

Bayesian Estimation of Endogenous Network Effects in SAR Models: Application to R&D Collaboration

Yajie Xu

University of South Florida

yajie@usf.edu

Murat Munkin

University of South Florida

mmunkin@usf.edu

October 19, 2026

[\[Click here for the most recent version\]](#)

Abstract

Selection biases pose a big challenge in estimating spillover effects in spatial autoregressive (SAR) models, primarily because network formation often depends on unobservable individual preferences affecting outcomes. This paper proposes a novel Bayesian framework to jointly model the dynamics of network effects and their formation and extends it to panel data. Specifically, we introduce an explicit latent structure, which preserves conjugacy, connects outcomes and network selection equations in the SAR model through the error terms and allows computationally more efficient estimation than previous studies. This novel structure simultaneously models endogenous formation and captures the associated spillover effects. Estimation is based on a Markov Chain Monte Carlo (MCMC) algorithm, all steps of which except one are the Gibbs sampler, a benefit of the modified model specification. We first validate the model through a simulation study and then apply it to data on R&D collaboration networks to estimate technological spillovers and assess the role of immigrant participation in network formation. The paper intends to contribute a novel general framework for estimating endogenous spillover effects and to promote broader, computationally simpler and more efficient application of Bayesian methods in network modeling.

Keywords: Bayesian, Networks, Spillover Effects, Endogeneity, R&D, Collaboration

JEL Codes: C11, L14, L24, J15

1 Introduction

This paper develops a SAR model with an endogenous spatial weight matrix to estimate peer effects of R&D collaboration among pharmaceutical firms on their annual performance for a sample of 173 firms observed from 2003 to 2022 and obtained from the SDC Joint Ventures & Strategic Alliances Database. The proposed SAR approach builds on the cross-sectional model of (Hsieh and Lee, 2016) and extends it to the case of panel data. It allows for potential endogeneity through an explicit latent structure designed in a way that leads to a computationally more efficient estimation algorithm than those in the existing studies. Specifically, the latent utility that generates the endogenous weight matrix is assumed to have a probit structure instead of the logit choice adopted by the existing literature. That preserves conjugacy and results in more efficient Gibbs sampler steps of the MCMC algorithm instead of computationally demanding Metropolis-Hasting (MH) steps utilized in the logit functional forms. Secondly, we directly place the error terms from the outcome equations into the network selection equations to control for the effects of the unobservable factors. Since the underlying distributions are normal the posterior distribution can be augmented with the corresponding latent variables which simplifies the MCMC algorithm. Finally, instead of generating endogeneity by introducing several unobserved variables into the network equations representing various unobservable human traits as in (Hsieh and Lee, 2016) we do so by placing a single error term from the outcome equation in the latent variables of the network equations. The main concern is that having multiple unobservable variables can potentially lead to identification issues similar to those present in finite mixture models.

The modeling approach in this paper builds on the Spatial Autoregressive (SAR) model, introduced into the field of social interaction to address the *reflection problem*. This problem was defined by Manski (1993) as the inability to distinguish the endogenous peer effects from the contextual effects. The causal effect of peers' endogenous social interactions on the agents' outcomes is commonly referred to as the *spillover effect*. Detailed data on network topology can theoretically help (see Bramoullé et al. (2009)), however, the observed social interactions introduce a more fundamental challenge: network endogeneity. Because agents strategically form ties based on unobserved preferences and endowments, the network structure is likely to be correlated with those unobservables that also influence the agents' outcomes. Empirical applications of the SAR approach include studies on how friendship influences academic achievements and health behavior (Hsieh and Lee, 2016; Boucher et al., 2014; Hsieh and Lin, 2017; Chung, 2020; Han et al., 2021), microfinance participation Lambotte (2025), from the micro-level, such as firms' performance and innovation activities (König et al., 2019; Hsieh et al., 2025); and to the macro-level, including systemic risk in

interbank market (Acemoglu et al., 2015; Elliott et al., 2014), as well as studies on European sovereign debt crisis D’Innocenzo et al. (2024), shock propagation through input-output networks Acemoglu et al. (2012) and international trade Kireyev and Leonidov (2021).

Our model is a panel data variant of (Hsieh and Lee, 2016) in which individual fixed effects can represent either actual individual firms or industry sectors to which the firms belong. However, unlike (Hsieh and Lee, 2016), the selection equations are governed by a Probit structure instead of Logit. The corresponding MCMC algorithm produces a posterior distribution augmented with latent data similar to the approach of Albert and Chib (1993) for posterior sampling. The approach of (Hsieh and Lee, 2016) uses the logit specification together with Metropolis-Hastings steps of the MCMC algorithm. We view such choices as leading to unnecessary computational inefficiencies that could be avoided even within the logit framework. Recent developments in Bayesian inference of the logit model by Frühwirth-Schnatter et al. (2008), Frühwirth-Schnatter and Frühwirth (2009) and Polson et al. (2013) use new data augmentation samplers more efficient than the MH steps utilized in (Hsieh and Lee, 2016). Our choice of the probit specification aims to simplify the estimation algorithm and make it more appealing to practitioners.

Instead of specifying the outcome equation conditional on a set of latent variables representing unobservable individual characteristics that generate endogeneity in the network formation matrix as in (Hsieh and Lee, 2016) we explicitly put the error term (ε) from the SAR outcome equation into the latent utility function of the Probit model to control for the effect of those unobservable factors that affect both the outcome and the weight matrix of friendship formation. In addition to modeling endogeneity this strategy achieves two goals. It leads to a relative computational simplicity of the resulting full conditional distributions. It also removes the need to introduce several unobservable latent variables to capture individual characteristics something that potentially leads to identification issues in estimating the effects of such latent factors. The latent variables in Hsieh and Lee (2016) effectively specify a finite mixture model of variables constructed as absolute values of normally distributed random variables. However, identification of finite mixtures can be challenging. See Gormley and Frühwirth-Schnatter (2019) for the discussion of three types of non-identifiability: invariance to relabeling the components, potential overfitting and generic non-identifiability. Munkin (2022) observes identification difficulties of finite mixture models when the outcome variable is an observed overdispersed count. The problem is likely to persist or even get worse in the case of latent utility generating the binary logit outcomes.

We develop a Bayesian Markov Chain Monte Carlo (MCMC) algorithm for estimation and test it first with a set of numerical examples. The simulation experiment shows that the estimation method recovers the unknown parameters with a good finite-sample perfor-

mance. Then the estimation method is used in the application of inter-firm research and development (R&D) collaboration networks. The empirical results reveal significant positive spillover effects on firm output through the co-research relationship. Crucially, the estimated endogeneity parameter δ is statistically very strong, rejecting the null hypothesis of exogenous network formation.

Comparing our results with the standard SAR model underscores the importance of modeling endogeneity. While the magnitudes of the spillover effects are similar, the standard SAR model with exogenous network formation produces inflated standard errors (inefficiency), especially in the presence of contextual effects. It can potentially lead to wrong conclusions that R&D cooperation yields minimal returns. These findings highlight the importance of the adopted modeling approach to ensure valid causal inference.

The remainder of the paper is organized as follows. Section 2 defines the model. The Bayesian MCMC algorithm is derived in section 3. Section 4 presents a simulation study to evaluate model performance. Section 5 deals with an application on R&D collaboration networks. Finally section 6 concludes the paper.

2 SAR Model with endogenous network formation

Assume that we have n independent pharmaceutical companies making individual choices of their friendship networks represented by a $n \times n$ spatial weight matrix W_t , which evolves over time t and may not be symmetric due to lack of reciprocity. In a given period t , each entry w_{ijt} represents the relationship from individual i to individual j , and is equal to 1 if i claims its relationship with j and 0 otherwise, so the link runs from sender i to receiver j . The diagonal entries are all restricted to zeros, which avoids a self-loop and is mathematically written as $w_{ijt} = 0$ when $i = j$.

Let $Y_t = (y_{1t}, y_{2t}, \dots, y_{nt})'$ be an $n \times 1$ vector of outcomes (company's revenue in our application) at time t , $X_t = (x_{1t}, x_{2t}, \dots, x_{nt})'$ be an $n \times k$ matrix of exogenous characteristics at time t for these n independent pharmaceutical companies. In the matrix form, the SAR model can be written as:

$$Y_t = \lambda W_t Y_t + \beta_0 \mathbf{1}_n + X_t \beta_1 + W_t X_t \beta_2 + \boldsymbol{\alpha} + \tau_t \mathbf{1}_n + \varepsilon_t, \quad t = 1, \dots, T \quad (1)$$

where λ represents the endogenous peer effect, the key parameter of interest. β_0 is the global intercept, and $\mathbf{1}_n$ is an $n \times 1$ vector of ones. β_1 and β_2 are coefficients for the direct (own) and contextual (friends) effects of the exogenous characteristics. The term $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)'$ is an $n \times 1$ vector capturing time-invariant individual fixed effect. Also $\tau = (\tau_1, \tau_2, \dots, \tau_t)'$

is a $T \times 1$ vector of time fixed effects, where each τ_t is a time-specific scalar that equals 1 at time t . To ensure identification of the sector and time fixed effects alongside the global intercept, we impose the normalizations $\alpha_1 = 0$ and $\tau_1 = 0$, treating individual (sector) 1 and period 1 as the baseline (Han et al. 2021; Greene 2018)¹ Finally, $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{nt})'$ is a vector of stochastic error terms, where $\varepsilon_{it} \stackrel{i.i.d.}{\sim} N(0, \sigma_\varepsilon^2)$. Equation (1) in the vector form for each i ($i = 1, \dots, n$) and t ($t = 1, \dots, T$) can be written as

$$y_{it} = \lambda \sum_{j=1}^n w_{ijt} y_{jt} + \beta_0 + x_{it} \beta_1 + \sum_{j=1}^n w_{ijt} x_{jt} \beta_2 + \alpha_i + \tau_t + \varepsilon_{it} \quad (2)$$

Note that while we specify individual fixed effects here, it can be easily converted to sector fixed effects in case there is a need to control for S sectors instead of n individual firms. In such a case, we replace α_i in Equation (2) with $\alpha_{s(i)}$, where $s(i) \in \{1, \dots, S\}$ denotes the sector to which firm i belongs. Then α_s is the corresponding sector effect with the normalization $\alpha_1 = 0$ retained. This change requires a very small change to the MCMC algorithm described in Section 3.

2.1 Network Formation

We model potential endogeneity of each entry of network matrix W_t to the outcome variable Y_t through a latent structure representing unobserved heterogeneity using the Probit framework. Assume that the binary choice variable w_{ijt} is determined based on the difference in levels of latent utility

$$u_{ijt}(w_{ijt} = 1) - u_{ijt}(w_{ijt} = 0) = \Delta u_{ijt} \quad (3)$$

derived by individual i from his friendship with individual j and specified as

$$\Delta u_{ijt} = \psi_{ijt} + \zeta_{ijt} \quad (4)$$

where

$$\psi_{ijt} = c_{it}\gamma_1 + c_{jt}\gamma_2 + c_{ijt}\gamma_3$$

is the deterministic part and ζ_{ijt} is the error term. Vectors c_{it} and c_{jt} are individual-specific variables explaining the friendship link decision, which may overlap with x_{it} and x_{jt} . How-

¹These restrictions resolve the perfect multicollinearity that arises from including a full set of sector dummies, a full set of time dummies, and a global intercept. Specifically, without normalizations, the sum of the sector indicators and the sum of the time indicators would both be perfectly collinear with the constant vector $\mathbf{1}_n$, violating the full-rank condition. By setting $\alpha_1 = 0$ and $\tau_1 = 0$, the intercept β_0 captures the expected baseline outcome for sector 1 in period 1. The remaining sector effects α_s ($s = 2, \dots, S$) measure sector-specific deviations relative to sector 1, and the remaining time effects τ_t ($t = 2, \dots, T$) measure period-specific deviations relative to period 1.

ever, vector c_{ijt} are observed dyad-specific variables, a natural choice for the exclusion restrictions. The time and individual fixed effects are not included because they cancel out when computing the difference Δu_{ijt} . Random variables ε_{it} and ζ_{ijt} are assumed to be jointly distributed as bivariate normal such that the correlation between them is generated by unobserved factors simultaneously affecting the network choice and economic outcome variables. However, instead of introducing specific latent variables controlling for such unobservables as in (Hsieh and Lee, 2016), we write a regression-like equation

$$\zeta_{ijt} \equiv \delta \varepsilon_{it} + \eta_{ijt} \quad \eta_{ijt} \stackrel{i.i.d.}{\sim} N(0, 1) \text{ and independent of } \varepsilon_{it}, \quad (5)$$

and ε_{it} is derived from equation (2) as

$$\varepsilon_{it} = y_{it} - \lambda \sum_{j=1}^n w_{ijt} y_{jt} - \left(\beta_0 + x_{it} \beta_1 + \sum_{j=1}^n w_{ijt} x_{jt} \beta_2 + \alpha_i + \tau_t \right).$$

This equation can also be interpreted as defined by the conditional distribution $\zeta_{ijt} \mid \varepsilon_{it} \sim N(\delta \varepsilon_{it}, 1)$ based on the joint distribution of ε_{it} and ζ_{ijt} . The variance of the conditional distribution of Δu_{ijt} given ε_{it} is fixed for identification of the Probit part. Then

$$\Delta u_{ijt} = c_{it} \gamma_1 + c_{jt} \gamma_2 + c_{ijt} \gamma_3 + \delta \varepsilon_{it} + \eta_{ijt}, \quad \eta_{ijt} \stackrel{i.i.d.}{\sim} N(0, 1) \quad (6)$$

and the observability condition is specified as

$$w_{ijt} = \begin{cases} 1 & \text{if } \Delta u_{ijt} > 0 \\ 0 & \text{if } \Delta u_{ijt} < 0 \end{cases} \quad (7)$$

In essence, endogeneity of the spatial choice matrix W_t is modeled through a joint distribution of the errors in the outcomes equation (1) and network choice equation (5) where

$$(\varepsilon_{it}, \zeta_{ijt}) \sim N_n \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon\eta} \\ \sigma_{\varepsilon\eta} & \sigma_\eta^2 \end{pmatrix} \right)$$

such that $\zeta_{ijt} \mid \varepsilon_{it} \sim N(\sigma_{\varepsilon\eta} \sigma_\varepsilon^{-2} \varepsilon_{it}, \sigma_\eta^2 - \sigma_{\varepsilon\eta}^2 \sigma_\varepsilon^{-2})$, where we further denote $\delta = \sigma_{\varepsilon\eta} \sigma_\varepsilon^{-2}$ and set $\sigma_\eta^2 = 1 + \sigma_{\varepsilon\eta}^2 \sigma_\varepsilon^{-2}$.

The sign of δ has an economic interpretation. The specification is a selection model in the sense of Heckman (1979): the network is observed as a choice affected by the unobservable ε_{it} , so that conditioning on the observed links is informative about ε_{it} .

3 MCMC algorithm

In this section, we outline the steps of the MCMC algorithm for the SAR model with endogenous network formation with derivation details presented in the Appendix. First, we write the joint likelihood of the observed and latent data. [LeSage and Pace \(2009\)](#) write the SAR model in the reduced form (p. 46). Similarly, Equation (1) can be written as

$$Y_t = (I_n - \lambda W_t)^{-1}(\beta_0 \mathbf{1}_n + X_t \beta_1 + W_t X_t \beta_2 + \alpha + \tau_t \mathbf{1}_n) + (I_n - \lambda W_t)^{-1} \varepsilon_t, \quad (8)$$

where I_n denotes the $n \times n$ identity matrix, and the inverse matrix $(I_n - \lambda W_t)^{-1}$ represents the spatial multiplier that captures the global spillover effects throughout the network. This equation facilitates calculation of the Jacobian of transformation for the substitution of ε_{it} with y_{it} in the corresponding likelihood derived below. Denote variables $Z_t = (\mathbf{1}_n, X_t, W_t X_t)$, $Z_{it} = \left(1, x_{it}, \sum_{j=1}^n w_{ijt} x_{jt}\right)$, $C_{ijt} = (c_{it}, c_{jt}, c_{ijt})$, $Q_{ijt} = (C_{ijt}, \varepsilon_{it})$, and parameters $\Gamma = (\gamma_1, \gamma_2, \gamma_3)'$, $\theta' = (\Gamma', \delta)$, $\beta = (\beta_0, \beta_1, \beta_2)'$, $\Theta = \{\lambda, \beta, \alpha, \tau, \sigma_\varepsilon, \theta\}$. Based on Equation (8) and (6), the joint likelihood can be written as

$$\begin{aligned} & P(Y, \Delta u \mid W, \Theta) \\ \propto & \left\{ (2\pi\sigma_\varepsilon^2)^{-\frac{Tn}{2}} |I - \lambda W_t|^T \right. \\ & \times \exp \left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^T \sum_{i=1}^n [y_{it} - \lambda \sum_{j=1}^n w_{ijt} y_{jt} - \alpha_i - \tau_t - Z_{it} \beta]^2 \right) \\ & \left. \times \exp \left[-\frac{1}{2} \sum_{t=1}^T \sum_{i=1}^n \sum_{j \neq i}^n \left(\Delta u_{ijt} - C_{ijt} \Gamma - \delta [y_{it} - \lambda \sum_{j=1}^n w_{ijt} y_{jt} - \alpha_i - \tau_t - Z_{it} \beta] \right)^2 \right] \right\}. \end{aligned} \quad (9)$$

The full conditional kernel is the product of the likelihood and the prior distributions of the model parameters, specified in the steps of the MCMC algorithm. We augment the posterior distribution with the latent variables Δu_{ijt} . The MCMC algorithm is fully derived in the Appendix. Here we just outline the steps of the algorithm as follows:

