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

Yajie Xu
University of South Florida

Murat Munkin
University of South Florida

October 24, 2025

[Click here for the most recent version]

Abstract

Selection biases pose a central challenge in estimating spillover effects in spatial autoregressive (SAR) models, primarily because network formation often depends on unobservable individual preferences. This paper proposes a novel Bayesian framework to jointly model both the dynamics of network effects and the formation process. Specifically, we introduce an explicit latent structure that connects outcomes and network selection equations in the SAR model through the error terms. This structure captures both endogenous formation and associated spillover effects simultaneously while enabling computationally efficient estimation. Estimation is performed via Markov Chain Monte Carlo (MCMC) methods. We validate the model through a simulation study and apply it to real-world data on R&D strategic alliance networks. We find that both the standard and the corrected model confirm the significant roles of immigrant executives in fostering collaboration and the positive spillover effects resulting from these alliances. However, the standard deviations are much higher in the standard model, indicating that ignoring network endogeneity can lead to imprecise estimates and identification challenges. This paper contributes to a general framework for estimating endogenous spillover effects and promotes the broader application of Bayesian methods in causal inference for network data.

Keywords: Bayesian, Networks, Spillover Effects, Endogeneity
JEL Codes: C11, L14, L24, J15

1 Introduction

"We Sapiens rule the world not because we are so wise but because we are the only animals that can cooperate flexibly in large numbers."
— Yuval Noah Harari, *Nexus*

Social and economic networks play a central role in shaping agents' decisions and outcomes. With the rapid expansion of information technologies, interactions among individuals and firms have become increasingly observable, allowing researchers to model interdependence explicitly. Decisions are rarely made in isolation—agents respond to incentives and behaviors of their peers, collaborators, and competitors. Recognizing this interdependence, recent econometric research has focused on identifying and estimating network effects, or *spillover effects*, that arise through these linkages. While disciplines such as physics and computer science emphasize network structure and topology, economists are primarily concerned with the human dimension—strategic decision-making and the causal impact of network structure on individual performance and social welfare.

A central issue in the empirical study of social interactions is the identification of causality. The workhorse–standard linear-in-means model–seeks to ascertain whether individual behavior is causally influenced by the average behavior of a reference group. However, as first formalized by Manski (1993), this model is subject to the "reflection problem." In its general form, the reflection problem makes it impossible to distinguish between three competing hypotheses: endogenous effects (the influence of peers' average outcomes), contextual effects (the influence of peers' average characteristics), and correlated effects (the influence of shared unobservables or sorting). This is due to the perfect collinearity that arises between the average group outcome and the average group characteristics, rendering the structural parameters inseparable.

Even under the restrictive assumption that contextual and correlated effects are absent—the "pure endogenous-effects model"—identification is still not guaranteed. Manski (1993) demonstrates that identification fails under two opposing conditions. First, if individual characteristics and group-level average characteristics are statistically independent, there is insufficient variation to identify the endogenous effect. Conversely, identification also fails if these characteristics are linearly or functionally dependent. This latter issue arises, for example, if social groups are defined by an attribute like income, which is also included as an individual characteristic. Likewise, if students in a classroom share highly similar individual traits, or if their individual characteristics are perfectly correlated with the factors that define their group membership, the endogenous effect cannot be identified.

The spatial autoregressive (SAR) model circumvents the reflection problem by leveraging the detailed structure of the network. As underscored by Bramoullé et al. (2009), identification is achieved if the network exhibits intransitivity—that is, if the friends of an individual's friends are not necessarily their own friends.1 This property, common in most real-world networks, breaks the perfect collinearity inherent in the linear-in-means model, thus resolving the reflection problem. However, the SAR specification introduces a different endogeneity concern: simultaneity. An individual's outcome influences their peers' outcomes, and simultaneously, their peers' outcomes influence the individual's outcome. In the regression, this means the spatial lag term (WY) is, by construction, correlated with the error term (ϵ), rendering standard OLS estimates biased and inconsistent.

Fortunately, the same network structure that solves the reflection problem also provides a source for valid instrumental variables (IVs). The exogenous characteristics of individuals' friend (WX) and at a network distance of two or more (e.g., friends' friends, captured by terms like W^2X) serve as natural instruments for the endogenous spatial lag. The intuition for this exclusion restriction is that the characteristics of one's friends are assumed to affect an individual's outcome only indirectly, through their influence on the outcomes of one's friends. In addition to this IV strategy, Maximum Likelihood (ML) and the Generalized Method of Moments (GMM) are also standard approaches used to obtain consistent estimates for SAR models, assuming the network itself is exogenous.

