi = \left(\sigma^{-2} \sum_{i=1}^N (W_{it} \mathbf{Y}_t)' (W_{it} \mathbf{Y}_t) \right)^{-1} \quad (25)$$

$$\phi = \Phi \sigma^{-2} \sum_{i=1}^{N_t} (W_{it} \mathbf{Y}_t)' (Y_{it} - \mathbf{X}_t \boldsymbol{\beta} - \mathbf{D}_t \mathbf{c}_t - \mathbf{b}_i) \quad (26)$$

$$\alpha = \frac{|\mathbf{A}(\rho_t^*)| \exp \left[-\frac{1}{2\sigma^2} \{ \mathbf{A}(\rho_t^*) \mathbf{Y}_t - \mathbf{X}_t \boldsymbol{\beta} - \mathbf{D}_t \mathbf{c}_t - \mathbf{b}_i \}' \{ \mathbf{A}(\rho_t^*) \mathbf{Y}_t - \mathbf{X}_t \boldsymbol{\beta} - \mathbf{D}_t \mathbf{c}_t - \mathbf{b}_i \} \right]}{|\mathbf{A}(\rho_t)| \exp \left[-\frac{1}{2\sigma^2} \{ \mathbf{A}(\rho_t) \mathbf{Y}_t - \mathbf{X}_t \boldsymbol{\beta} - \mathbf{D}_t \mathbf{c}_t - \mathbf{b}_i \}' \{ \mathbf{A}(\rho_t) \mathbf{Y}_t - \mathbf{X}_t \boldsymbol{\beta} - \mathbf{D}_t \mathbf{c}_t - \mathbf{b}_i \} \right]} \quad (27)$$

The determinant $|\mathbf{A}(\rho_t)| = \prod_{i=1}^m (1 - \rho_t \lambda_i)$ where λ_i is an eigenvalue of \mathbf{W}_t . Then if $\alpha > 1$ accept ρ_t^* ; otherwise, reject the proposal and keep ρ_t as the updated value

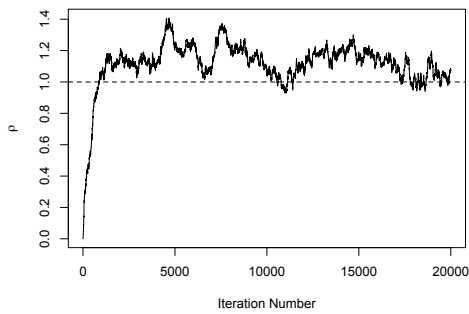
(h) Repeat the process from (a) to (g) till convergence.

3 Stationarity Check

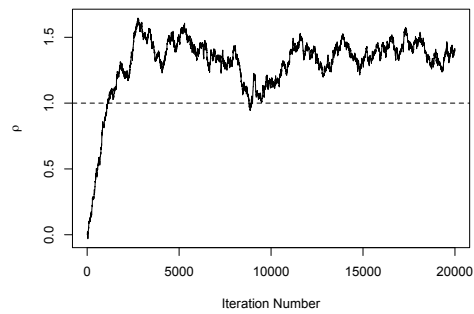
To assess stationarity concerns with the time-invariant ρ model, we re-estimated time-invariant ρ model without placing any restrictions on the parameter space of ρ . We started the chain for ρ at zero. Figure 1 shows the trace plots of draws of ρ in two such MCMC simulations without a burn-in stage.

The plot on the left-side is based on a spatial model using mutual GATT/WTO membership to construct the matrix of spatial weights. After approximately one hundred iterations, the chain went out of the stationarity space, and stayed in the area above 1 for most of the time. There is no sign that the chain would return in the stationarity space by increasing the number of iterations.

Figure 1: System-Specific/Time-Specific Effect of Regime



Spatial Weight: Mutual GATT/WTO Membership



Spatial Weight: Geographical Distance

To further check this problem, we re-estimated the same model, but used a time-invariant matrix of spatial weights, where the weight for two countries consists of their geographical distance. The trace plot on the right-side hand shows that it is very unlikely that the process is stationary, either. While the chain for ρ moves above 1 slightly less quickly (though still pretty quickly), the chain is very rarely below 1.