1. Conditionally on ε_{it} and individual- and dyad-level regressors, the latent utility variables Δu_{ijt} are drawn from the truncated normal distributions:

$$\Delta u_{ijt} \mid Q_{ijt}, \theta, w_{ijt} \stackrel{\text{ind}}{\sim} \begin{cases} \text{TN}_{(-\infty, 0]}(Q_{ijt} \theta, 1) & \text{if } w_{ijt} = 0, \\ \text{TN}_{(0, \infty)}(Q_{ijt} \theta, 1) & \text{if } w_{ijt} = 1. \end{cases}$$

2. Given the prior distribution of $\gamma \sim N(\underline{\Gamma}, \underline{H}_\Gamma^{-1})$ and $\delta \sim N(\underline{\delta}, \underline{H}_\delta^{-1})$ we form the prior

$\theta \sim \mathcal{N}(\underline{\theta}, \underline{H}_\theta^{-1})$ such that

$$\begin{aligned}\underline{\theta} &= \begin{pmatrix} \underline{\Gamma} \\ \underline{\delta} \end{pmatrix}, \\ \underline{H}_\theta &= \begin{pmatrix} \underline{H}_\Gamma & 0 \\ 0 & \underline{H}_\delta \end{pmatrix}.\end{aligned}$$

Then the full conditional distribution of θ is $\theta \sim N(\bar{\theta}, \bar{H}_\theta^{-1})$ where

$$\begin{aligned}\bar{H}_\theta &= H_\theta + \sum_{t=1}^T \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n Q_{ijt}^\top Q_{ijt}, \\ \bar{\theta} &= \bar{H}_\theta^{-1} \left(H_\theta \underline{\theta} + \sum_{t=1}^T \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n Q_{ijt}^\top \Delta u_{ijt} \right)\end{aligned}$$

3. Given the prior distribution of $\beta \sim N(\underline{\beta}, \underline{H}_\beta^{-1})$, the full conditional distribution of β is:

$$\beta \sim N(\bar{\beta}, \bar{H}_\beta^{-1}),$$

where

$$\begin{aligned}\bar{H}_\beta &= \underline{H}_\beta + \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n-1) \right) \sum_{t=1}^T \sum_{i=1}^n Z'_{it} Z_{it}, \\ \bar{\beta} &= \bar{H}_\beta^{-1} \left[\underline{H}_\beta \underline{\beta} + \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n-1) \right) \sum_{t=1}^T \sum_{i=1}^n Z'_{it} \left(y_{it} - \lambda \sum_{j=1}^n w_{ijt} y_{jt} - \alpha_i - \tau_t \right) \right. \\ &\quad \left. - \delta \sum_{t=1}^T \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n Z'_{it} (\Delta u_{ijt} - C_{ijt} \Gamma) \right]\end{aligned}$$

4. For each individual $i = 1, \dots, n$, we specify a prior $\alpha_i \sim N(\underline{\alpha}_i, \underline{H}_{\alpha_i}^{-1})$, with α_1 fixed at zero. Then the full conditional distribution of α_i is

$$\alpha_i \sim N(\bar{\alpha}_i, \bar{H}_{\alpha_i}^{-1}),$$

where

$$\begin{aligned}\overline{H}_{\alpha_i} &= \underline{H}_{\alpha_i} + T \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n-1) \right), \\ \overline{\alpha}_i &= \overline{H}_{\alpha_i}^{-1} \left[\underline{H}_{\alpha_i} \underline{\alpha}_i + \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n-1) \right) \sum_{t=1}^T \left(y_{it} - \lambda \sum_{j=1}^n w_{ijt} y_{jt} - Z_{it} \boldsymbol{\beta} - \tau_t \right) \right. \\ &\quad \left. - \delta \sum_{t=1}^T \sum_{\substack{j=1 \\ j \neq i}}^n (\Delta u_{ijt} - C_{ijt} \Gamma) \right]\end{aligned}$$

5. Given the prior distribution of $\tau_t \sim N(\underline{\tau}_t, \underline{H}_{\tau_t}^{-1})$, the full conditional distribution of each τ_t is:

$$\tau_t \sim N(\overline{\tau}_t, \overline{H}_{\tau_t}^{-1}),$$

where

$$\begin{aligned}\overline{H}_{\tau_t} &= \underline{H}_{\tau_t} + n \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n-1) \right), \\ \overline{\tau}_t &= \overline{H}_{\tau_t}^{-1} \left[\underline{H}_{\tau_t} \underline{\tau}_t + \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n-1) \right) \sum_{i=1}^n \left(y_{it} - \lambda \sum_{j=1}^n w_{ijt} y_{jt} - Z_{it} \boldsymbol{\beta} - \alpha_i \right) \right. \\ &\quad \left. - \delta \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n (\Delta u_{ijt} - C_{ijt} \Gamma) \right]\end{aligned}$$

6. The full conditional of σ_ε^{-2} , given the prior

$$\sigma_\varepsilon^{-2} \sim \text{Gamma} \left(\frac{\nu}{2}, \frac{g}{2} \right),$$

is

$$\sigma_\varepsilon^{-2} \sim \text{Gamma} \left(\frac{\nu + nT}{2}, \frac{g}{2} + \frac{1}{2} \sum_{t=1}^T \sum_{i=1}^n \varepsilon_{it}^2 \right).$$

7. Given the uniform prior $\lambda \sim \mathcal{U}(-1, 1)$, the full conditional distribution of λ is:

$$P(\lambda \mid Y, \Delta u, W, \Theta) \propto \pi(\lambda) \cdot P(Y, \Delta u \mid W, \Theta),$$

where the likelihood is given by Equation (9). To sample from the full conditional of λ , we design a Metropolis-Hastings (MH) step with a proposal density defined as in Lesage (1997); LeSage and Pace (2009); Hsieh and Lee (2016); Han et al. (2021).

Step 1: Propose $\tilde{\lambda} \sim N(\lambda^{(q-1)}, v_\lambda)$, where $\lambda^{(q-1)}$ is the current state of the chain and the proposal variance v_λ is adaptively tuned to maintain an acceptance rate between 30% and 40%. Specifically, if the acceptance rate falls below 30%, the variance is decreased by setting $v'_\lambda = v_\lambda/1.1$; if it exceeds 40%, the variance is increased by setting $v'_\lambda = v_\lambda \times 1.1$. If $|\tilde{\lambda}| \geq 1$, which violates both the uniform prior support and the stability condition², the proposal is rejected and set $\lambda^{(q)} = \lambda^{(q-1)}$.

Step 2: Otherwise, compute the acceptance probability

$$\Pr(\lambda^{(q-1)}, \tilde{\lambda}) = \min \left\{ 1, \frac{\pi(\tilde{\lambda})}{\pi(\lambda^{(q-1)})} \frac{p(Y, \Delta u \mid \tilde{\lambda}, \beta, \alpha, \tau, \sigma_\varepsilon^{-2}, \Gamma, \delta)}{p(Y, \Delta u \mid \lambda^{(q-1)}, \beta, \alpha, \tau, \sigma_\varepsilon^{-2}, \Gamma, \delta)} \right\}.$$

If a uniform random draw $u \sim \mathcal{U}(0, 1)$ satisfies $u < \Pr$, set $\lambda^{(q)} = \tilde{\lambda}$; otherwise, $\lambda^{(q)} = \lambda^{(q-1)}$.

4 Numerical Examples

This section tests the model, the MCMC algorithm and the programs with numerical examples in which the data are generated according to the data-generating process (d.g.p.) designed to mirror the empirical application introduced in the next section. More specifically, we generate a balanced, directed panel of $N = 200$ firms distributed across $S = 5$ sectors, observed over $T = 20$ time periods. As mentioned in the definition of the model we impose a standard normalization restriction on the sector and time effects, restricting to zero the first group and time effects ($\alpha_1 = 0$ and $\tau_1 = 0$). The remaining unobserved heterogeneity components are drawn from the standard normal distribution, such that $\alpha_s \sim \mathcal{N}(0, 1)$ for $s \in \{2, \dots, 5\}$ and $\tau_t \sim \mathcal{N}(0, 1)$ for $t \in \{2, \dots, 20\}$.

The directed network adjacency matrix W_t is constructed by first generating latent utilities and evaluating their differences Δu_{ijt} . We generate two-dimensional firm-level covariates, $c_{it} \sim \mathcal{N}(0, I_2)$, representing attributes of sender (i). The covariates of receiver

²The stability condition ensures that the spatial autoregressive process is stationary and that the matrix $I - \lambda W$ is non-singular. For a row-normalised spatial weight matrix W (with eigenvalues inside the unit circle), the condition reduces to $|\lambda| < 1$. In our MCMC sampler this is automatically enforced by the uniform prior $\lambda \sim \mathcal{U}(-1, 1)$, and any proposal with $|\tilde{\lambda}| \geq 1$ is rejected.

(j) enter Δu_{ijt} in the network formation equation. We also draw three-dimensional dyad-specific covariates, $c_{ijt} \sim \mathcal{N}(0, I_3)$, to capture homophily or synergistic effects for all $N(N - 1)$ pairs in each period. The structural error term ε_{it} is drawn from $\mathcal{N}(0, 1)$ and it also enters the network formation equations to introduce endogeneity. Finally, the idiosyncratic edge-specific shock is drawn as $\eta_{ijt} \sim \mathcal{N}(0, 1)$. The network formation parameters are chosen as follows: the baseline intercept is set to -4.5 to govern the baseline sparsity of the network; the coefficients for the sender, receiver, and dyadic covariates (θ) are randomly drawn from a discrete uniform distribution $\mathcal{U}_d\{0.1, 0.2, \dots, 0.9\}$; and the coefficient capturing the endogenous selection effect of the structural error is set to $\delta = -0.5$. Network indicators equal $w_{ijt} = 1$ if the latent utility gain is strictly positive ($\Delta u_{ijt} > 0$) and 0 otherwise. Because the structural error ε_{it} is specific to sender i , $\Delta u_{ijt} \neq \Delta u_{jit}$, naturally generating an asymmetric directed network. Under these parameter values, the simulated network density is approximately 0.59%, closely aligning with the 0.51% density observed in our empirical application.

Next, we generate the outcome vector Y_t . The exogenous covariate matrix X_t includes an intercept and four explanatory variables drawn independently from the standard normal distribution, $x_{it} \sim \mathcal{N}(0, I_4)$. The true parameters for the SAR outcome equation are specified as follows: the global intercept is $\beta_0 = 0.5$, the coefficients for the exogenous characteristics are $\beta_1 = (1, 1, 1, 1)'$, and the endogenous peer effect, the key parameter of interest, is set to $\lambda = 0.3$. The outcome equation uses the row-normalized version of the simulated network, $\bar{W}_t$, the entries of which are $\bar{w}_{ijt} = w_{ijt} / \sum_{k=1}^n w_{ikt}$, with $\bar{w}_{ijt} = 0$ for a firm that has no links.³ Combining all exogenous covariates, the unobserved sector (α_s) and time (τ_t) fixed effects, $\bar{W}_t$, and the stochastic error term ε_{it} , the outcome vector Y_t is constructed according to Equation (2).

Table 1 reports the posterior means, standard deviations, and 95% credible intervals together with the true parameter values used in the data-generating process (DGP). To reduce high autocorrelation present for some parameters we collect every 10th draw after the burn-in period, yielding a final posterior sample of 1,800 draws.

Overall, the spatial and structural parameters are estimated with high accuracy. The posterior mean of the endogenous peer effect is $\hat{\lambda} = 0.289$ (95% CI is $[0.274, 0.304]$) given the true value of 0.3. Similarly, the estimated endogeneity parameter is $\hat{\delta} = -0.514$ (95% CI $[-0.531, -0.498]$) and the true value is -0.5 . This shows that the

³We row-normalize the adjacency matrix so that the spatial lag represents the average outcome of a firm's neighbors. Firms without links have zero rows. This normalization also ensures that $I_n - \lambda W_t$ is invertible for $|\lambda| < 1$.

Table 1: Monte Carlo experiment results

Parameter	True	Mean	SD	Ratio	ACF20	CI95
<i>Spatial Autoregressive (SAR) Outcome Model</i>						
λ	0.300	0.289	0.008	37.305	0.096	0.274, 0.304
β_0	0.500	0.559	0.058	9.679	0.715	0.430, 0.661
β_1	1.000	1.014	0.011	90.091	0.026	0.992, 1.035
β_2	1.000	1.004	0.011	91.255	-0.021	0.982, 1.026
β_3	1.000	1.015	0.011	89.290	0.025	0.994, 1.038
β_4	1.000	0.990	0.012	84.268	0.003	0.968, 1.014
σ^2	1.000	1.026	0.023	43.875	-0.005	0.981, 1.075
<i>Network Formation Model</i>						
δ	-0.500	-0.514	0.008	-61.712	0.103	-0.531, -0.498
γ_0	-4.500	-4.542	0.030	-151.860	0.789	-4.598, -4.476
γ_1	0.100	0.102	0.008	12.099	0.018	0.087, 0.119
γ_2	0.700	0.700	0.010	71.030	0.413	0.681, 0.721
γ_3	0.100	0.091	0.008	11.537	-0.011	0.075, 0.107
γ_4	0.500	0.506	0.009	58.735	0.174	0.489, 0.524
γ_5	0.400	0.415	0.008	51.213	0.113	0.399, 0.430
γ_6	0.500	0.502	0.009	55.039	0.264	0.484, 0.520
γ_7	0.900	0.901	0.011	81.916	0.542	0.879, 0.922

Notes: Results are based on 20,000 MCMC draws with a burn-in of 2,000 and a thinning factor of 10, resulting in 1,800 retained draws.

estimation method performs well even in the extreme cases of low network density. The coefficients for the exogenous characteristics are also centered at their true values, exhibiting high precision with mean-to-standard-deviation ratios approaching 90.

Figure 1 displays trace plots and autocorrelation functions (ACF) for the thinned posterior sample. The MCMC algorithm achieves an acceptance rate of 40.07% for the MH step of the spatial parameter λ , within the optimal target rate range for the random walk Metropolis-Hastings sampler. The trace plots indicate that the Markov chains are stable and center accurately on the true parameter values. Following the thinning procedure, the lag-20 autocorrelation drops to near-zero levels for the primary spatial and structural parameters. Moderate autocorrelation persists for the baseline intercept (γ_0) and the global outcome intercept (β_0), an expected outcome for an MCMC algorithm with extensive fixed effects. Overall, the simulation results provide strong evidence that the proposed method yields reliable inferences for moderately sized directed network panels. We estimated various specifications with a range of parameter and network density values. The MCMC algorithm produces reasonable results but we report only one set for the d.g.p. consistent with the application.

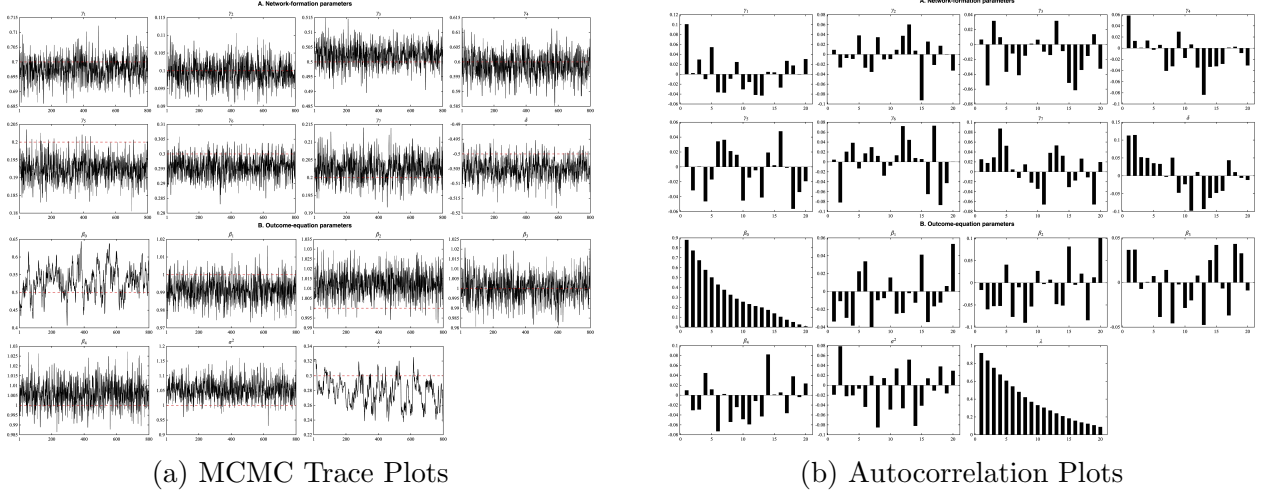


Figure 1: MCMC diagnostics for the simulated data-generating exercise. (a) Trace plots, (b) Autocorrelation functions

Notes: Diagnostics are computed based on 20,000 MCMC draws with a burn-in of 2,000 draws, retaining every tenth draw. Panels are grouped by equations: the network-formation parameters γ and the endogeneity parameter δ , followed by the outcome-equation parameters β , the disturbance variance σ_ε^2 , and the endogenous peer effect λ . Panel titles use the same parameter symbols reported in Table 1.

5 Empirical Application

The structural model is developed in this paper to study the effects of R&D strategic alliances on firm performance. More generally the existing literature demonstrates that R&D cooperation fundamentally drives firm performance, industrial competitiveness, economic growth, and social welfare improvement (Schilling and Phelps, 2007; Cassiman and Veugelers, 2002; Ahuja, 2000; Powell et al., 1996). In this application, we pursue two primary objectives. First, we evaluate the impact of cooperation on firm performance, measured by revenue. Second, we want to understand the underlying mechanism that drives formation of such networks. Specifically, we are interested in whether presence of immigrant executives acts as a catalyst for alliance formation. More specifically, we estimate the effects of homophily and synergistic characteristics in shaping the architecture of R&D strategic alliances.