However, a more fundamental challenge is the endogeneity of the network itself. The estimation strategies discussed above, are valid only under the assumption that the network matrix, W, is exogenous. It is is often plausible in the SAR model's original context of spatial econometrics, where networks represent fixed geographic structures like the adjacency of cities or states. In most social and economic contexts, however, this exogeneity assumption is highly questionable. Individuals are not randomly assigned to peer groups; they actively form and dissolve ties through a process of self-selection. If there are unobserved factors—such as innate ability, motivation, or personal tastes—that influence both an individual's propensity to form friendships and their outcome of interest, the network matrix W will be correlated with the error term ϵ_g. This correlation renders the previously mentioned solutions inconsistent and biased.

A primary strategy for addressing network endogeneity is to adopt a joint modeling approach that explicitly specifies the network formation process. One prominent method within this framework is the two-stage instrumental variable (IV) approach. Recognizing the difficulty of finding valid external instruments for an entire network matrix,König et al. (2019), for example, propose a two-stage strategy to identify the causal effect of R&D spillovers.

In the first stage, they estimate a model of link formation (a logistic regression) to generate a predicted R&D network. This prediction is based on predetermined dyadic characteristics, such as whether two firms had a past collaboration, shared a common partner, are located in the same city, or have similar technological profiles. These variables are assumed to satisfy the exclusion restriction; that is, they influence the current probability of a link forming but are assumed not to affect the current outcome of interest directly, other than through their effect on the network structure. In the second stage, this predicted network matrix is used to construct valid instruments for the main outcome equation, purging the estimates of endogeneity bias.

However, this method is not without its challenges. The consistency of the second-stage estimates is highly dependent on the validity of the first stage. A poorly specified or weakly predictive first-stage model can lead to weak instruments, which can introduce finite-sample bias into the final results.

Other prominent methods for addressing network endogeneity include likelihood-based and control function approaches, both of which often lead to a Bayesian estimation strategy due to the models' complexity. For likelihood-based methods, the central challenge is that the full joint likelihood of the network and outcome equations is often computationally intractable for large networks. This intractability makes both standard Maximum Likelihood (ML) and full-information Bayesian estimation infeasible. To overcome this, Hsieh et al. (2025) propose a composite likelihood estimation method. This approach remains computationally feasible by maximizing an objective function composed of a product of simpler, conditional log-likelihoods (e.g., the likelihood of the outcomes given the network, and vice-versa) rather than the full, complex one. Then they implement this by using a Bayesian Markov chain Monte Carlo (MCMC) approach to estimate the parameters of the composite likelihood model. This hybrid strategy combines the computational scalability of the composite likelihood framework with the Bayesian paradigm's ability to handle unobserved latent variables through data augmentation. While this tractability comes at the cost of some statistical inefficiency compared to a full-information method, it makes an otherwise unsolvable problem estimable.

The control function approach, in the spirit of the classic Heckman (1979) selection model, addresses endogeneity by modeling its source directly. Pioneered in this context by Goldsmith-Pinkham and Imbens (2013) and Hsieh and Lee (2016), this method introduces unobserved latent variables into both the network formation and outcome models. Hsieh and Lee (2016) term their specific implementation the Selection-Corrected SAR (SC-SAR) model. The introduction of these high-dimensional latent variables is what makes the model's likelihood intractable for classical estimation methods like ML, which would require a computationally prohibitive integration step. This is precisely why they turn to Bayesian estimation. The Bayesian MCMC framework makes the model tractable by treating the latent variables as parameters to be estimated via data augmentation, thus avoiding the impossible integration problem. This framework has since been extended to dynamic settings, combining the selection-correction process with the Spatial Dynamic Panel Data (SDPD) model (Han et al., 2021).

While these latent variable models are intuitive, their implementation can be computationally demanding and relies on specific distributional assumptions about the unobservables. This paper, therefore, proposes a more direct and computationally efficient method to address this endogeneity. Following the joint modeling literature, our approach explicitly models the correlation between the network formation and outcome equations. However, instead of relying on high-dimensional latent variables, we model this dependency by incorporating the error term from the outcome equation directly into the specification for network formation.

Given the complex stochastic structure this introduces, we develop a Bayesian MCMC algorithm for estimation. We first validate the model's performance through Monte Carlo simulations. We then apply the model to a network of inter-firm research and development (R&D) collaborations and compare our estimates to those from a standard SAR model. Our results reveal a significant correlation on the dynamic of two models, highlighting the importance of our proposed correction in analyzing network data.

The remainder of the paper is organized as follows. Section 2 presents the model specification for network formation and the SAR panel data model. The Bayesian identification strategy is propsed in section 3. Section 4 presnets the simulation studies to test the model performance. Section 5 includes an empirical study on R&D collaboration networks, then section 6 concludes the paper.

2 Model

2.1 SAR Model

Consider an environment where individuals form network links and their activity outcomes are subject to social interactions (peer effects). Let n be the number of individuals.

Let $Y_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ be the $n \times 1$ outcome vector of individuals at time t. Let $X_t = (x_{1t}, x_{2t}, \ldots, x_{nt})'$ denote the $n \times k$ matrix of exogenous characteristics at time t. W_t represents the $n \times n$ adjacency matrix at time t, which evolves over time and may not be symmetric due to a 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. And to avoid a self-loop, the diagonal entries are all zeros. Mathematically,

$w_{ijt} = 0$ when $i = j$.

In the matrix form, the model can be written as:

$$Y_t = \lambda W_t Y_t + X_t \beta_1 + W_t X_t \beta_2 + \alpha + \ell \tau_t + \varepsilon_t, \quad t = 1, \ldots, T \tag{1}$$

where λ represents the contemporaneous peer effect, β_1 and β_2 are the coefficients for the direct and contextual effects of the covariates; $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)'$ is the $n \times 1$ vector of time-invariant individual effects, where each α_i is the dummy variable equals to 1 for individual i, and it is time invariant. Define also $\tau = (\tau_1, \tau_2, \ldots, \tau_t)'$ as the $T \times 1$ vector of time fixed effects, where each τ_t is the dummy variable equals to 1 at time t; ℓ is an $n \times 1$ vector of ones; and $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{nt})'$ is the vector of stochastic error terms, where $\varepsilon_{it} \overset{i.i.d.}{\sim} N(0, \sigma_\varepsilon^2)$.

For observation i $(i = 1, ..., n)$ we can write

$$y_{it} = \lambda \sum_{j=1}^{n} w_{ijt} y_{jt} + x_{it} \beta_1 + \sum_{j=1}^{n} w_{ijt} x_{jt} \beta_2 + \alpha_i + \tau_t + \varepsilon_{it}, \quad t = 1, \ldots, T \tag{2}$$

Following LeSage and Pace (2009), we can rewrite Equation (1) as

$$\begin{aligned}
Y_t - \lambda W_t Y_t &= X_t \beta_1 + W_t X_t \beta_2 + \alpha + \ell_n \tau_t + \varepsilon_t \\
(I_n - \lambda W_t) Y_t &= X_t \beta_1 + W_t X_t \beta_2 + \alpha + \ell_n \tau_t + \varepsilon_t \\
Y_t &= (I_n - \lambda W_t)^{-1} (X_t \beta_1 + W_t X_t \beta_2 + \alpha + \ell_n \tau_t + \varepsilon_t) \\
&= (I_n - \lambda W_t)^{-1} (X_t \beta_1 + W_t X_t \beta_2 + \alpha + \ell_n \tau_t) + (I_n - \lambda W_t)^{-1} \varepsilon_t \qquad (3) \\
\varepsilon_t &\sim N(0, \sigma_\varepsilon^2 I_n), \quad t = 1, \ldots, T
\end{aligned}$$

2.2 Latent Utility

However, due to self-selection, each entry of W_t might be endogenous to Y_t. We take a standard approach of modeling endogeneity through unobserved heterogeneity.