5.1 Data Summary

The R&D alliance data from 2003 to 2022 are obtained from the SDC *Joint Ventures & Strategic Alliances* Database (König et al. 2019; Schilling 2009). This database is chosen for its comprehensive coverage of inter-firm R&D collaborations, drawing from a wide range of sources, including SEC filings, international equivalents, trade

publications, and news reports. To focus on innovation-oriented partnerships, only alliances explicitly classified as R&D collaborations are included in the sample.

Over the past two decades, several firms in the dataset underwent mergers and acquisitions (M&A). Following the approach of [König et al. \(2019\)](#), it is assumed that acquiring firms inherit all existing R&D collaborations of the target firms. Information on M&A events is obtained from Thomson Reuters’ SDC M&A and S&P Compustat databases, while firm-level financial data are collected from S&P Compustat and Financial Modeling Prep.

The analysis focuses exclusively on publicly listed pharmaceutical firms with Standard Industrial Classification (SIC) code 283, which represents the industry with the highest number of R&D collaborations ([Hsieh et al. 2025](#)).⁴ Following [Hsieh et al. \(2025\)](#) and [König et al. \(2019\)](#), monetary variables are converted to millions of U.S. dollars and deflated by the CPI of the firm’s home country. All continuous variables are transformed using a logarithmic transformation to reduce skewness and facilitate comparison across variables. The outcome variable is the natural logarithm of annual revenue. Total assets, R&D expenditure and cost of revenue enter in natural logarithms, whereas leverage (long-term debt divided by assets) and return on assets enter in levels. Debt/assets and return on assets are winsorized at the first and ninety-ninth percentiles. To address the concern of reverse causality, all financial controls are lagged one year. 2003 observations are used to construct the first lags, so the estimation sample covers 2004–2022. Table 2 presents descriptive statistics of all variables used in the outcome equation.

Imm is a time-varying dummy variable for a firm that equals one when at least one senior executive was identified as an immigrant in that year. Information about them is collected from multiple sources. The primary source is firms’ annual reports filed with the Securities and Exchange Commission (SEC), which contain officers’ names and brief biographies. This information is further cross-referenced and supplemented using official company websites, Wikipedia, LinkedIn, WikiTree, Bloomberg, Encyclopedia.com, [NNDB](#), [WBE](#), and various media reports.

Because birthplace information is occasionally unavailable, an inferential approach is adopted: the country of the university where an executive obtained their bachelor’s degree is used as a proxy for birthplace ([Mahroum and Ansari 2017](#)). Executives who obtained their undergraduate degrees outside the country where their firm is head-

⁴According to [Bloom et al. \(2013\)](#), R&D activities are predominantly concentrated among publicly traded firms, suggesting that such samples capture the majority of innovation efforts.

Table 2: Summary statistics for firm-level variables (SAR sample)

Variable	Mean	SD	Median	Min	Max
Output (log revenue)	4.195	3.660	4.090	-7.724	11.222
US	0.500	0.500	1.000	0.000	1.000
Imm	0.273	0.446	0.000	0.000	1.000
Imm (Contextual)	0.137	0.305	0.000	0.000	1.000
Assets	5.538	3.018	5.269	-8.067	12.285
Leverage	0.130	0.202	0.032	0.000	1.141
ReturnOnAssets	-0.350	0.971	-0.056	-7.155	0.483
R&D	3.248	2.681	3.117	-8.023	16.913
CostOfRevenue	3.011	3.516	2.330	-8.510	10.233
Observations	3,080				

Five financial regressors are lagged one year relative to the output year. Monetary variables are CPI-deflated and expressed in millions of U.S. dollars. *Output*, *Assets*, *R&D* and *CostOfRevenue* are in natural logarithms; *Leverage* is the ratio of long-term debt to assets and *ReturnOnAssets* is net income over assets, both in levels. *US* equals one for U.S.-headquartered firms. *Imm* is the time-varying U.S.-firm indicator defined in the text. It equals zero for non-U.S. firms. *Imm (Contextual)* is the row-normalized share of a firm’s current collaborators that are immigrant-led; it is zero otherwise. The sample is 3,080 firm-years from 172 SIC-283 firms over 2004–2022.

quartered are classified as immigrants. Since some individuals may have immigrated at a young age and completed their education in the United States, this approach may undercount immigrant executives; therefore, the resulting estimates represent a lower bound of immigrants’ contribution to U.S. executive leadership.

To isolate the influence of immigrant executives in the U.S. context, all executives of non-U.S. firms are assumed to have been born in the same country where their firm is located. This assumption is supported by several observations. First, immigrant executives are considerably more prevalent in the United States than elsewhere, particularly in technology and healthcare sectors (Mahroum and Ansari 2017). In contrast, firms in many Asian and African countries tend to hire domestic executives due to cultural, linguistic, or policy-related barriers (Arp et al. 2013; Platonova and Urso 2013; Flahaux and Haas 2016). In Europe, while cross-country mobility exists, cultural proximity often limits executive diversity. Moreover, research indicates that compared to the United States, European countries have historically maintained less attractive immigration policies for highly skilled workers, leading to a persistent “brain drain” toward the U.S. (Mahroum, 1999, 2000; Prato, 2025).

The dyadic variables, reported in Table 3, capture relational, demographic and financial

Table 3: Summary statistics for dyadic/network variables

Variable	Mean	SD	Median	Min	Max
R&DAlliance	0.005	0.071	0.000	0.000	1.000
IM	0.472	0.499	0.000	0.000	1.000
BothUS	0.249	0.432	0.000	0.000	1.000
SameCountryOfBirth	0.290	0.454	0.000	0.000	1.000
SameCity	0.007	0.084	0.000	0.000	1.000
SameIndustry	0.496	0.500	0.000	0.000	1.000
OldFriends	0.015	0.136	0.000	0.000	4.000
CommonFriends	0.087	0.282	0.000	0.000	1.000
DifAssets	3.347	2.579	2.791	0.000	19.852
DifLeverage	0.180	0.221	0.113	0.000	1.141
DifReturnOnAssets	0.689	1.188	0.348	0.000	7.638
DifR&D	2.967	2.302	2.531	0.000	24.785
DifCost	3.913	3.034	3.495	0.000	18.565
Observations	248,226				

R&DAlliance is a binary variable representing pairwise relationships: it equals one when a pair initiated an alliance in the current or preceding four years. Binary regressors (*IM*, *BothUS*, *SameCountryOfBirth*, *SameCity*, *SameIndustry*, and *CommonFriends*) equal one when the stated relational, geographic or industry condition holds for a pair and zero otherwise. In particular, *IM* equals one when at least one firm in a pair is immigrant-led. *OldFriends* is the cumulative number of a pair's prior alliance initiations; its observed range is from zero to four. Variables prefixed *Dif* are absolute differences between two firms' lagged attributes. *DifAssets* is computed on natural logarithms of deflated total assets; *DifLeverage* on ratios of long-term debt to assets; *DifReturnOnAssets* on returns on assets; and *DifR&D* and *DifCost* on log-transformed R&D expenditure and cost of revenue. The summary statistics is computed over 248,226 dyad-years of the estimation sample (2004–2022) of which 1,272 observations are linked.

characteristics of each firm pair. *IM* equals one if at least one senior executive in either firm is an immigrant. *BothUS* equals one if both firms are headquartered in the United States, and *SameCountryOfBirth* equals one if at least one pair of executives from two firms was born in the same country. *OldFriends* counts a pair’s alliance initiations before the current year, and *CommonFriends* equals one if they shared at least one collaborator before the current alliance is considered. *SameCity* equals one if two firms are located in the same city, and *SameIndustry* equals one if they share the same four-digit SIC code. Together these indicators characterize social, geographic and technological proximity that may influence formation of interfirm R&D collaborations.

The annual network variable *R&DAlliance* in Table 3 follows the five-year duration assumption of König et al. (2019): an alliance is treated as active in the year it is initiated and in the four subsequent years. A pair is therefore linked in year t if it initiated an alliance in t or in any of the four preceding years, and a pair that initiates repeatedly may remain linked for longer than five consecutive years. The resulting undirected adjacency matrix is row-normalized separately by year, so λ is interpreted as the spillover from the average outcome of a firm’s current collaborators rather than from their sum (Hsieh and Lee, 2016). Because alliances enter and expire, each W_t is a yearly snapshot of the extant collaboration structure. Firms without a collaborator in a given year remain in the sample and contribute a zero row to W_t .

Homophily, the tendency to link with similar partners, is among the most important factors in deciding whom to form an alliance with (Mele, 2022; Hsieh and Lin, 2017). Therefore, we utilize five lagged financial attributes in the network-formation equation constructed as absolute differences measuring the gap between partners in size, leverage, profitability, R&D expenditure and cost of revenue. These are variables prefixed *Dif*: *DifAssets*, *DifLeverage*, *DifReturnOnAssets*, *DifR&D* and *DifCost*. The data determine the direction along each dimension: a negative coefficient indicates homophily, since a wider gap lowers the propensity to link, whereas a positive coefficient indicates complementarity, with firms favoring partners whose attributes differ from their own.

5.2 Estimation Results

The model parameters are estimated using the developed Markov chain Monte Carlo (MCMC) algorithm based on 100,000 iterations discarding the initial 5,000 iterations of the burn-in phase. We retain every tenth draw to reduce serial autocorrelation. The posterior means and standard deviations are calculated based on this output.

Convergence is verified using the standard diagnostics tools presented in Tables 4, 5 and plots 2 (Geweke, 1992; Raftery and Lewis, 1992; Heidelberger and Welch, 1983).

For comparison, the tables also report estimates obtained for the model that ignores endogeneity estimating two equations separately: OLS for the outcome equation and maximum-likelihood Probit for the network equation. The columns titled "Bayesian" report estimates based on the the joint model developed in Section 3, which allows for endogenous network formation.

The coefficient of main interest is the endogenous spillover parameter λ , representing the strength of peer or network interactions in their effect on the output. Because revenue is log-transformed and W_t is row-normalized, λ is the structural elasticity with respect to the average output of a firm's current collaborators, conditional on the included regressors and fixed effects (LeSage and Pace, 2009). Its estimated posterior mean is approximately 0.046, with zero being outside the 95% credible intervals (0.027, 0.064), indicating statistically strong spillovers. This estimate suggests that a 1% increase in the average output of a firm's collaborators is associated with roughly a 0.046% increase in the firm's own output, holding other factors constant. The standard OLS model with exogenous networks underestimates the spillover effect (0.034), consistent with evidence from König et al. (2019).

For a firm connected through a row-normalized component, the implied social multiplier is $(1 - \lambda)^{-1} \approx 1.048$ at the posterior mean: a shock to such a firm's output ultimately generates an equilibrium response about 4.8% larger than the initial impact, occurring through feedback among connected firms. The estimates are interpreted as propagation of revenue shocks through observed R&D alliances, and highlight the economic significance of interfirm interdependency: local improvements in productivity travel beyond the directly affected firms and reinforce aggregate output through repeated spillovers.

ImmigrantExecutive (*Imm*) has a posterior mean of 0.113 (0.055) with posterior standard deviation of 0.055, and a 95% credible interval [0.007, 0.224]. At the same time the posterior mean for *US* is -0.372 (0.076). Together these two binary variables partition the entire sample into three groups. The reference group is non-U.S. firms ($US = 0$) and then $Imm = 0$. The second group is U.S. firms without an immigrant executive ($US = 1$ and $Imm = 0$). The *US* coefficient implies that their expected log revenue is lower by 0.372 log points, corresponding to a level about $e^{-0.372} - 1 \approx 31.1\%$ lower than that of the reference group. The third group is U.S. firms with an identified immigrant executive. Relative to the second group, the expected log revenue is

higher by 0.113 log points, corresponding to a level roughly $e^{0.113} - 1 \approx 12.0\%$ higher. Measured relative to the non-U.S. reference group, the combined conditional difference is $-0.372 + 0.113 = -0.259$ log points, or about $e^{-0.259} - 1 \approx 22.8\%$ lower overall. The positive immigrant-executive coefficient therefore offsets part, but not all, of the negative conditional difference associated with being a U.S. firm, when holding the other regressors, fixed effects, network structure, and network exposure constant.

The contextual regressor $W_t Imm$ has a posterior mean of 0.265, with a 95% credible interval of [0.053, 0.468]. Holding the network matrix fixed, a 10-percentage-point increase in the immigrant-executive share among a firm’s collaborators leads to an increase of about $e^{0.1 \times 0.265} - 1 \approx 2.7\%$ in expected revenue. Beyond this direct impact, a change in Imm also propagates through the feedback loop in the network. The corresponding impact matrix is $S(W_t) = (I - \lambda W_t)^{-1}(\beta_{Imm}I + \beta_{WImm}W_t)$, the average row sum of which gives the total effect (LeSage and Pace, 2009). As an illustrative back-of-the-envelope calculation, if in the row-normalized matrix W_t every firm in a connected network hypothetically switches from zero to one, the average total response is approximately $(0.113 + 0.265)/(1 - 0.046) = 0.396$ log points, corresponding to revenue about 48.6% higher.

Among the lagged own-firm characteristics, *Assets* and *CostOfRevenue* produce the strongest impacts in the outcome equation, with posterior means of 0.523 and 0.530 and 95% credible intervals of [0.427, 0.618] and [0.450, 0.606] respectively. Because both the regressors and outcome are log-transformed, these are conditional elasticities: a one-percent increase in a firm’s lagged assets or cost of revenue results in roughly a half-percent increase in current revenue, indicating that both firm size and operating scale are substantial drivers of output. The remaining three financial controls produce weaker effects. *Leverage* has a posterior mean of 0.127 with interval $[-0.094, 0.348]$, and *R&D* a posterior mean of 0.026 with interval $[-0.040, 0.090]$; both intervals include zero. *ReturnOnAssets* has a negative impact of -0.078 with interval $[-0.164, 0.006]$ and a posterior probability of 0.967 of being negative.

The estimated coefficients in the probit equations are similar in both magnitude and statistical significance across both estimation methods, Bayesian and MLE. However, the endogeneity term δ can be estimated only in the model that allows endogeneity of network selection. The posterior mean of δ is negative and statistically very strong in both specifications, with and without contextual effects. This provides evidence of non-negligible endogeneity between the latent alliance-formation process and firms’ output decisions captured in the SAR model. The negative sign of δ indicates that unobserved

factors such as realized or expected productivity shocks that increase output, at the same time lower the probability of alliance formation. This ultimately results in a negative correlation between the residuals.

The large and negative constant term (posterior mean ≈ -3.26) is reflective of a very low proportion of observed network connections, consistent with the practical difficulties of establishing successful R&D strategic alliances among firms: only 1,272 of 248,226 dyad-years are linked. The positive coefficient of *Imm*, with posterior mean 0.259 and a 95% credible interval of [0.196, 0.322], highlights the role of immigrant executives in facilitating cross-firm cooperation, potentially through broader international networks and cultural openness. By far the strongest predictor is *OldFriends*, at 1.862 with a tight interval [1.809, 1.914], underscoring the importance of relational trust and accumulated experience in sustaining interfirm collaboration. *CommonFriends* is also positive, at 0.131 with interval [0.060, 0.199], indicating that indirect acquaintances and third-party introductions do help bring pharmaceutical firms together, over and above their own prior alliance history. By contrast, the demographic and proximity indicators are all weak and statistically indistinguishable from zero: *SameCountryOf-Birth* (0.003), *SameIndustry* (0.000), *SameCity* (-0.048) and *BothUS* (-0.092), so shared nationality, industry overlap, geographic proximity and common U.S. headquarters play little role once executives' demographic information and firms' relational history are controlled for, consistent with the dominance of global principles among collaborating firms.

The five dyadic financial regressors reveal whether similarity or dissimilarity along those dimensions attracts partners. *DifReturnOnAssets* is negative at -0.193 with interval $[-0.259, -0.133]$, the strongest in magnitude of the financial terms: a wider gap in profitability lowers the propensity to link, which is evidence of homophily, or assortative matching on financial performance. *DifR&D* at 0.031 [0.016, 0.046] and *DifCost* at 0.030 [0.015, 0.043] are smaller in magnitude but their positive sign carries the opposite implication: firms favour partners whose research intensity and cost structure differ from their own, a pattern of strategic complementarity in which each party contributes strengths the other lacks. *DifAssets* (0.001) and *DifLeverage* (-0.001) are statistically indistinguishable from zero, so differences in scale and in indebtedness do not by themselves shape partner choice. Taken together, the financial terms suggest that alliance formation reflects a balance between similarity and complementarity: firms sort towards partners of comparable profitability while seeking diversity in research and cost profiles.

Table 4: Outcome-equation estimates: OLS versus joint Bayesian

Variable	Parameter	OLS		Bayesian			
		Estimate	SE	Mean	SD	ACF20	CI95
Spillover effect	λ	0.034	0.008	0.046	0.009	0.042	0.027, 0.064
Constant	β_0	-0.325	0.334	-0.723	2.010	0.496	-4.616, 3.165
Imm	β_1	0.139	0.071	0.113	0.055	-0.012	0.007, 0.224
Imm (Contextual)	β_2	0.201	0.102	0.265	0.105	0.074	0.053, 0.468
US	β_3	-0.402	0.070	-0.372	0.076	0.090	-0.522, -0.227
Assets	β_4	0.519	0.024	0.523	0.049	0.013	0.427, 0.618
Leverage	β_5	0.144	0.124	0.127	0.112	-0.008	-0.094, 0.348
ReturnOnAssets	β_6	-0.080	0.033	-0.078	0.043	0.017	-0.164, 0.006
R&D	β_7	0.020	0.016	0.026	0.033	0.009	-0.040, 0.090
CostOfRevenue	β_8	0.530	0.014	0.530	0.039	0.019	0.450, 0.606
σ^2	σ_ε^2	—	—	1.795	0.060	0.066	1.691, 1.927
Observations							3,080

Notes: The dependent variable is log revenue. The outcome equation includes four-digit SIC and year fixed effects. OLS columns report coefficient estimates and conventional regression standard errors. Bayesian columns use the 9,500 draws retained after a 5,000-draw burn-in and thinning by ten; they report posterior means, posterior SDs, lag-20 autocorrelations of the thinned chain (ACF20), and 95% equal-tailed credible intervals. *Imm (Contextual)* is $W_t \text{Imm}$, the row-normalized share of current collaborators that are immigrant-led. Dashes indicate that a parameter is not estimated by the corresponding model.