Assume that the binary choice w_{ijt} is made based on the difference in latent utilities derived from the choices of individual i regarding potential friend j defined as

$$u_{ijt}(w_{ijt} = 1) - u_{ijt}(w_{ijt} = 0) = \Delta u_{ijt} = \psi_{ijt} + \zeta_{ijt} \tag{4}$$

where

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

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, we write a regression-like equation

$$\zeta_{ijt} = \delta \varepsilon_{it} + \eta_{ijt},$$

where from equation (2)

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

and $\eta_{ijt} \sim N(0,1)$ and the equation can be interpreted as defined by the conditional distibution $\zeta_{ijt} \mid \varepsilon_{it} \sim N(\delta\varepsilon_{it}, 1)$ based on the joint distribution of ε_{it} and ζ_{ijt}. Since the variance of the conditional distibution of Δu_{ijt} given ε_{it} is fixed identification should not be an issue. Then the observability condition is

$$\begin{array}{ll} w_{ijt} = 1 & \text{if } \psi_{ijt} + \zeta_{ijt} > 0 \\ w_{ijt} = 0 & \text{if } \psi_{ijt} + \zeta_{ijt} \leqslant 0 \end{array}.$$

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

Similar to (Koop et al., 2007), we can express latent variable Δu_{ijt} as a probit model:

$$\Delta u_{ijt} = c_{it}\gamma_1 + c_{jt}\gamma_2 + c_{ijt}\gamma_3 + \delta\varepsilon_{it} + \eta_{ijt}, \quad \eta_{ijt} \overset{i.i.d.}{\sim} N(0,1) \tag{5}$$

In this formulation the time and individual fixed effects would cancel out when computing the difference Δu_{ijt}. Then

$$w_{ijt} = \begin{cases} 1 & \text{if } \Delta u_{ijt} > 0, \\ 0 & \text{if } \Delta u_{ijt} \leq 0. \end{cases}$$

2.3 Likelihood

Denoting variables $Z_t = (X_t, W_t X_t)$, $Z_{it} = \left(x_{it}, \sum_{j=1}^n w_{ijt} x_{jt}\right)$, $C_{ijt} = (c_{it}, c_{jt}, c_{ijt})$, $Q_{ijt} = (C_{ijt}, \varepsilon_{it})$, and parameters $\Gamma = (\gamma_1, \gamma_2, \gamma_3)'$, $\theta' = (\Gamma', \delta)$, $\beta = (\beta_1, \beta_2)'$, $\Theta = \{\lambda, \beta, \alpha, \tau, \sigma_\varepsilon, \theta\}$.

Based on Equation (3) and (5), the full conditional kernel can be written as

$$\begin{aligned} & P(Y, \Delta u \mid W, \Theta) \qquad (6) \\ \propto \; & \Big\{ (2\pi\sigma_\varepsilon^2)^{-\frac{Tn}{2}} \, |I - \lambda W_t|^T \\ & \times \exp\left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^{T} \sum_{i=1}^{n} [y_{it} - \lambda \sum_{j=1}^{n} w_{ijt} y_{jt} - \alpha_i - \tau_t - Z_{it}\boldsymbol{\beta}]^2 \right) \\ & \times \exp\left[-\frac{1}{2} \sum_{t=1}^{T} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \left(\Delta u_{ijt} - C_{ijt}\Gamma - \delta[y_{it} - \lambda \sum_{j=1}^{n} w_{ijt} y_{jt} - \alpha_i - \tau_t - Z_{it}\boldsymbol{\beta}] \right)^2 \right] \Big\}. \end{aligned}$$

3 Bayesian Estimation

In this section, we provide details of the MCMC algorithm for the SAR model with the latent utility model to capture the endogenous network formation. The posterior is augmented with latent variables Δu_{ijt}. The steps of the MCMC algorithm are the following:

1. As denoted above $Q_{ijt} = (c_{it}, c_{jt}, c_{ijt}, \varepsilon_{it})$, $\theta' = (\Gamma', \delta)$. Following Koop et al. (2007), conditionally on ε_{it} as well as individual- and dyad-level regressors, the latent Δu_{ijt} can be drawn from the truncated normal distribution:

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

where the notation $\text{TN}_{[a,b]}(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2, truncated to the interval $[a, b]$.

2. Given the prior distribution of $\gamma \sim N(\underline{\Gamma}, \underline{H}_\Gamma^{-1})$ and $\delta \sim N\left(\underline{\delta}, \underline{H}_\delta^{-1}\right)$ form the prior for $\theta \sim \mathcal{N}(\underline{\theta}, \underline{H}_\theta^{-1})$ 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 $\theta \sim N(\overline{\theta}, \overline{H}_{\theta}^{-1})$ where

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

3. We previously defined $Z_{it} = \left(x_{it}, \sum_{j=1}^{n} w_{ijt} x_{jt} \right)$ and $\boldsymbol{\beta}' = (\beta_1, \beta_2)$. Given the prior distribution of $\boldsymbol{\beta} \sim N(\underline{\boldsymbol{\beta}}, \underline{H}_{\boldsymbol{\beta}}^{-1})$, the full conditional distribution of $\boldsymbol{\beta}$ is:

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

where

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

4. To ensure identification of the individual fixed and time fixed effects, we impose the normalization $\alpha_1 = 0$, treating individual 1 as the baseline (Han et al. 2021; Greene 2018). This restriction resolves perfect multicollinearity arising from the inclusion of full sets of both fixed effects in the absence of an intercept Specifically, the design matrix suffers from linear dependence because the sum of individual dummies ($\sum_{i=1}^{n} \alpha_i$) and time dummies ($\sum_{t=1}^{T} \tau_t$) both equal a constant vector 1, implying $\sum_{i=1}^{n} \alpha_i - \sum_{t=1}^{T} \tau_t = 0$. This violates the full-rank assumption, rendering parameters unidentifiable.

   By setting $\alpha_1 = 0$, we eliminate this dependence: the remaining individual dummies does not add up to 1. Additionally, no corresponding omission of time dummies is required. Because the model lacks an intercept, the time dummies retain the necessary role of capturing the overall level of the dependent variable across periods. All T time effects $(\tau_1, \ldots, \tau_T)$ remain freely estimated, measuring absolute temporal shifts relative to the baseline individual.

Under the prior $\alpha_i \sim N\left(\underline{\alpha}_i\,, \underline{H}_{\alpha_i}^{-1}\right)$, the full conditional distribution of α_i is

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

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 for 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 prior $\lambda \sim \text{Uniform}(-1, 1)$, we sample λ from the full conditional distribution:

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

where the likelihood is given by Equation (6):

To sample from the full conditional of λ, we designed a Metropolis-Hastings (MH) step with a proposal density following LeSage (1997); LeSage and Pace (2009); Hsieh and Lee (2016); Han et al. (2021).

S1.1: Propose $\tilde{\lambda} \sim N(\lambda^{(q-1)}, c_{\lambda})$, where c_{λ} is adjusted during iterations to achieve an acceptance rate between 40% and 60%. Check whether $\tilde{\lambda}$ satisfies the stability condition implied by the prior. If not, redraw $\tilde{\lambda}$ until it satisfies the condition.

S1.2: With acceptance probability

$$\Pr\left(\lambda^{(q-1)}, \tilde{\lambda}\right) = \min\left\{1, \prod_{t=1}^{T}\prod_{i=1}^{n} \frac{p\left(y_{it} \mid W_t, \tilde{\lambda}, \beta, \alpha_i, \tau_t, \sigma_{\varepsilon}^{-2}\right)}{p\left(y_{it} \mid W_t, \lambda^{(q-1)}, \beta, \alpha_i, \tau_t, \sigma_{\varepsilon}^{-2}\right)} \times \frac{\pi\left(\tilde{\lambda}\right)}{\pi\left(\lambda^{(q-1)}\right)}\right\}$$

update $\lambda^{(q)} = \tilde{\lambda}$; otherwise, set $\lambda^{(q)} = \lambda^{(q-1)}$.

4 Simulation Study

This simulation study jointly estimates the latent utility and SAR models. The data-generating process (DGP) follows the SAR specification in equation (1) and the latent utility model in equation (4). The simulated panel consists of 100 firms across 5 sectors observed over 20 years, controlling for both time and sector fixed effects. This setup is designed to approximate the structure of the empirical application discussed later.

Table 1 reports the posterior means, standard deviations, and credible intervals from the Bayesian estimation alongside the true parameter values used in the DGP. Overall, the model reproduces the true parameters with reasonable accuracy. The posterior means are close to the true values, and the 95% credible intervals typically include the true parameters. The mean-to-standard-deviation ratios suggest that most coefficients