Table 5: Network-formation equation estimates: Probit MLE versus joint Bayesian

Variable	Parameter	MLE		Bayesian			
		Estimate	SE	Mean	SD	ACF20	CI95
δ (Endogeneity)	δ	—	—	-0.058	0.016	0.215	-0.087, -0.024
Constant	γ_0	-3.250	0.040	-3.260	0.040	0.126	-3.341, -3.183
IM	γ_1	0.259	0.031	0.259	0.032	0.021	0.196, 0.322
BothUS	γ_2	-0.091	0.064	-0.092	0.061	0.035	-0.209, 0.028
SameCountryOfBirth	γ_3	-0.000	0.060	0.003	0.057	0.026	-0.111, 0.113
SameCity	γ_4	-0.035	0.176	-0.048	0.165	-0.003	-0.379, 0.263
SameIndustry	γ_5	0.001	0.028	0.000	0.028	-0.029	-0.055, 0.055
OldFriends	γ_6	1.858	0.030	1.862	0.027	0.015	1.809, 1.914
CommonFriends	γ_7	0.132	0.036	0.131	0.035	0.031	0.060, 0.199
DifAssets	γ_8	0.005	0.011	0.001	0.011	0.059	-0.021, 0.022
DifLeverage	γ_9	0.006	0.076	-0.001	0.075	0.010	-0.151, 0.145
DifReturnOnAssets	γ_{10}	-0.189	0.034	-0.193	0.033	0.529	-0.259, -0.133
DifR&D	γ_{11}	0.029	0.008	0.031	0.008	0.047	0.016, 0.046
DifCost	γ_{12}	0.027	0.007	0.030	0.007	0.060	0.015, 0.043
Observations							248,226

Notes: MLE columns report Probit maximum-likelihood estimates and conventional model-based standard errors. *IM* equals one for a pair involving at least one immigrant-led firm. Bayesian columns use the 9,500 draws retained after a 5,000-draw burn-in and thinning by ten; they report posterior means, posterior SDs, lag-20 autocorrelations of the thinned chain (ACF20), and 95% equal-tailed credible intervals. Probit coefficients are latent-index coefficients, not marginal probability effects. The endogeneity loading δ is estimated only by the joint model.

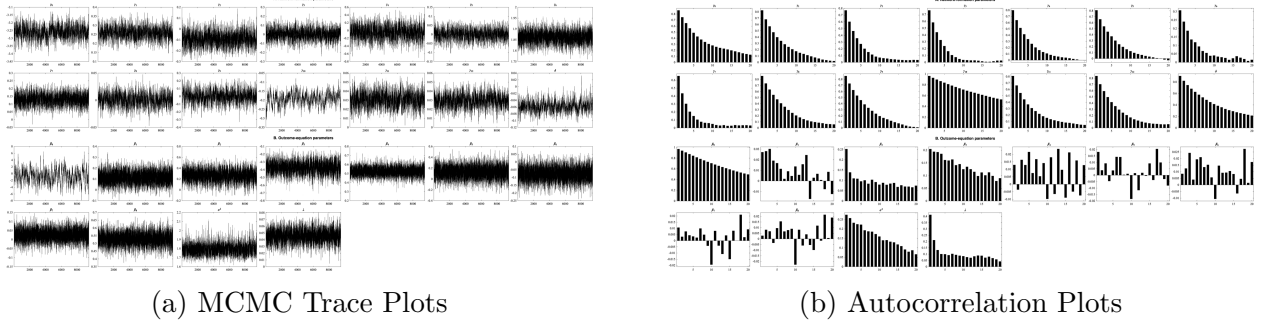


Figure 2: MCMC diagnostics for the estimated model. (a) Trace plots, (b) Autocorrelation functions

Notes: Diagnostics are computed based on 100,000 MCMC draws with 5,000 discarded draws of the burn-in phase, retaining every tenth draw, which leaves 9,500 draws for inference. Both panels use the same thinned sample. Panels are grouped by equation and labelled with the parameter symbols reported in Tables 4 and 5: panel A shows the network-formation coefficients γ_0 – γ_{12} and the endogeneity parameter δ , panel B the outcome-equation coefficients β_0 – β_8 , the disturbance variance σ_ε^2 , and the endogenous peer effect λ .

5.3 Marginal Effects of an Immigrant-Executive Policy

The results presented in Section 5.2 imply that presence of immigrant executives can affect firm performance through a firm’s own status (Imm), its collaborators’ immigrant-executive composition (W_tImm), and alliance formation (IM). Thus, the combined effect coming from all these channels can not be captured by a single parameter, but instead we adopt the approach of Han et al. (2021) to evaluate a counterfactual change in the share of immigrant-led U.S. firms, allowing this intervention to affect both the outcome and network-formation equations and propagate through the spatial multiplier.

The counterfactual computation is conducted for 2022, the last year in our dataset. The year has a total of 156 firms, 79 are U.S. firms, of which 41 are immigrant-led. The immigrant-led share is consequently $41/79 = 51.9\%$. We examine two distinct scenarios, when the share of immigrant-led firms reduces to 20% and when it increases to 80%. The contraction policy reduces the share to 16 out of 79 U.S. firms (20.3%) by changing the status of 25 baseline immigrant-led firms from one to zero. The expansion policy raises the share to 63 out of 79 firms (79.7%) by switching the status of 22 baseline non-immigrant-led firms from zero to one. The immigrant-leadership status of non-U.S. firms remains unchanged at zero under both policies. Firms to be switched are selected once with a fixed assignment seed and remain the same in every network simulation.

We closely follow the discrete-policy exercise of Han et al. (2021) given in Appendix F and evaluate policy effects at the posterior-mean of parameter estimates denoted $\hat{\Theta}$ and

reported in Tables 4 and 5. Since the counterfactual analysis is based on only 2022 data, the time subscript is suppressed. Let $a \in \{b, p\}$ index the immigrant executive assignment, where b is the observed baseline and p is the assignment under the target share, $p \in \{0.2, 0.8\}$. Let Imm^a denote the $n \times 1$ vector of immigrant-executive indicators under assignment a , and let IM_{ij}^a denote the corresponding dyadic regressor, which equals one when at least one of firms i and j is immigrant-led.

The assignments affect revenue through two routes. First, it changes Imm and, through it, the collaborators' exposure $WImm$ in the outcome equation. Second, it changes IM , which enters the latent utility in Equation (6) and may affect the links determined by Equation (7). We exercise two possibilities. In the fixed-network scenario, the network is held at the observed 2022 values, W^{obs} , under both assignments, so only the outcome equation is affected. In the rewired scenario, a new network is generated according to the mechanism defined by Equations (6) and (7) under each assignment, so that the effect can come from both routes. In other words, in the fixed-network case only the outcome variables are generated conditional on the new assignment of immigrant executives. In the rewired case, both outcomes and latent utilities are generated leading to rewiring the network matrix. We denote w_{ij}^a the generated link indicator and W^a the corresponding row-normalized matrix, such that $W^a Imm^a$ is recomputed for each generated network. All other covariates, sector and year effects remain at their 2022 values.

Define firm i 's network degree as a measure of collaboration activity under assignment a as $D_i^a = \sum_{j \neq i} w_{ij}^a$. Further, define the firm-level expected network degree effect as

$$ME_i^D(p) = \mathbb{E}[D_i^p | \hat{\Theta}] - \mathbb{E}[D_i^b | \hat{\Theta}]. \quad (10)$$

The network degree effect is defined only for the rewired scenario, because each firm's degree is unchanged when holding the network fixed.

Let Y_i^a be firm i 's log revenue under assignment a in the rewired scenario, according to Equation (8) conditional on Imm^a and W^a . The firm-level expected revenue effect under endogenous rewiring is

$$ME_i^{\text{rewired}}(p) = \mathbb{E}[Y_i^p | \hat{\Theta}] - \mathbb{E}[Y_i^b | \hat{\Theta}]. \quad (11)$$

The marginal effect in Equations (10) and (11) are approximated by:

$$\widehat{ME}_i^D(p) = \frac{1}{R} \sum_{r=1}^R \left(D_i^{p,r} - D_i^{b,r} \right), \quad \widehat{ME}_i^{\text{rewired}}(p) = \frac{1}{R} \sum_{r=1}^R \left(Y_i^{p,r} - Y_i^{b,r} \right), \quad (12)$$

where replication r changes such that $r = (1, \dots, R)$ and $R = 1,000$. Across replications, $\widehat{\Theta}$, the covariates, and the set of switched firms are held fixed, and only the shocks are redrawn. Within each replication, the baseline and policy scenarios use the same draws of ε_i and η_{ij} , so that the difference between them is only due to the change in assignment.

In the fixed-network scenario, the marginal effect can be calculated directly without simulation as

$$ME^{\text{fixed}}(p) = \left(I_n - \widehat{\lambda} W^{\text{obs}} \right)^{-1} \left(\widehat{\beta}_{Imm} I_n + \widehat{\beta}_{WImm} W^{\text{obs}} \right) (Imm^p - Imm^b). \quad (13)$$

Figures 3 and 4 plot the firm-level expected effects for all 156 firms observed in 2022. Each marker is one firm's effect averaged over 1,000 simulations, so the spread of the markers describes heterogeneity across firms. Figure 3 reports the effect on collaboration activity, measured by network degree, and Figure 4 the effect on log revenue. In both figures, red circles correspond to the contraction policy lowering the share of immigrant-led U.S. firms to 20%, and blue triangles correspond to the expansion policy, which raises it to 80%. Firms are divided into four mutually exclusive groups: U.S. firms the immigrant-executive status of which changes (switchers), immigrant-led U.S. firms (no change in status), non-immigrant-led U.S. firms (no change in status), and 77 non-U.S. firms (no change in status).

Figure 3 shows that the changes in alliance formation are more prominent among switchers. Under contraction, switchers lose 0.457 links on average. The losses are strongly left-skewed: they range from 0.152 to 1.651 links, although most lose fewer than 0.4. Firms the status of which is unchanged are affected far less. The 16 immigrant-led U.S. stayers lose nothing, because for them $IM = 1$ keeps its value, independent of its partner's status. The 38 non-immigrant-led U.S. stayers lose only 0.052 links on average, and the 77 non-U.S. firms only 0.093. Under expansion the pattern is reversed, but the effects are smaller in magnitude. Switchers gain 0.192 links on average, with gains between 0.066 and 0.678. Immigrant-led U.S. stayers gain nothing, non-immigrant-led U.S. stayers gain 0.014 links on average, and non-U.S. firms only 0.046.

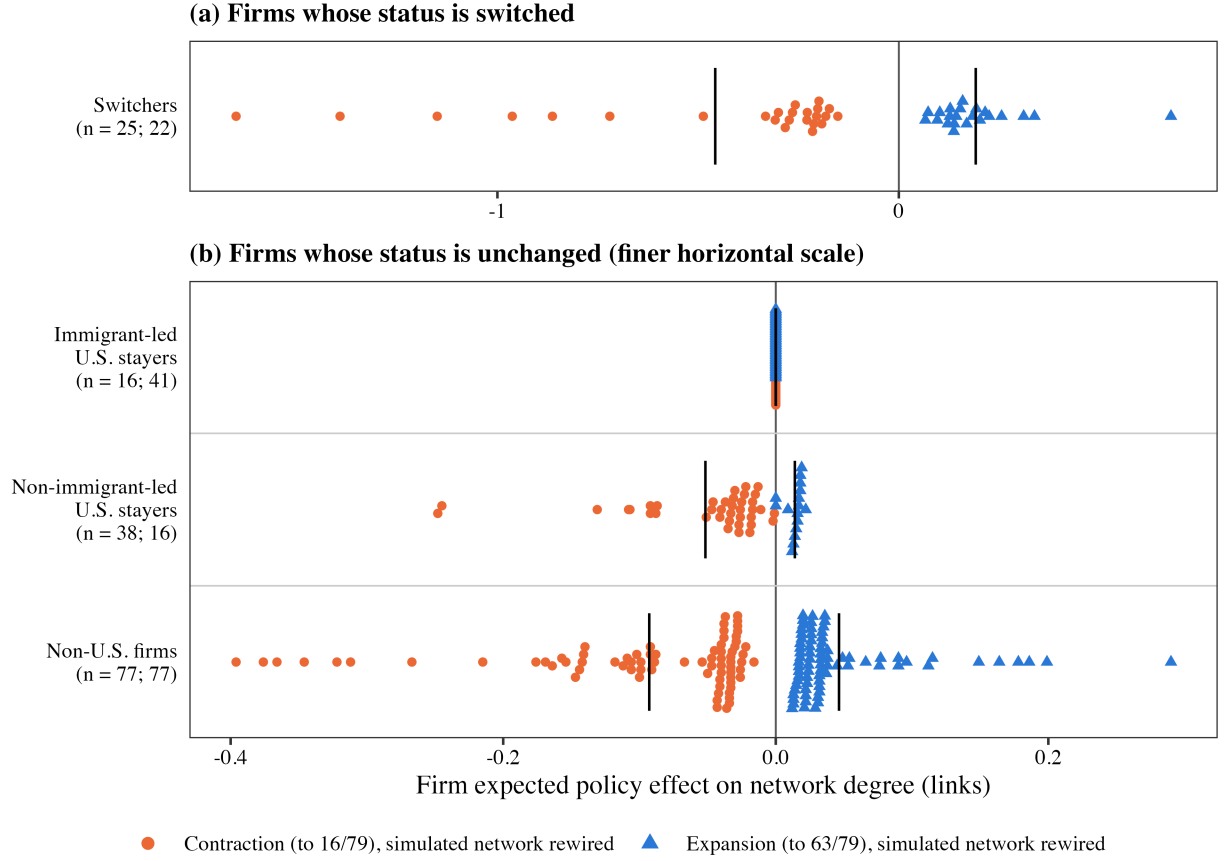


Figure 3: Firm-level expected policy effects on network degree under contraction and expansion

Notes: Each marker is one firm's expected change in network degree (number of collaborators), policy minus baseline, averaged over 1,000 simulations at fixed posterior-mean parameters. Red circles correspond to contraction from the observed 41/79 immigrant-led U.S. firms to 16/79 and blue triangles to expansion to 63/79. Policy assignments are fixed across simulations. The sample is the 156 firms observed in 2022. Groups are policy-specific. The numbers in parentheses below each group label give the number of firms in that group under contraction and under expansion, respectively. Vertical bars mark group means. Firms with nearly identical effects are stacked vertically, so the height of a stack indicates the number of firms. Panel (b) uses a finer horizontal scale than panel (a). Degree effects with the observed network held fixed are zero by construction and are not plotted.