Table 1: Monte Carlo experiment results

Parameter	**True**	**Mean**	**SD**	**Ratio**	**ACF20**	**CI95**
λ	0.300	0.279	0.018	15.507	0.089	0.247, 0.316
β_0	0.500	0.532	0.039	13.467	0.008	0.453, 0.609
β_1	1.000	0.991	0.007	147.120	0.053	0.978, 1.005
β_2	1.000	1.012	0.007	154.350	0.101	1.000, 1.026
β_3	1.000	1.000	0.007	150.490	−0.009	0.987, 1.014
β_4	1.000	1.006	0.007	153.140	0.024	0.993, 1.019
σ^2	1.000	1.053	0.033	31.477	0.029	0.992, 1.124
δ	−0.500	−0.505	0.004	−139.530	−0.012	−0.512, −0.498
γ_0	0.700	0.698	0.004	176.360	0.031	0.690, 0.706
γ_1	0.100	0.100	0.003	29.536	−0.033	0.093, 0.106
γ_2	0.500	0.502	0.004	135.480	−0.033	0.495, 0.510
γ_3	0.600	0.599	0.004	143.200	−0.031	0.591, 0.607
γ_4	0.200	0.192	0.003	56.188	−0.019	0.186, 0.200
γ_5	0.300	0.296	0.004	81.945	0.000	0.289, 0.303
γ_6	0.200	0.203	0.004	57.138	0.020	0.196, 0.210

Notes: MCMC diagnostics for the simulated SAR–MCMC model based on 10,000 total draws with a 2,000 burn-in period. Posterior samples are thinned by retaining every 10th iteration.

are estimated with relatively high precision. As shown in figure 1, the trace plots indicate stable chains, and after thinning, the autocorrelation functions decay rapidly, suggesting satisfactory convergence.

The estimated spillover parameter is $\hat{\lambda} = 0.279$ with a 95% credible interval of $[0.247, 0.316]$, which is close to the true value of 0.3. Similarly, the estimated endogeneity parameter $\hat{\delta} = -0.505$ (95% CI $[-0.512, -0.498]$) aligns closely with the true value of -0.5. These results indicate that the model adequately captures the dependence between the latent utility and the SAR disturbances.

Lag-20 autocorrelations are below 0.1 for all parameters, implying satisfactory chain independence under a thinning interval of 10. Slightly higher persistence is observed for λ and β_2, but both remain within acceptable limits. Overall, the simulation provides evidence that the proposed model yields reliable estimates for both spatial and structural parameters in moderately sized panel data.

5 Empirical Application

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).[1] Following Hsieh et al. (2025) and König et al. (2019), all financial variables are deflated by the Consumer Price Index (CPI) of the firm's home country, expressed in millions of U.S. dollars, and transformed into natural logarithms. *Output* is measured as the logarithm of annual revenue, *R&D effort* as the logarithm of R&D expenditures, and *Productivity* as the one-year lag of R&D capital stock, computed using a 15% annual depreciation rate. Table 2 presents the full list of variables and descriptive statistics.

Information about senior executive officers is compiled from multiple sources. The primary source is firms' annual reports filed with the Securities and Exchange Commission (SEC), which contain their 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

[1] According to Bloom et al. (2013), R&D activities are predominantly concentrated among publicly traded firms, suggesting that this sample captures the majority of innovation efforts.

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

Variable	**Mean**	**SD**	**Median**	**Min**	**Max**
Output	4.181	3.668	4.085	-7.724	11.231
LongTermDebt	2.678	3.636	1.559	-7.612	10.690
CostOfRevenue	3.072	3.526	2.458	-8.510	10.233
R&D	3.316	2.685	3.199	-8.023	16.913
Productivity	4.440	2.999	4.548	-5.521	16.913
Observations					3,121

Notes: All financial variables are deflated by the Consumer Price Index (CPI) of firm's home country, expressed in millions of U.S. dollars, and transformed into natural logarithms. *Output* is measured as the logarithm of firms' annual revenue, *R&D effort* as the logarithm of R&D expenditures, and *Productivity* as the one-year lag of R&D capital stock, computed using a 15% annual depreciation rate.

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 headquartered 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 De 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; Mahroum 2000; Mahroum 2001; Prato 2024).