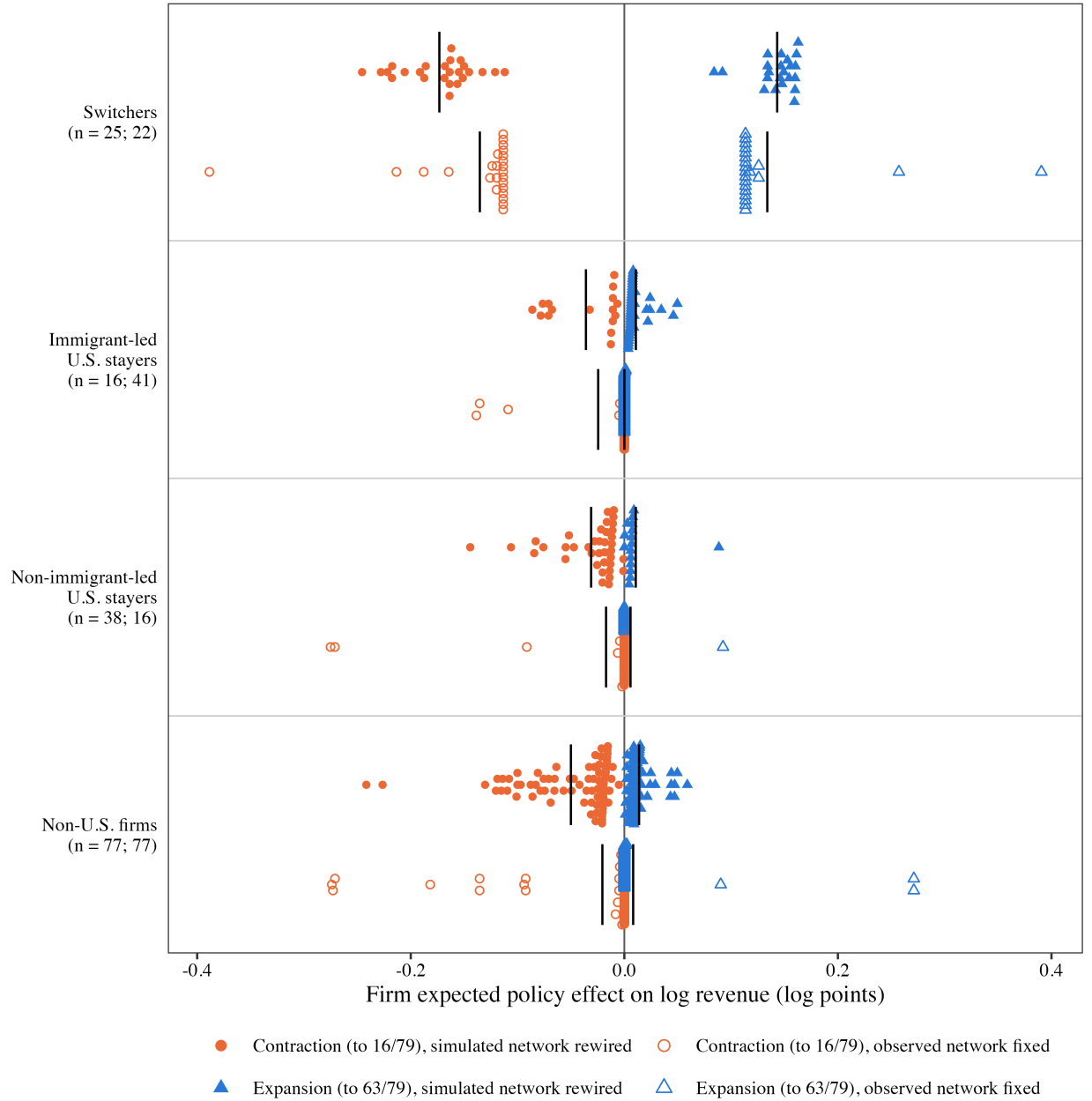


Figure 4: Firm-level expected policy effects on log revenue under contraction and expansion
Notes: Each marker is one firm's expected change in log revenue. Solid markers correspond to the rewired case averaged over 1,000 paired simulations at fixed posterior-mean parameters. Hollow markers correspond to the fixed case with the observed 2022 network fixed under both assignments. Circles correspond to contraction from the observed 41/79 immigrant-led U.S. firms to 16/79 and triangles correspond to expansion to 63/79. Policy assignments are fixed across simulations. The sample is the 156 firms observed in 2022. Groups are policy-specific. The numbers in parentheses below each group label give the number of firms in that group under contraction and under expansion, respectively. Vertical bars mark group means. Firms with nearly identical effects are stacked vertically, so the height of a stack indicates the number of firms. The spread of the markers measures cross-firm heterogeneity. All other covariates and fixed effects remain at their 2022 values.

The revenue effects are presented in Figure 4. Under contraction with network rewiring, switchers lose 0.173 log points on average, with losses between 0.112 and 0.246. Firms the status of which is unchanged also lose: immigrant-led U.S. stayers lose 0.036, non-immigrant-led U.S. stayers 0.031, and non-U.S. firms 0.050. Under expansion, switchers gain 0.143 log points on average, with gains between 0.084 and 0.163. Both groups of U.S. stayers gain 0.011 and non-U.S. firms gain 0.014. When the observed network is fixed, the four group means are smaller in magnitude in every case: -0.135 , -0.025 , -0.017 and -0.021 under contraction, and 0.134 , 0.000 , 0.006 and 0.008 under expansion.

Table 6: Expected effects of contraction and expansion in firms immigrant leadership

Group	N	Log revenue			Network degree	
		Fixed	Rewired	SD	Rewired	SD
<i>Contraction: target 20% (16/79 US firms)</i>						
All U.S. firms	79	-0.056	-0.077	0.073	-0.170	0.308
Switchers	25	-0.135	-0.173	0.034	-0.457	0.419
Immigrant-led stayers	16	-0.025	-0.036	0.032	0.000	0.000
Non-immigrant-led stayers	38	-0.017	-0.031	0.031	-0.052	0.056
Non-U.S. firms	77	-0.021	-0.050	0.045	-0.093	0.092
<i>Expansion: target 80% (63/79 US firms)</i>						
All U.S. firms	79	0.038	0.048	0.062	0.056	0.108
Switchers	22	0.134	0.143	0.021	0.192	0.129
Immigrant-led stayers	41	0.000	0.011	0.011	0.000	0.000
Non-immigrant-led stayers	16	0.006	0.011	0.021	0.014	0.006
Non-U.S. firms	77	0.008	0.014	0.012	0.046	0.050

Notes: Each entry is the average over N firms in a group. Model parameters are fixed at the posterior means reported in Tables 4 and 5. Revenue effects are in log points and network degree effects in the number of links per firm.

Fixed holds the observed 2022 network fixed under both assignments. *Rewired* simulates both the baseline and the policy networks and averages each firm’s effect over 1,000 paired simulations. Network degree cannot change when the network is fixed, so only the rewired degree effect is reported.

SD is the standard deviation of the firm-level expected rewired effects within the group. It measures how much firms differ from one another; it is not the standard error of the average.

Four groups are defined for each policy. Switchers are the U.S. firms the immigrant-executive status of which is changed: 25 under contraction and 22 under expansion. Stayers are U.S. firms the status of which is unchanged. All U.S. firms combine all three U.S. groups. Non-U.S. firms never change their status.

Table 6 summarizes the firm-level effects by groups. Each entry is the average of the expected effects plotted in Figures 3 and 4 taken over all firms in a group. The revenue columns present results for both scenarios. Under contraction, the average log

revenue of U.S. firms falls by 0.056 when the observed network is fixed and by 0.077 when the network is rewired. The losses and gains are greater in magnitude for the rewired scenario. The losses under contraction are 0.173 for the 25 switchers, 0.036 for immigrant-led stayers, 0.031 for non-immigrant-led stayers and 0.050 for non-U.S. firms. Under expansion, switchers gain 0.143 log points, each stayer group gains about 0.011 and non-U.S. firms gain 0.014, which gives a U.S.-firm average of 0.048. The network degree columns show that switchers lose 0.457 links under contraction and gain 0.192 under expansion, immigrant-led stayers do not change, and non-immigrant-led stayers and non-U.S. firms change by less than 0.1 links on average.

Table 7: Own-status benchmarks and policy propagation

Group	N	Benchmark	Effect	SD	MCSE	Ratio
<i>Contraction: target 20%</i>						
Switchers	25	-0.113	-0.173	0.034	0.00095	1.527
Immigrant-led stayers	16	0.000	-0.036	0.032	0.00052	—
Non-immigrant-led stayers	38	0.000	-0.031	0.031	0.00052	—
Non-U.S. firms	77	0.000	-0.050	0.045	0.00043	—
<i>Expansion: target 80%</i>						
Switchers	22	0.113	0.143	0.021	0.00081	1.262
Immigrant-led stayers	41	0.000	0.011	0.011	0.00019	—
Non-immigrant-led stayers	16	0.000	0.011	0.021	0.00035	—
Non-U.S. firms	77	0.000	0.014	0.012	0.00023	—

Notes: *Benchmark* reports the effect on log revenue that a firm would experience if the policy acted only through the firm’s own immigrant-executive status. For switchers it is $\hat{\beta}_{Imm} = 0.113$, with a minus sign under contraction and a plus sign under expansion. For every firm the status of which is unchanged it is zero.

Effect reports the group-average expected effect on log revenue with network rewiring, and *SD* the standard deviation of the firm-level expected effects within the group; both are the same as in Table 6.

MCSE is the Monte Carlo standard errors of what *Effect* reports: the standard deviation of the group-average effect across the 1,000 simulations, divided by $\sqrt{1,000}$. It measures the numerical precision of simulation only and should not be used as standard errors for inference.

Ratio is *Effect* divided by *Benchmark*. It is reported for switchers only, because the benchmark of the other groups is zero and their entire effects are due to propagation.

Table 7 analyzes how much the total revenue effect is different from that coming off a firm’s change in status, if the policy acted only through a firm’s own immigrant-executive status. For switchers, this is the direct coefficient $\hat{\beta}_{Imm}$: -0.113 under contraction and 0.113 under expansion. For a firm the status of which is unchanged, it is zero. The rewired total effects for switchers are -0.173 and 0.143 , which are 1.527 and 1.262 times the benchmark. The differences of 0.060 and 0.030 log points respectively come as additional effects generated through the network, by three different channels.

First, the policy changes which firms a switcher collaborates with, which changes the value of variable IM that enters the link-formation equation. Second, it changes the share of the switcher’s collaborators that are immigrant-led, $WImm$, because other firms will have a switched status so that the switcher’s collaborators are no longer the same firms. Third, the revenue of those collaborators also changes, and this feeds back into the switcher’s own revenue through the spillover parameter λ . For the two stayer groups and for non-U.S. firms their entire policy effect comes from these three network channels: between -0.031 and -0.050 log points under contraction and between 0.011 and 0.014 under expansion. The ratio column measures the total amplification relative to the direct coefficient; it does not separate the three channels. The Monte Carlo standard error (MCSE) column shows that the simulation averages are numerically precise: every value is below 0.001 , which is much smaller than the effects themselves. As an additional exercise, Appendix 6.1 repeats the same calculations over a different range of target shares.

6 Conclusion

We develop a structural model that extends the Spatial Autoregressive (SAR) framework to account for endogenous network formation in a panel data application to estimate peer effects of R&D collaboration among pharmaceutical firms on their annual performance. When the spatial weight matrix is endogenously determined ignoring it can result in selection biases.

This paper proposes a direct and computationally tractable joint modeling approach in which the error term from the outcome equation enters directly the Probit endogenous network selection equations. That preserves conjugacy and the developed MCMC algorithm has all steps, handled by the Gibbs sampler, except one. The aim is to introduce a computationally simpler and more efficient approach to facilitate a broader application of Bayesian methods in network modeling by applied scientists.

We examine the developed Bayesian MCMC algorithm and its finite-sample performance through multiple Monte Carlo experiments and report the results in which the d.g.p. is consistent with the observed network density in the application of inter-firm R&D collaborations .

Our empirical results confirm presence of substantial endogeneity. A possible interpretation of this can be that unobserved factors, such as realized or expected productivity shocks, that increase output also lower the probability of alliance formation.

In contrast to the SAR model specifications that ignore endogeneity, our proposed framework identifies a strong, positive spillover effect, even when contextual effects are controlled for, demonstrating modeling robustness in overcoming the identification challenges posed by endogenous network formation.

Conflict of Interest

The authors declare no conflict of interest.

References

- Acemoglu, D., Carvalho, V. M., Ozdaglar, A., and Tahbaz-Salehi, A. (2012). The Network Origins of Aggregate Fluctuations. *Econometrica*, 80(5):1977–2016.
- Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2015). Systemic Risk and Stability in Financial Networks. *American Economic Review*, 105(2):564–608.
- Ahuja, G. (2000). Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study. *Administrative Science Quarterly*, 45(3):425–455.
- Albert, J. H. and Chib, S. (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association*, 88(422):669–679.
- Arp, F., Hutchings, K., and Smith, W. A. (2013). Foreign Executives in Local Organisations: An Exploration of Differences to Other Types of Expatriates.
- Bloom, N., Schankerman, M., and Van Reenen, J. (2013). Identifying Technology Spillovers and Product Market Rivalry. *Econometrica*, 81(4):1347–1393.
- Boucher, V., Bramoullé, Y., Djebbari, H., and Fortin, B. (2014). Do Peers Affect Student Achievement? Evidence from Canada Using Group Size Variation. *Journal of Applied Econometrics*, 29(1):91–109.
- Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150(1):41–55.
- Cassiman, B. and Veugelers, R. (2002). R&D Cooperation and Spillovers: Some Empirical Evidence from Belgium. *American Economic Review*, 92(4):1169–1184.
- Chung, B. W. (2020). Peers’ parents and educational attainment: The exposure effect. *Labour Economics*, 64:101812.
- D’Innocenzo, E., Lucas, A., Opschoor, A., and Zhang, X. (2024). Heterogeneity and dynamics in network models. *Journal of Applied Econometrics*, 39(1):150–173.
- Elliott, M., Golub, B., and Jackson, M. O. (2014). Financial Networks and Contagion. *American Economic Review*, 104(10):3115–3153.
- Flahaux, M.-L. and Haas, H. D. (2016). African migration: Trends, patterns, drivers. *Comparative Migration Studies*, 4(1):1.

- Frühwirth-Schnatter, S. and Frühwirth, R. (2009). Data Augmentation and MCMC for Binary and Multinomial Logit Models.
- Frühwirth-Schnatter, S., Frühwirth, R., Held, L., and Rue, H. (2008). Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data. *Statistics and Computing*, 19(4):479.
- Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In Bernardo, J. M., Berger, J. O., Dawid, P., Smith, A. F. M., Bernardo, J. M., Berger, J. O., Dawid, P., and Smith, A. F. M., editors, *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting, Dedicated to the Memory of Morris H. DeGroot, 1931–1989*, page 0. Oxford University Press.
- Gormley, I. C. and Frühwirth-Schnatter, S. (2019). Mixture of Experts Models. In *Handbook of Mixture Analysis*. Chapman and Hall/CRC.
- Greene, W. H. (2018). *Econometric Analysis*. Pearson.
- Han, X., Hsieh, C.-S., and Ko, S. I. M. (2021). Spatial Modeling Approach for Dynamic Network Formation and Interactions. *Journal of Business & Economic Statistics*, 39(1):120–135.
- Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1):153–161.
- Heidelberger, P. and Welch, P. D. (1983). Simulation Run Length Control in the Presence of an Initial Transient. *Operations Research*, 31(6):1109–1144.
- Hsieh, C.-S., König, M. D., and Liu, X. (2025). Endogenous Technology Spillovers in R&D Collaboration Networks. *The RAND Journal of Economics*, 56(4):419–443.
- Hsieh, C.-S. and Lee, L. F. (2016). A Social Interactions Model with Endogenous Friendship Formation and Selectivity. *Journal of Applied Econometrics*, 31(2):301–319.
- Hsieh, C.-S. and Lin, X. (2017). Gender and racial peer effects with endogenous network formation. *Regional Science and Urban Economics*, 67:135–147.
- Kireyev, A. and Leonidov, A. (2021). Twin trade shocks: Spillovers from US-China trade tensions. *International Economics*, 167:174–188.

- König, M. D., Liu, X., and Zenou, Y. (2019). R&D Networks: Theory, Empirics, and Policy Implications. *The Review of Economics and Statistics*, 101(3):476–491.
- Lambotte, M. (2025). Peer Effects in Binary Outcomes: Strategic Complementarity and Taste for Conformity With Endogenous Networks. *Journal of Applied Econometrics*, 40(6):608–626.
- LeSage, J. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Chapman and Hall/CRC, New York.
- Lesage, J. P. (1997). Bayesian Estimation of Spatial Autoregressive Models. *International Regional Science Review*, 20(1-2):113–129.
- Mahroum, S. (1999). Competing for the highly skilled: Europe in perspective. *Science and Public Policy*, 26(1):17–25.
- Mahroum, S. (2000). Highly skilled globetrotters: Mapping the international migration of human capital. *R&D Management*, 30(1):23–32.
- Mahroum, S. and Ansari, R. (2017). What the Data Tells Us About Immigrant Executives in the U.S. *Harvard Business Review*.
- Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3):531–542.
- Mele, A. (2022). A Structural Model of Homophily and Clustering in Social Networks. *Journal of Business & Economic Statistics*, 40(3):1377–1389.
- Munkin, M. K. (2022). Count Roy model with finite mixtures. *Journal of Applied Econometrics*, 37(6):1160–1181.
- Platonova, A. and Urso, G. (2013). Asian Immigration to the European Union, United States and Canada: An Initial Comparison. *Journal of Global Policy and Governance*, 1(2):143–156.
- Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian Inference for Logistic Models Using Pólya–Gamma Latent Variables. *Journal of the American Statistical Association*, 108(504):1339–1349.
- Powell, W. W., Koput, K. W., and Smith-Doerr, L. (1996). Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology. *Administrative Science Quarterly*, 41(1):116–145.

- Prato, M. (2025). The Global Race for Talent: Brain Drain, Knowledge Transfer, and Growth. *The Quarterly Journal of Economics*, 140(1):165–238.
- Raftery, A. E. and Lewis, S. M. (1992). Comment: One Long Run with Diagnostics: Implementation Strategies for Markov Chain Monte Carlo. *Statistical Science*, 7(4):493–497.
- Schilling, M. A. (2009). Understanding the alliance data. *Strategic Management Journal*, 30(3):233–260.
- Schilling, M. A. and Phelps, C. C. (2007). Interfirm Collaboration Networks: The Impact of Large-Scale Network Structure on Firm Innovation. *Management Science*, 53(7):1113–1126.

Appendix

6.1 Additional Marginal Effects: Immigrant-Executive Policy

The main-text policy exercise follows [Han et al. \(2021\)](#) by estimating the marginal effects at the posterior means and averaging over simulated networks. As an alternative exercise, this appendix calculates the marginal effects at 1,000 draws taken from the MCMC output approximating the entire posterior distribution, simulating 200 paired baseline and policy networks. Thus, the averages are taken with respect to the network replications and over the posterior draws. Consequently, the shaded bands represent 95% posterior credible intervals.

The horizontal axis gives target immigrant-led shares. The observed share, $41/79 = 51.9\%$, serves as the baseline (null policy), at which no firm changes status and all effects are zero. Targets below the baseline remove immigrant-led status from a fixed set of baseline immigrant-led firms, determined by a random seed, while targets above the baseline assign immigrant-led status to a fixed set of baseline non-immigrant-led firms. Non-U.S. firms remain coded as zero.