The dyadic variables, reported in Table 3, capture both relational and demographic characteristics between firm pairs. The variable *ImmigrantExecutive* equals one if at least one senior executive in either firm is an immigrant. *SameCountryOfBirth* equals one if at least one pair of executives from the two firms was born in the same country. Similarly, *OldFriends* equals one if the two firms had an R&D alliance prior to the

Table 3: Summary statistics for dyadic/network variables

Variable	**Mean**	**SD**	**Median**	**Min**	**Max**
ImmigrantExecutive	0.470	0.499	0.000	0.000	1.000
SameCountryOfBirth	0.290	0.454	0.000	0.000	1.000
OldFriends	0.014	0.136	0.000	0.000	4.000
CommonFriends	0.087	0.281	0.000	0.000	1.000
SameCity	0.007	0.084	0.000	0.000	1.000
AveOutput	4.172	2.602	4.132	-6.928	11.194
AveDebt	2.663	2.604	2.529	-6.000	10.614
AveCost	3.062	2.504	3.043	-7.479	10.176
AveR&D	3.307	1.915	3.215	-6.754	13.109
AveProductivity	4.418	2.237	4.486	-4.345	13.797
DifOutput	4.182	3.029	3.633	0.000	18.679
DifDebt	3.937	3.182	3.443	0.000	18.244
DifCost	3.929	3.030	3.525	0.000	18.565
DifR&D	2.972	2.300	2.535	0.000	24.785
DifProductivity	3.128	2.493	2.620	0.000	20.604
NetworkDensity	0.005	0.071	0.000	0.000	1.000
Observations					254,910

Notes: Binary indicators (*ImmigrantExecutive*, *SameCountryOfBirth*, *OldFriends*, *CommonFriends*, and *SameCity*) take the value one when the corresponding relational or spatial condition holds between two firms and zero otherwise. Variables prefixed by *Ave* represent the average of firm-level financial and innovation characteristics within each dyad, while those prefixed by *Dif* denote the absolute differences between the two firms, capturing heterogeneity in size, resources, and technological capability. All continuous variables are expressed in natural logarithms and deflated via the Consumer Price Index (CPI) of the firm's home country, and converted to U.S. dollars. *NetworkDensity* measures the share of realized links among all possible firm pairs in the observed R&D collaboration network. Summary statistics are computed for the unbalanced panel of 254,910 firm–pair observations covering 2004–2022.

current period, and *CommonFriends* equals one if they shared at least one collaborator before forming the current alliance. *SameCity* equals one if the two firms are located in the same city. Together, these binary indicators characterize the social and spatial proximity that may influence the formation of interfirm R&D collaborations.

In addition to these binary measures, the dataset includes continuous variables that describe firm-level characteristics averaged or differenced across each dyad. Variables beginning with *Ave* (e.g., *AveOutput*, *AveDebt*, *AveCost*, *AveR&D*, and *AveProductivity*) denote the average levels of the corresponding firm attributes between the two firms in each pair, capturing the overall scale or intensity of their financial and innovation activities. Variables beginning with *Dif* (e.g., *DifOutput*, *DifDebt*, *DifCost*, *DifR&D*, and *DifProductivity*) measure the absolute differences in these attributes between the two firms, reflecting heterogeneity in size, resource endowment, and technological capability. Finally, *NetworkDensity* represents the proportion of realized connections among all potential firm pairs within the observed network, serving as a measure of the overall connectivity of the R&D collaboration network. All continuous variables are transformed using logarithmic transformations to reduce skewness and facilitate comparison across variables.

5.2 Estimation Results

The Bayesian estimation is implemented using Markov chain Monte Carlo (MCMC) sampling with 100,000 iterations, discarding the initial 15,000 iterations as burn-in and retaining every tenth draw to reduce serial correlation. Posterior means and standard deviations computed from the retained draws are reported as point estimates. Convergence is verified following standard diagnostics (Geweke, 1992; Raftery and Lewis, 1992; Heidelberger and Welch, 1983). Table 4 and 6 reports the results without contextual effects, while Table 5 and 7 incorporates spatially lagged (contextual) covariates.