The response is monotone across the spectrum. Moving all U.S. firms to immigrant-led status increases average U.S.-firm log revenue by 0.083 under rewiring, with a 95% credible interval of (0.031, 0.137), compared with 0.071 (0.019, 0.125) when the observed network is fixed. The rewired network degree effect is 0.085 links per U.S. firm (0.065, 0.106). At the opposite endpoint, removing immigrant-led status from all U.S. firms changes average log revenue by -0.132 ($-0.194, -0.068$) and network degree by -0.300 links ($-0.367, -0.229$). These endpoint comparisons involve different numbers of switchers and therefore should not be interpreted as equal-sized directional treatments.

Table 8 reports results for representative points along the spectrum. At the 20% target, the rewired average U.S.-firm revenue effect is -0.078 , with a 95% credible interval of $(-0.12, -0.04)$, and average network degree falls by 0.169 links. At the 75% target, the corresponding effects are 0.040 (0.02, 0.07) and 0.050 links. The difference between the rewired and fixed scenarios is descriptive rather than an exact additive decomposition because the two scenarios employ different network baselines.

The group results show that the spectrum is not driven only by firms the status of which changes. At the 75% target, switchers gain 0.143 log points (0.037, 0.254), while both immigrant-led and non-immigrant-led stayers gain approximately 0.010. Since stayers

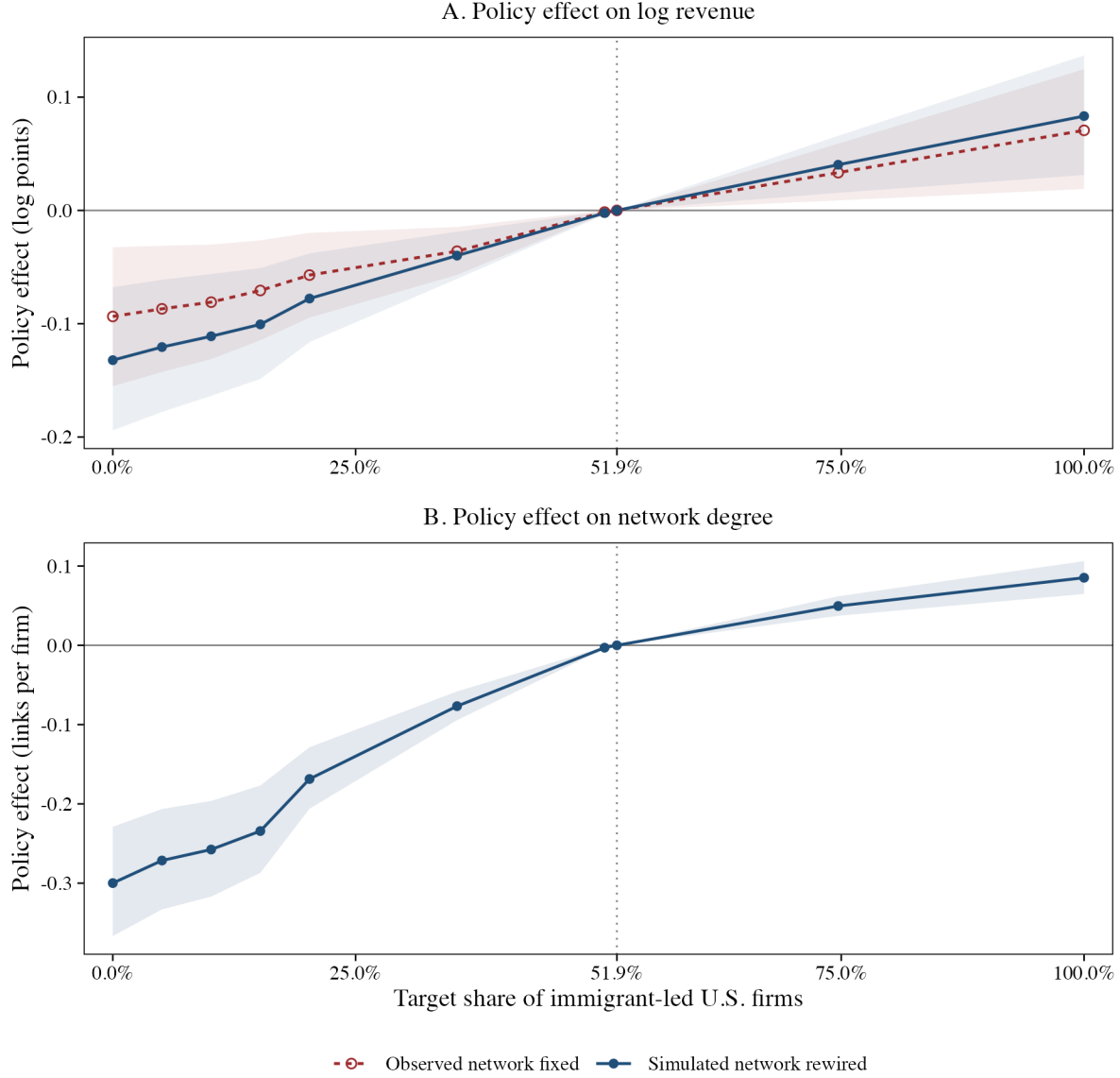


Figure 5: Posterior marginal effects under fixed observed and simulated rewired networks
Notes: Panel (a) reports average marginal effects of U.S.-firm log revenue and Panel (b) reports network degree. The dotted vertical line marks the observed 51.9% share. Results use 1,000 posterior draws and 200 paired baseline-policy network simulations within each draw. Shaded regions are 95% posterior credible intervals. The observed-network-fixed scenario uses the observed 2022 network in both calculations. The simulated-network-rewired scenario generates paired baseline and policy networks.

Table 8: Immigrant-executive policy effects across the U.S. target share spectrum

		Target share of U.S. firms that are immigrant-led				
Group	Scenario	0%	20%	50%	75%	100%
<i>Panel A. Log revenue</i>						
All U.S. firms	Fixed network	-0.094	-0.057	-0.001	0.033	0.071
		(-0.16, -0.03)	(-0.09, -0.02)	(-0.00, -0.00)	(0.01, 0.06)	(0.02, 0.12)
	Rewired network	-0.132	-0.078	-0.002	0.040	0.083
		(-0.19, -0.07)	(-0.12, -0.04)	(-0.00, -0.00)	(0.02, 0.07)	(0.03, 0.14)
Scenario difference	-0.039	-0.021	-0.001	0.007	0.013	
	(-0.05, -0.03)	(-0.03, -0.01)	(-0.00, -0.00)	(0.00, 0.01)	(0.01, 0.02)	
Switchers	Fixed network	-0.164	-0.138	-0.116	0.141	0.135
		(-0.28, -0.05)	(-0.25, -0.03)	(-0.22, -0.01)	(0.03, 0.25)	(0.03, 0.24)
	Rewired network	-0.206	-0.175	-0.134	0.143	0.155
		(-0.32, -0.09)	(-0.28, -0.07)	(-0.24, -0.02)	(0.04, 0.25)	(0.05, 0.26)
Scenario difference	-0.042	-0.037	-0.018	0.002	0.020	
	(-0.06, -0.03)	(-0.05, -0.03)	(-0.04, 0.00)	(-0.02, 0.03)	(0.01, 0.03)	
Stayers: imm.	Fixed network	—	-0.025	0.000	0.000	0.011
			(-0.04, -0.01)	(0.00, 0.00)	(0.00, 0.00)	(0.00, 0.02)
	Rewired network	—	-0.037	-0.001	0.010	0.017
			(-0.07, -0.01)	(-0.00, -0.00)	(0.00, 0.02)	(0.00, 0.03)
Scenario difference	—	-0.012	-0.001	0.010	0.006	
		(-0.02, -0.00)	(-0.00, -0.00)	(0.00, 0.02)	(0.00, 0.01)	
Stayers: non-imm.	Fixed network	-0.018	-0.017	0.000	0.005	—
		(-0.03, -0.00)	(-0.03, -0.00)	(0.00, 0.00)	(0.00, 0.01)	
	Rewired network	-0.053	-0.031	-0.000	0.010	—
		(-0.08, -0.02)	(-0.05, -0.01)	(-0.00, -0.00)	(0.00, 0.02)	
Scenario difference	-0.035	-0.014	-0.000	0.005	—	
	(-0.05, -0.02)	(-0.02, -0.01)	(-0.00, -0.00)	(0.00, 0.01)		
<i>Panel B. Network degree (links per firm)</i>						
All U.S. firms	Rewired network	-0.300	-0.169	-0.003	0.050	0.085
		(-0.37, -0.23)	(-0.21, -0.13)	(-0.00, -0.00)	(0.04, 0.06)	(0.06, 0.11)
Switchers	Rewired network	-0.507	-0.458	-0.205	0.204	0.177
		(-0.62, -0.39)	(-0.56, -0.35)	(-0.28, -0.13)	(0.15, 0.26)	(0.13, 0.22)
Stayers: imm.	Rewired network	—	0.000	0.000	0.000	0.000
			(0.00, 0.00)	(0.00, 0.00)	(0.00, 0.00)	(0.00, 0.00)
Stayers: non-imm.	Rewired network	-0.076	-0.049	-0.001	0.012	—
		(-0.09, -0.06)	(-0.06, -0.04)	(-0.00, -0.00)	(0.01, 0.02)	

Notes: Posterior means with 95% credible intervals in parentheses, from 1000 posterior draws and 200 simulated networks per draw. Effects are policy minus baseline in the 2022 cross-section. The target share is of immigrant-led status among U.S. firms only: 41 of 79 U.S. firms are immigrant-led at baseline (51.9%), and non-U.S. firms are always assigned $Imm = 0$. At the observed baseline share, no firm switches and every effect is identically zero. *Switchers* are U.S. firms the immigrant-executive status of which changes at that target share; they gain one above the baseline share and lose one below it. *Stayers* are U.S. firms the status of which is unchanged *at that specific target share*, not a fixed baseline cohort, so their group size varies along the spectrum. *Fixed network* holds the observed 2022 network fixed, so only the own and contextual channels move. *Rewired network* re-simulates link formation under the baseline and policy assignments. *Scenario difference* is rewired minus fixed. Because it compares a simulated-network scenario with an observed-network scenario, it is descriptive and is not an exact decomposition of the policy effect. Degree effects are reported for the rewired arm only; with the network held fixed, no link can change. Dashes mark groups that are empty at that target share.

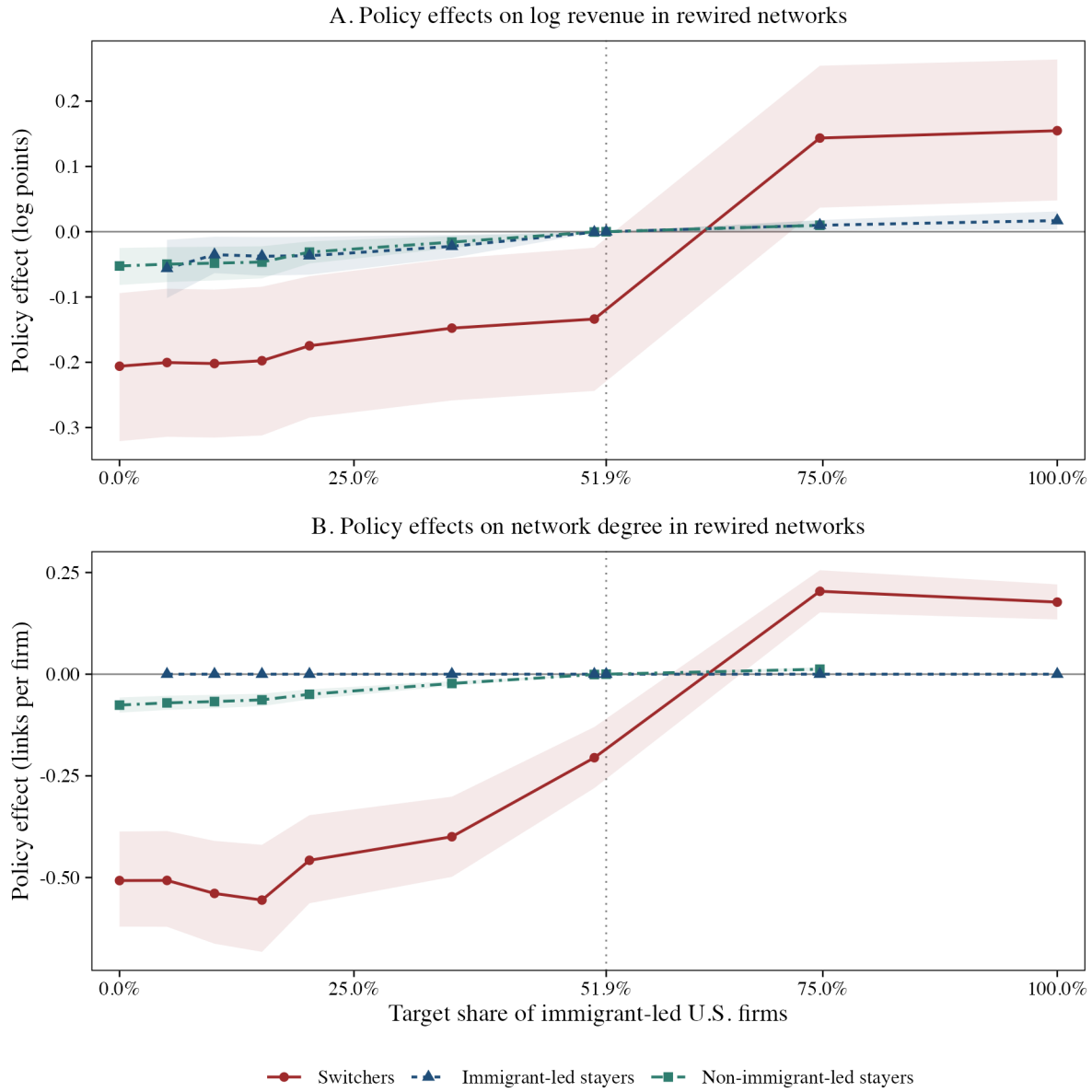


Figure 6: Posterior marginal effects group decomposition under simulated network rewiring
Notes: Switchers, immigrant-led stayers, and non-immigrant-led stayers are defined separately at each target and partition all 79 U.S. firms. The aggregate U.S.-firm effect equals the group-size-weighted average of the three group effects. Credible-interval endpoints should not be combined using those weights.

experience no change in their own *Imm* indicator, their response reflects propagation through contextual exposure, endogenous network adjustment, and spatial outcome interaction.

Table 9: Multiplier for the immigrant-executive policy at 75% target share

Group	No-multiplier benchmark	Estimated effect	95% credible interval	Ratio to benchmark
Switchers	0.113	0.143	(0.037, 0.254)	1.26
Stayers: immigrant-led	0.000	0.010	(0.002, 0.018)	—
Stayers: non-immigrant-led	0.000	0.010	(0.003, 0.017)	—

Notes: Rewired-network outcome effects at a nominal target share of 75% of U.S. firms, realized as 59 of 79 firms (74.7%), against the benchmark that would be obtained with no social interaction. Following Han et al. (2021), absent any multiplier, the switchers' effect would equal the own coefficient on the policy variable ($\hat{\beta}_{\text{Imm}} = 0.1134$), and the stayers' effect would be exactly zero. Stayer effects are therefore attributable to network spillovers; no ratio is defined against a zero benchmark.

At the 75% target, the switcher effect is 1.26 times the posterior-mean own-status coefficient of 0.113. This ratio is a compact amplification measure, not a decomposition of the separate mechanisms.

6.2 Posterior Distribution for θ

The full-conditional density for $\theta = (\Gamma', \delta)'$ is proportional to its block-diagonal Gaussian prior and to the dyadic (link) likelihood that contains θ :

- **Prior:**

$$\theta \sim \mathcal{N}(\underline{\theta}, \underline{H}_{\theta}^{-1}), \quad \underline{\theta} = \begin{pmatrix} \underline{\Gamma} \\ \underline{\delta} \end{pmatrix}, \quad \underline{H}_{\theta} = \begin{pmatrix} \underline{H}_{\Gamma} & 0 \\ 0 & \underline{H}_{\delta} \end{pmatrix}.$$

- **Link likelihood Δu_{ijt} :**

For every dyad (i, j) with $j \neq i$ and every period $t = 1, \dots, T$. The latent-utility equation is

$$\Delta u_{ijt} = C_{ijt}\Gamma + \delta \varepsilon_{it} + \eta_{ijt}, \quad \eta_{ijt} \stackrel{iid}{\sim} \mathcal{N}(0, 1),$$

which is linear in θ . Denoting $Q_{ijt} = (C_{ijt}, \varepsilon_{it})$. The SAR model does *not* involve θ once ε_{it} is treated as known in this step.

1. Log-Prior for θ

$$\log p(\theta) = -\frac{1}{2}(\theta - \underline{\theta})^{\top} \underline{H}_{\theta}(\theta - \underline{\theta}) + \text{const}$$

2. Log-Likelihood from Δu_{ijt} (all $i, j \neq i, t$)

The model is $\Delta u_{ijt} = Q_{ijt}\boldsymbol{\theta} + \eta_{ijt}$, with $\eta_{ijt} \sim \mathcal{N}(0, 1)$. Log-likelihood:

$$\log p(\Delta \mathbf{u} \mid \boldsymbol{\theta}, \cdot) = -\frac{1}{2} \sum_{t=1}^T \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n (\Delta u_{ijt} - Q_{ijt}\boldsymbol{\theta})^2 + \text{const}$$

3. Combined Log-Posterior

$$\begin{aligned} \log p(\boldsymbol{\theta} \mid \cdot) &= -\frac{1}{2}(\boldsymbol{\theta} - \underline{\boldsymbol{\theta}})^\top H_{\boldsymbol{\theta}}(\boldsymbol{\theta} - \underline{\boldsymbol{\theta}}) \\ &\quad -\frac{1}{2} \sum_{t=1}^T \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n (\Delta u_{ijt} - Q_{ijt}\boldsymbol{\theta})^2 + \text{const} \end{aligned}$$

4. Expand and Collect Terms

- **Quadratic term** (coefficient of $\boldsymbol{\theta}^\top \boldsymbol{\theta}$):

$$H_{\boldsymbol{\theta}} + \sum_{t=1}^T \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n Q_{ijt}^\top Q_{ijt}$$

- **Linear term** (coefficient of $\boldsymbol{\theta}$):

$$H_{\boldsymbol{\theta}}\underline{\boldsymbol{\theta}} + \sum_{t=1}^T \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n Q_{ijt}^\top \Delta u_{ijt}$$

5. Posterior Parameters

$$\begin{aligned} \overline{H}_{\boldsymbol{\theta}} &= H_{\boldsymbol{\theta}} + \sum_{t=1}^T \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n Q_{ijt}^\top Q_{ijt}, \\ \overline{\boldsymbol{\theta}} &= \overline{H}_{\boldsymbol{\theta}}^{-1} \left(H_{\boldsymbol{\theta}}\underline{\boldsymbol{\theta}} + \sum_{t=1}^T \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n Q_{ijt}^\top \Delta u_{ijt} \right) \end{aligned}$$

6.3 Posterior Distribution of β

The full conditional density for the slope vector $p(\beta \mid \mathbf{y}, \Delta \mathbf{u}, \cdot)$ is proportional to its Gaussian prior and to every likelihood term that contains β :

- **Prior:** $\beta \sim \mathcal{N}(\underline{\beta}, \underline{H}_\beta^{-1})$
- **Outcome likelihood y_{it} :** for every individual $i = 1, \dots, n$ and every period $t = 1, \dots, T$, because β enters the mean through the regressor block $Z_{it}\beta$.
- **Dyadic likelihood Δu_{ijt} :** for every sender–receiver pair (i, j) with $j \neq i$ across the same T periods, since the sender’s error term $\varepsilon_{it} = r_{it} - Z_{it}\beta$ appears in the latent utility equation whenever i is the sender.

1. Log-Prior for β

$$\log p(\beta) = -\frac{1}{2}(\beta - \underline{\beta})' H_\beta (\beta - \underline{\beta}) + \text{const}$$

2. Log-Likelihood from y_{it} (all i, t)

$$\begin{aligned} r_{it} &= y_{it} - \lambda \sum_{j=1}^n w_{ijt} y_{jt} - \alpha_i - \tau_t \\ \varepsilon_{it} &= r_{it} - Z_{it}\beta \\ \log p(\mathbf{y} \mid \cdot) &= -\frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^n \sum_{t=1}^T (r_{it} - Z_{it}\beta)^2 + \text{const} \end{aligned}$$

3. Log-Likelihood from Δu_{ijt} (all $i, j \neq i, t$)

$$\begin{aligned} s_{ijt} &= \Delta u_{ijt} - c_{it}\gamma_1 - c_{jt}\gamma_2 - c_{ijt}\gamma_3 \\ \Delta u_{ijt} &= s_{ijt} + \delta\varepsilon_{it} + \eta_{ijt}, \quad \varepsilon_{it} = r_{it} - Z_{it}\beta \\ \log p(\Delta \mathbf{u} \mid \cdot) &= -\frac{1}{2} \sum_{i=1}^n \sum_{t=1}^T \sum_{j \neq i} (s_{ijt} - \delta(r_{it} - Z_{it}\beta))^2 + \text{const} \end{aligned}$$

4. Combined Log-Posterior

$$\begin{aligned}\log p(\boldsymbol{\beta} \mid \cdot) &= -\frac{1}{2}(\boldsymbol{\beta} - \underline{\boldsymbol{\beta}})' H_{\boldsymbol{\beta}} (\boldsymbol{\beta} - \underline{\boldsymbol{\beta}}) \\ &\quad - \frac{1}{2\sigma_{\varepsilon}^2} \sum_{i=1}^n \sum_{t=1}^T (r_{it} - Z_{it}\boldsymbol{\beta})^2 \\ &\quad - \frac{1}{2} \sum_{i=1}^n \sum_{t=1}^T \sum_{j \neq i} (s_{ijt} - \delta(r_{it} - Z_{it}\boldsymbol{\beta}))^2 + \text{const}\end{aligned}$$

5. Expand and Collect Terms

Quadratic Term:

$$H_{\boldsymbol{\beta}} + \frac{1}{\sigma_{\varepsilon}^2} \sum_{i=1}^n \sum_{t=1}^T Z'_{it} Z_{it} + \delta^2 \sum_{i=1}^n \sum_{t=1}^T \sum_{j \neq i} Z'_{it} Z_{it} = H_{\boldsymbol{\beta}} + \left(\frac{1}{\sigma_{\varepsilon}^2} + \delta^2(n-1) \right) \sum_{i=1}^n \sum_{t=1}^T Z'_{it} Z_{it}$$

Linear Term:

$$H_{\boldsymbol{\beta}} \underline{\boldsymbol{\beta}} + \left(\frac{1}{\sigma_{\varepsilon}^2} + \delta^2(n-1) \right) \sum_{i=1}^n \sum_{t=1}^T r_{it} Z'_{it} - \delta \sum_{i=1}^n \sum_{t=1}^T \sum_{j \neq i} s_{ijt} Z'_{it}$$

6. Posterior Parameters

$$\begin{aligned}\overline{H}_{\boldsymbol{\beta}} &= H_{\boldsymbol{\beta}} + \left(\frac{1}{\sigma_{\varepsilon}^2} + \delta^2(n-1) \right) \sum_{i=1}^n \sum_{t=1}^T Z'_{it} Z_{it} \\ \overline{\boldsymbol{\beta}} &= \overline{H}_{\boldsymbol{\beta}}^{-1} \left[H_{\boldsymbol{\beta}} \underline{\boldsymbol{\beta}} + \left(\frac{1}{\sigma_{\varepsilon}^2} + \delta^2(n-1) \right) \sum_{i=1}^n \sum_{t=1}^T r_{it} Z'_{it} - \delta \sum_{i=1}^n \sum_{t=1}^T \sum_{j \neq i} s_{ijt} Z'_{it} \right]\end{aligned}$$

Substituting r_{it} and s_{ijt} :

$$\begin{aligned}\overline{\boldsymbol{\beta}} &= \overline{H}_{\boldsymbol{\beta}}^{-1} \left[H_{\boldsymbol{\beta}} \underline{\boldsymbol{\beta}} + \left(\frac{1}{\sigma_{\varepsilon}^2} + \delta^2(n-1) \right) \sum_{i=1}^n \sum_{t=1}^T \left(y_{it} - \lambda \sum_{j=1}^n w_{ijt} y_{jt} - \alpha_i - \tau_t \right) Z'_{it} \right. \\ &\quad \left. - \delta \sum_{i=1}^n \sum_{t=1}^T \sum_{j \neq i} (\Delta u_{ijt} - c_{it}\gamma_1 - c_{jt}\gamma_2 - c_{ijt}\gamma_3) Z'_{it} \right]\end{aligned}$$

6.4 Posterior Distribution of α_i

The full conditional density for a single individual fixed effect $p(\alpha_i \mid \mathbf{y}, \Delta \mathbf{u}, \cdot)$ is proportional to its Gaussian prior and to every likelihood term that contains α_i :

- **Prior:** $\alpha_i \sim \mathcal{N}(\underline{\alpha}_i, H_{\alpha_i}^{-1})$
- **Outcome likelihood y_{it} :** for the same individual i across all time periods $t = 1, \dots, T$, because α_i appears in the mean of each y_{it} only through the index i .
- **Dyadic likelihood Δu_{ijt} :** for each pair (i, j) with sender i and $j \neq i$, across the same T periods, because $\varepsilon_{it} = y_{it} - \lambda \sum_{j=1}^n w_{ijt} y_{jt} - \dots - \alpha_i$ enters the latent utility model whenever i is the sender.

1. Log-Prior:

$$\log p(\alpha_i) = -\frac{1}{2}(\alpha_i - \underline{\alpha}_i)^T H_{\alpha_i}(\alpha_i - \underline{\alpha}_i) + \text{const}$$

2. Log-Likelihood from y_{it} :

Define residual:

$$r_{it} = y_{it} - \lambda \sum_j w_{ijt} y_{jt} - x_{it} \beta_1 - \sum_j w_{ijt} x_{jt} \beta_2 - \tau_t,$$

so that $\varepsilon_{it} = r_{it} - \alpha_i$.

Then the log-likelihood becomes:

$$\log p(y \mid \cdot) = -\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^T (r_{it} - \alpha_i)^2 + \text{const}$$

3. Log-Likelihood from Δu_{ijt} :

Define:

$$s_{ijt} = \Delta u_{ijt} - c_{it} \gamma_1 - c_{jt} \gamma_2 - c_{ijt} \gamma_3,$$

so that the model is:

$$\Delta u_{ijt} = s_{ijt} + \delta \varepsilon_{it} + \eta_{ijt},$$

and substituting $\varepsilon_{it} = r_{it} - \alpha_i$, we get:

$$\log p(\Delta u \mid \cdot) = -\frac{1}{2} \sum_{t=1}^T \sum_{j \neq i} (s_{ijt} - \delta(r_{it} - \alpha_i))^2 + \text{const}$$

4. Combined Log-Posterior:

$$\begin{aligned}\log p(\alpha_i \mid \cdot) &= -\frac{1}{2}H_{\alpha_i}(\alpha_i - \underline{\alpha}_i)^2 - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^T (r_{it} - \alpha_i)^2 \\ &\quad - \frac{1}{2} \sum_{t=1}^T \sum_{j \neq i} (s_{ijt} - \delta(r_{it} - \alpha_i))^2 + \text{const}\end{aligned}$$

Expanding and collecting terms:

5. Posterior Quadratic Term (coefficient of α_i^2):

$$H_{\alpha_i} + \frac{T}{\sigma_\varepsilon^2} + \delta^2 \sum_{t=1}^T \sum_{j \neq i} 1 = H_{\alpha_i} + \frac{T}{\sigma_\varepsilon^2} + \delta^2 T(n-1)$$

Linear Term (coefficient of α_i):

$$H_{\alpha_i} \underline{\alpha}_i + \frac{1}{\sigma_\varepsilon^2} \sum_{t=1}^T r_{it} - \delta \sum_{t=1}^T \sum_{j \neq i} (s_{ijt} - \delta r_{it})$$

6. Posterior Parameters:

The full conditional is:

$$\alpha_i \sim \mathcal{N}(\bar{\alpha}_i, \bar{H}_{\alpha_i}^{-1}),$$

where:

$$\begin{aligned}\bar{H}_{\alpha_i} &= H_{\alpha_i} + T \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n-1) \right), \\ \bar{\alpha}_i &= \bar{H}_{\alpha_i}^{-1} \left[H_{\alpha_i} \underline{\alpha}_i + \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n-1) \right) \sum_{t=1}^T r_{it} - \delta \sum_{t=1}^T \sum_{j \neq i} s_{ijt} \right]\end{aligned}$$

Substituting expressions for r_{it} and s_{ijt} :

$$\begin{aligned}\bar{\alpha}_i &= \bar{H}_{\alpha_i}^{-1} \left[\underline{H}_{\alpha_i} \underline{\alpha}_i + \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n-1) \right) \sum_{t=1}^T \left(y_{it} - \lambda \sum_{j=1}^n w_{ijt} y_{jt} - x_{it} \beta_1 - \sum_{j=1}^n w_{ijt} x_{jt} \beta_2 - \tau_t \right) \right. \\ &\quad \left. - \delta \sum_{t=1}^T \sum_{\substack{j=1 \\ j \neq i}}^n (\Delta u_{ijt} - c_{it} \gamma_1 - c_{jt} \gamma_2 - c_{ijt} \gamma_3) \right]\end{aligned}$$

6.5 Posterior Distribution of τ_t

The full conditional density for a single time fixed effect $p(\tau_t \mid \mathbf{y}, \Delta \mathbf{u}, \cdot)$ is proportional to its Gaussian prior and to every likelihood term that contains τ_t :

- **Prior:** $\tau_t \sim \mathcal{N}(\underline{\tau}_t, H_{\tau_t}^{-1})$
- **Outcome likelihood y_{it} :** for all individuals $i = 1, \dots, n$ in the same period t , because τ_t appears in the mean of every y_{it} only through the time index t .
- **Dyadic likelihood Δu_{ijt} :** for every dyad (i, j) with $i \neq j$ in that same period t , since the sender's structural error $\varepsilon_{it} = y_{it} - \lambda \sum_{j=1}^n w_{ijt} y_{jt} - Z_{it} \beta - \alpha_i - \tau_t$ appears in the latent utility equation whenever the time index is t .

1. Log-Prior for τ_t :

$$\log p(\tau_t) = -\frac{1}{2}(\tau_t - \underline{\tau}_t)^T H_{\tau_t} (\tau_t - \underline{\tau}_t) + \text{const}$$

2. Log-Likelihood from y_{it} at time t :

Define the residual excluding τ_t :

$$r_{it} = y_{it} - \lambda \sum_{j=1}^n w_{ijt} y_{jt} - x_{it} \beta_1 - \sum_{j=1}^n w_{ijt} x_{jt} \beta_2 - \alpha_i$$

Then $\varepsilon_{it} = r_{it} - \tau_t$, so the log-likelihood becomes:

$$\log p(y_t \mid \cdot) = -\frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^n (r_{it} - \tau_t)^2 + \text{const}$$

3. Log-Likelihood from Δu_{ijt} at time t :

Define:

$$s_{ijt} = \Delta u_{ijt} - c_{it} \gamma_1 - c_{jt} \gamma_2 - c_{ijt} \gamma_3$$

Then:

$$\Delta u_{ijt} = s_{ijt} + \delta \varepsilon_{it} + \eta_{ijt}$$

Substitute $\varepsilon_{it} = r_{it} - \tau_t$:

$$\log p(\Delta u_t \mid \cdot) = -\frac{1}{2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n (s_{ijt} - \delta(r_{it} - \tau_t))^2 + \text{const}$$

4. Combined Log-Posterior:

$$\begin{aligned} \log p(\tau_t \mid \cdot) &= -\frac{1}{2}H_{\tau_t}(\tau_t - \underline{\tau}_t)^2 - \frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^n (r_{it} - \tau_t)^2 \\ &\quad - \frac{1}{2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n (s_{ijt} - \delta(r_{it} - \tau_t))^2 + \text{const} \end{aligned}$$

5. Expand and Collect Terms

Quadratic term (coefficient of τ_t^2):

$$H_{\tau_t} + \frac{n}{\sigma_\varepsilon^2} + \delta^2 \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n 1 = H_{\tau_t} + \frac{n}{\sigma_\varepsilon^2} + \delta^2 n(n-1)$$

Linear term (coefficient of τ_t):

$$H_{\tau_t} \underline{\tau}_t + \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^n r_{it} - \delta \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n (s_{ijt} - \delta r_{it})$$

6. Posterior Parameters

The full conditional is:

$$\tau_t \sim \mathcal{N}(\bar{\tau}_t, \bar{H}_{\tau_t}^{-1})$$

where:

$$\begin{aligned} \bar{H}_{\tau_t} &= H_{\tau_t} + \frac{n}{\sigma_\varepsilon^2} + \delta^2 n(n-1), \\ \bar{\tau}_t &= \bar{H}_{\tau_t}^{-1} \left[H_{\tau_t} \underline{\tau}_t + \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n-1) \right) \sum_{i=1}^n r_{it} - \delta \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n s_{ijt} \right] \end{aligned}$$

Optionally, substitute expressions for r_{it} and s_{ijt} :

$$\begin{aligned} \bar{\tau}_t = & \bar{H}_{\tau_t}^{-1} \left[H_{\tau_t} \underline{\tau}_t + \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n-1) \right) \sum_{i=1}^n \left(y_{it} - \lambda \sum_{j=1}^n w_{ijt} y_{jt} - x_{it} \beta_1 - \sum_{j=1}^n w_{ijt} x_{jt} \beta_2 - \alpha_i \right) \right. \\ & \left. - \delta \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n (\Delta u_{ijt} - c_{it} \gamma_1 - c_{jt} \gamma_2 - c_{ijt} \gamma_3) \right] \end{aligned}$$

6.6 Posterior Distribution of σ_ε^{-2}

1. Prior Distribution

Given the prior:

$$\sigma_\varepsilon^{-2} \sim \text{Gamma} \left(\frac{\nu}{2}, \frac{g}{2} \right),$$

the prior density is:

$$p(\sigma_\varepsilon^{-2}) \propto (\sigma_\varepsilon^{-2})^{\frac{\nu}{2}-1} \exp \left(-\frac{g}{2} \sigma_\varepsilon^{-2} \right).$$

2. Likelihood Contribution

The Gaussian likelihood for ε_{it} (from the SAR model) is:

$$\varepsilon_{it} \sim \mathcal{N}(0, \sigma_\varepsilon^2), \quad \text{independent for } i = 1, \dots, n; t = 1, \dots, T.$$

The joint likelihood for all ε_{it} is:

$$p(y \mid \cdot) \propto (\sigma_\varepsilon^2)^{-nT/2} \exp \left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^T \sum_{i=1}^n \varepsilon_{it}^2 \right).$$

Substituting $\sigma_\varepsilon^2 = (\sigma_\varepsilon^{-2})^{-1}$, we get:

$$p(y \mid \cdot) \propto (\sigma_\varepsilon^{-2})^{nT/2} \exp \left(-\frac{\sigma_\varepsilon^{-2}}{2} \sum_{t=1}^T \sum_{i=1}^n \varepsilon_{it}^2 \right).$$

3. Full Conditional (Proportionality)

The full conditional satisfies:

$$p(\sigma_\varepsilon^{-2} \mid \cdot) \propto p(\sigma_\varepsilon^{-2}) \cdot p(y \mid \cdot).$$

Combining Steps 1 and 2:

$$\begin{aligned} p(\sigma_\varepsilon^{-2} \mid \cdot) &\propto (\sigma_\varepsilon^{-2})^{\frac{\nu}{2}-1} \exp\left(-\frac{g}{2}\sigma_\varepsilon^{-2}\right) \cdot (\sigma_\varepsilon^{-2})^{nT/2} \exp\left(-\frac{\sigma_\varepsilon^{-2}}{2} \sum_{t=1}^T \sum_{i=1}^n \varepsilon_{it}^2\right) \\ &\propto (\sigma_\varepsilon^{-2})^{\frac{\nu+nT}{2}-1} \exp\left(-\sigma_\varepsilon^{-2} \left[\frac{g}{2} + \frac{1}{2} \sum_{t=1}^T \sum_{i=1}^n \varepsilon_{it}^2\right]\right). \end{aligned}$$

4. Posterior Distribution

Thus, the full conditional posterior is:

$$\sigma_\varepsilon^{-2} \mid \cdot \sim \text{Gamma}\left(\frac{\nu + nT}{2}, \frac{g}{2} + \frac{1}{2} \sum_{t=1}^T \sum_{i=1}^n \varepsilon_{it}^2\right).$$

6.7 Unbalanced Panel

Observed sets and counts. For each period $t = 1, \dots, T$, let

$$\mathcal{I}_t = \{i \in \{1, \dots, n\} : (i, t) \text{ is observed}\}, \quad n_t = |\mathcal{I}_t|.$$

For each individual i , let

$$\mathcal{T}_i = \{t \in \{1, \dots, T\} : (i, t) \text{ is observed}\}.$$

The total number of observed outcomes is $N_{\text{obs}} = \sum_{t=1}^T n_t$.

Denote $Z_t = (X_t, W_t X_t)$, $Z_{it} = (x_{it}, \sum_j w_{ijt} x_{jt})$, $C_{ijt} = (c_{it}, c_{jt}, c_{ijt})$, $Q_{ijt} = (C_{ijt}, \varepsilon_{it})$, $\Gamma = (\gamma_1, \gamma_2, \gamma_3)'$, $\theta' = (\Gamma', \delta)$, $\beta = (\beta_1, \beta_2)'$, and $\Theta = \{\lambda, \beta, \alpha, \tau, \sigma_\varepsilon, \theta\}$. Let $\mathcal{I}_t = \{i : (i, t) \text{ is observed}\}$ and $n_t := |\mathcal{I}_t|$. For each t , W_t is the $n_t \times n_t$ submatrix of the spatial weight matrix whose rows/columns are indexed by $\mathcal{I}_t$ in the same order as the stacked outcome vector at t . Recall the SAR residual for each observed (i, t) :

$$\varepsilon_{it} := y_{it} - \lambda \sum_{r=1}^{n_t} w_{irt} y_{rt} - \alpha_i - \tau_t - Z_{it} \beta.$$

The joint kernel of the observed data

$$\begin{aligned}
& P(Y, \Delta u \mid W, \Theta) \\
& \propto \prod_{t=1}^T \left\{ (2\pi\sigma_\varepsilon^2)^{-n_t/2} |I_{n_t} - \lambda W_t| \exp\left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^{n_t} \varepsilon_{it}^2\right) \times \exp\left(-\frac{1}{2} \sum_{i=1}^{n_t} \sum_{j \neq i}^{n_t} [\Delta u_{ijt} - C_{ijt}\Gamma - \delta \varepsilon_{it}]^2\right) \right\} \\
& = \prod_{t=1}^T \left\{ (2\pi\sigma_\varepsilon^2)^{-n_t/2} |I_{n_t} - \lambda W_t| \exp\left(-\frac{1}{2} \left[\frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^{n_t} \varepsilon_{it}^2 + \sum_{i=1}^{n_t} \sum_{j \neq i}^{n_t} [\Delta u_{ijt} - C_{ijt}\Gamma - \delta \varepsilon_{it}]^2 \right] \right) \right\} \\
& = \prod_{t=1}^T \left\{ (2\pi\sigma_\varepsilon^2)^{-n_t/2} |I_{n_t} - \lambda W_t| \exp\left(-\frac{1}{2} \left[\sum_{i=1}^{n_t} \sum_{j \neq i}^{n_t} [\Delta u_{ijt} - C_{ijt}\Gamma]^2 \right. \right. \right. \\
& \quad \left. \left. - 2\delta \sum_{i=1}^{n_t} \varepsilon_{it} \sum_{j \neq i}^{n_t} [\Delta u_{ijt} - C_{ijt}\Gamma] + \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n_t - 1)\right) \sum_{i=1}^{n_t} \varepsilon_{it}^2 \right] \right) \right\} \\
& = (2\pi\sigma_\varepsilon^2)^{-\frac{1}{2} \sum_{t=1}^T n_t} \left[\prod_{t=1}^T |I_{n_t} - \lambda W_t| \right] \exp\left(-\frac{1}{2} \left[\sum_{t=1}^T \sum_{i=1}^{n_t} \sum_{j \neq i}^{n_t} [\Delta u_{ijt} - C_{ijt}\Gamma]^2 \right. \right. \\
& \quad \left. \left. - 2\delta \sum_{t=1}^T \sum_{i=1}^{n_t} \varepsilon_{it} \sum_{j \neq i}^{n_t} [\Delta u_{ijt} - C_{ijt}\Gamma] + \sum_{t=1}^T \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n_t - 1)\right) \sum_{i=1}^{n_t} \varepsilon_{it}^2 \right] \right).
\end{aligned}$$

Substituting ε_{it} :

$$\begin{aligned}
& P(Y, \Delta u \mid W, \Theta) \\
& \propto (2\pi\sigma_\varepsilon^2)^{-\frac{1}{2} N_{\text{obs}}} \left[\prod_{t=1}^T |I_{n_t} - \lambda W_t| \right] \\
& \quad \times \exp\left(-\frac{1}{2} \left[\sum_{t=1}^T \sum_{i=1}^{n_t} \sum_{j \neq i}^{n_t} (\Delta u_{ijt} - C_{ijt}\Gamma)^2 \right. \right. \\
& \quad \left. - 2\delta \sum_{t=1}^T \sum_{i=1}^{n_t} \left(y_{it} - \lambda \sum_{k=1}^{n_t} w_{ikt} y_{kt} - Z_{it}\boldsymbol{\beta} - \alpha_i - \tau_t \right) \sum_{j \neq i}^{n_t} (\Delta u_{ijt} - C_{ijt}\Gamma) \right. \\
& \quad \left. \left. + \sum_{t=1}^T (\sigma_\varepsilon^{-2} + \delta^2(n_t - 1)) \sum_{i=1}^{n_t} \left(y_{it} - \lambda \sum_{k=1}^{n_t} w_{ikt} y_{kt} - Z_{it}\boldsymbol{\beta} - \alpha_i - \tau_t \right)^2 \right] \right).
\end{aligned}$$

Posterior of β :

$$\begin{aligned}\overline{H}_\beta &= \underline{H}_\beta + \sum_{t=1}^T \sum_{i=1}^{n_t} (\sigma_\varepsilon^{-2} + \delta^2(n_t - 1)) Z'_{it} Z_{it}, \\ \overline{\beta} &= \overline{H}_\beta^{-1} \left[\underline{H}_\beta \underline{\beta} + \sum_{t=1}^T \sum_{i=1}^{n_t} \left\{ (\sigma_\varepsilon^{-2} + \delta^2(n_t - 1)) Z'_{it} \left(y_{it} - \lambda \sum_{k=1}^{n_t} w_{ikt} y_{kt} - \alpha_i - \tau_t \right) \right. \right. \\ &\quad \left. \left. - \delta Z'_{it} \sum_{j \neq i}^{n_t} (\Delta u_{ijt} - C_{ijt} \Gamma) \right\} \right].\end{aligned}$$

Posterior of α_i (normalization $\alpha_1 = 0$):

$$\begin{aligned}\overline{H}_{\alpha_i} &= \underline{H}_{\alpha_i} + \sum_{t=1}^T (\sigma_\varepsilon^{-2} + \delta^2(n_t - 1)), \\ \overline{\alpha}_i &= \overline{H}_{\alpha_i}^{-1} \left[\underline{H}_{\alpha_i} \underline{\alpha}_i + \sum_{t=1}^T \left\{ (\sigma_\varepsilon^{-2} + \delta^2(n_t - 1)) \left(y_{it} - \lambda \sum_{k=1}^{n_t} w_{ikt} y_{kt} - Z_{it} \beta - \tau_t \right) \right. \right. \\ &\quad \left. \left. - \delta \sum_{j \neq i}^{n_t} (\Delta u_{ijt} - C_{ijt} \Gamma) \right\} \right].\end{aligned}$$

Posterior of τ_t (normalization $\tau_1 = 0$):

$$\begin{aligned}\overline{H}_{\tau_t} &= \underline{H}_{\tau_t} + \sum_{i=1}^{n_t} (\sigma_\varepsilon^{-2} + \delta^2(n_t - 1)) = \underline{H}_{\tau_t} + n_t (\sigma_\varepsilon^{-2} + \delta^2(n_t - 1)), \\ \overline{\tau}_t &= \overline{H}_{\tau_t}^{-1} \left[\underline{H}_{\tau_t} \underline{\tau}_t + \sum_{i=1}^{n_t} \left\{ (\sigma_\varepsilon^{-2} + \delta^2(n_t - 1)) \left(y_{it} - \lambda \sum_{k=1}^{n_t} w_{ikt} y_{kt} - Z_{it} \beta - \alpha_i \right) \right. \right. \\ &\quad \left. \left. - \delta \sum_{j \neq i}^{n_t} (\Delta u_{ijt} - C_{ijt} \Gamma) \right\} \right].\end{aligned}$$

Posterior for σ_ε^{-2} (shape-rate). With the prior $\sigma_\varepsilon^{-2} \sim \text{Gamma}(\frac{\nu}{2}, \frac{g}{2})$,

$$\sigma_\varepsilon^{-2} \mid \text{rest} \sim \text{Gamma} \left(\frac{\nu + \sum_{t=1}^T n_t}{2}, \frac{g + \sum_{t=1}^T \sum_{i=1}^{n_t} \varepsilon_{it}^2}{2} \right).$$

Posterior of $\theta = (\Gamma', \delta)'$. For each observed dyad (i, j, t) with $i, j \in \mathcal{I}_t$ and $j \neq i$, let $Q_{ijt} = (C_{ijt}, \varepsilon_{it})$ and $\Delta u_{ijt} = Q_{ijt} \theta + \eta_{ijt}$, where $\eta_{ijt} \sim \mathcal{N}(0, 1)$. With prior $\theta \sim \mathcal{N}(\underline{\theta}, \underline{H}_\theta^{-1})$,

$$\begin{aligned}\overline{H}_\theta &= \underline{H}_\theta + \sum_{t=1}^T \sum_{i=1}^{n_t} \sum_{j \neq i}^{n_t} Q'_{ijt} Q_{ijt}, \\ \bar{\theta} &= \overline{H}_\theta^{-1} \left[\underline{H}_\theta \underline{\theta} + \sum_{t=1}^T \sum_{i=1}^{n_t} \sum_{j \neq i}^{n_t} Q'_{ijt} \Delta u_{ijt} \right].\end{aligned}$$

Full conditional density of λ

$$\begin{aligned}& \pi(\lambda \mid Y, \Delta u, W, \Theta \mid \{\lambda\}) \\ & \propto \prod_{t=1}^T |I_{n_t} - \lambda W_t| \times \exp \left\{ \delta \sum_{t=1}^T \sum_{i=1}^{n_t} \left(y_{it} - \lambda \sum_{k=1}^{n_t} w_{ikt} y_{kt} - Z_{it} \boldsymbol{\beta} - \alpha_i - \tau_t \right) \sum_{j \neq i}^{n_t} (\Delta u_{ijt} - C_{ijt} \Gamma) \right. \\ & \quad \left. - \frac{1}{2} \sum_{t=1}^T \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n_t - 1) \right) \sum_{i=1}^{n_t} \left(y_{it} - \lambda \sum_{k=1}^{n_t} w_{ikt} y_{kt} - Z_{it} \boldsymbol{\beta} - \alpha_i - \tau_t \right)^2 \right\}.\end{aligned}$$

This can be reduced to

$$\begin{aligned}& \pi(\lambda \mid Y, \Delta u, W, \Theta \mid \{\lambda\}) \\ & \propto \prod_{t=1}^T |I_{n_t} - \lambda W_t| \times \exp \left\{ -\lambda \delta \sum_{t=1}^T \sum_{i=1}^{n_t} \left(\sum_{k=1}^{n_t} w_{ikt} y_{kt} \right) \sum_{j \neq i}^{n_t} (\Delta u_{ijt} - C_{ijt} \Gamma) \right. \\ & \quad \left. - \frac{1}{2} \sum_{t=1}^T \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n_t - 1) \right) \sum_{i=1}^{n_t} \left(\lambda \sum_{k=1}^{n_t} w_{ikt} y_{kt} - (y_{it} - Z_{it} \boldsymbol{\beta} - \alpha_i - \tau_t) \right)^2 \right\}.\end{aligned}$$

Log of density:

$$\begin{aligned}& \log \pi(\lambda \mid Y, \Delta u, W, \Theta \mid \{\lambda\}) \\ & \propto \sum_{t=1}^T \log |I_{n_t} - \lambda W_t| - \lambda \delta \sum_{t=1}^T \sum_{i=1}^{n_t} \left(\sum_{k=1}^{n_t} w_{ikt} y_{kt} \right) \sum_{j \neq i}^{n_t} (\Delta u_{ijt} - C_{ijt} \Gamma) \\ & \quad - \frac{1}{2} \sum_{t=1}^T \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n_t - 1) \right) \sum_{i=1}^{n_t} \left(\lambda \sum_{k=1}^{n_t} w_{ikt} y_{kt} - (y_{it} - Z_{it} \boldsymbol{\beta} - \alpha_i - \tau_t) \right)^2.\end{aligned}$$

Quadratic expansion in λ

$$\begin{aligned}
& \log \pi(\lambda \mid Y, \Delta u, W, \Theta \mid \{\lambda\}) \\
& \propto \sum_{t=1}^T \log |I_{n_t} - \lambda W_t| - \frac{1}{2} \sum_{t=1}^T \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n_t - 1) \right) \sum_{i=1}^{n_t} \left(\sum_{k=1}^{n_t} w_{ikt} y_{kt} \right)^2 \lambda^2 \\
& + \left[\sum_{t=1}^T \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n_t - 1) \right) \sum_{i=1}^{n_t} (y_{it} - Z_{it} \boldsymbol{\beta} - \alpha_i - \tau_t) \left(\sum_{k=1}^{n_t} w_{ikt} y_{kt} \right) \right. \\
& \quad \left. - \delta \sum_{t=1}^T \sum_{i=1}^{n_t} \left(\sum_{k=1}^{n_t} w_{ikt} y_{kt} \right) \sum_{j \neq i}^{n_t} (\Delta u_{ijt} - C_{ijt} \Gamma) \right] \lambda.
\end{aligned}$$

Since

$$\begin{aligned}
& \frac{\partial}{\partial \lambda} \sum_{t=1}^T \log |I_{n_t} - \lambda W_t| = \frac{\partial}{\partial \lambda} \sum_{t=1}^T \log \left(\prod_{r=1}^{n_t} (1 - \lambda \mu_{rt}) \right) \\
& = \sum_{t=1}^T \sum_{r=1}^{n_t} \frac{\partial}{\partial \lambda} \log(1 - \lambda \mu_{rt}) = - \sum_{t=1}^T \sum_{r=1}^{n_t} \frac{\mu_{rt}}{1 - \lambda \mu_{rt}},
\end{aligned}$$

$$\begin{aligned}
& \frac{\partial}{\partial \lambda} \log \pi(\lambda \mid Y, \Delta u, W, \Theta \mid \{\lambda\}) \\
& = - \sum_{t=1}^T \sum_{r=1}^{n_t} \frac{\mu_{rt}}{1 - \lambda \mu_{rt}} - \left\{ \sum_{t=1}^T \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n_t - 1) \right) \sum_{i=1}^{n_t} \left(\sum_{k=1}^{n_t} w_{ikt} y_{kt} \right)^2 \right\} \lambda \\
& + \sum_{t=1}^T \left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n_t - 1) \right) \sum_{i=1}^{n_t} (y_{it} - Z_{it} \boldsymbol{\beta} - \alpha_i - \tau_t) \left(\sum_{k=1}^{n_t} w_{ikt} y_{kt} \right) \\
& - \delta \sum_{t=1}^T \sum_{i=1}^{n_t} \left(\sum_{k=1}^{n_t} w_{ikt} y_{kt} \right) \sum_{j \neq i}^{n_t} (\Delta u_{ijt} - C_{ijt} \Gamma)
\end{aligned}$$

$$\begin{aligned}
& \frac{\partial^2}{\partial \lambda^2} \log \pi(\lambda \mid Y, \Delta u, W, \Theta \mid \{\lambda\}) \\
& = - \sum_{t=1}^T \sum_{r=1}^{n_t} \frac{\mu_{rt}^2}{(1 - \lambda \mu_{rt})^2} - \sum_{t=1}^T \left[\left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n_t - 1) \right) \sum_{i=1}^{n_t} \left(\sum_{k=1}^{n_t} w_{ikt} y_{kt} \right)^2 \right].
\end{aligned}$$