Table 4: SAR (no contextual effects): OLS vs. Bayesian

	OLS		**Bayesian**				
Variable	**Estimate**	**SD**	**Mean**	**SD**	**Ratio**	**ACF20**	**CI95**
λ	0.046	0.012	0.050	0.007	7.692	0.013	0.037, 0.064
Constant	0.967	0.323	-0.228	0.988	-0.231	0.496	-2.144, 1.774
LongTermDebt	0.082	0.012	0.113	0.007	15.957	0.004	0.099, 0.127
CostOfRevenue	0.794	0.011	0.514	0.007	75.919	-0.010	0.501, 0.527
R&D	0.246	0.019	0.144	0.010	14.284	0.010	0.125, 0.164
Productivity	-0.038	0.016	0.016	0.013	1.275	0.025	-0.009, 0.041
σ^2	1.888	–	0.607	0.027	22.677	0.123	0.563, 0.668
Observations							3,121

Notes: Dependent variable: Output. Fixed effects: sector and year. Standard errors for MLE are IID. σ^2 is the square of the empirical residual. RMSE (OLS): 1.368; Adjusted R^2: 0.860; Within R^2: 0.858. Bayesian columns report posterior means (Mean), posterior standard deviations (SD), mean-to-SD ratios (Ratio), lag-20 autocorrelation (ACF20), and 95% credible intervals (CI95), shown as lower and upper bounds separated by a comma.

The coefficient of primary interest is the endogenous spillover parameter λ, representing the strength of peer or network interactions in output among firms. Under the normalized network specification, λ measures the average percentage change in a firm's output in response to a one-percent change in the average output of its connected peers (LeSage and Pace, 2009). In both the baseline and contextual specifications, the

Table 5: SAR with Contextual Effects: OLS vs. Bayesian

	OLS		Bayesian				
Variable	**Estimate**	**SD**	**Mean**	**SD**	**Ratio**	**ACF20**	**CI95**
λ	0.055	0.122	0.054	0.011	5.077	-0.000	0.033, 0.075
Constant	0.973	0.324	-0.055	1.066	-0.052	0.535	-2.064, 2.174
LongTermDebt	0.082	0.012	0.113	0.007	15.686	-0.018	0.099, 0.127
CostOfRevenue	0.794	0.011	0.514	0.007	74.996	0.023	0.501, 0.528
R&D	0.246	0.019	0.145	0.010	14.045	-0.017	0.125, 0.165
Productivity	-0.038	0.016	0.016	0.013	1.262	0.047	-0.009, 0.041
LongTermDebt (Contextual)	-0.009	0.072	-0.078	0.020	-3.942	-0.023	-0.117, -0.039
CostOfRevenue (Contextual)	-0.038	0.124	0.048	0.032	1.474	0.019	-0.018, 0.110
R&D (Contextual)	0.138	0.112	-0.021	0.038	-0.556	0.030	-0.093, 0.055
Productivity (Contextual)	-0.086	0.084	0.030	0.044	0.684	0.032	-0.057, 0.116
σ^2	1.890		0.612	0.028	21.645	0.130	0.567, 0.679
Observations							3,121

Notes: Dependent variable: Output. Fixed effects: sector and year. Standard errors for MLE are IID. OLS SDs are from the regression output. Bayesian columns report posterior means (Mean), posterior standard deviations (SD), mean-to-SD ratios (Ratio), lag-20 autocorrelation of the MCMC chain (ACF20), and 95% credible intervals (CI95), shown as lower and upper bounds separated by commas. Contextual variables (e.g., LongTermDebt (Contextual)) correspond to the average of cooperators' covariates.

posterior mean of λ is approximately 0.05 with no zero included in the 95% credible intervals (0.037, 0.064) and (0.033, 0.075), indicating statistically significant endogenous spillovers. Substantively, these estimates suggest that a 1% increase in the average output of a firm's collaborators is associated with roughly a 0.05% increase in the firm's own output, holding other factors constant.

The implied social multiplier, defined as $(1-\lambda)^{-1}$ under a normalized network, is approximately 1.053. This indicates that, on average, a one-unit exogenous increase in output at the firm level ultimately generates a 5.3% larger equilibrium response in the aggregate network through feedback effects among connected firms. Such amplification highlights the economic significance of interfirm interdependencies in R&D collaboration networks: local improvements in productivity propagate beyond the directly affected firms, reinforcing aggregate output through repeated spillovers.

The Bayesian results are overally consistent with the OLS estimates, though the Bayesian posteriors yield slightly larger and more precise estimates for the spillover effect. This finding reflects the Bayesian model's ability to account for uncertainty in the joint distribution of the parameters and to mitigate small-sample bias through hierarchical shrinkage. When contextual effects are included, the OLS estimate of λ becomes statistically insignificant, likely because the contextual covariates absorb some of the between-firm variation. However, the Bayesian estimate of λ remains credibly different from zero, underscoring the robustness of the spillover mechanism even after controlling for network-level covariates.

For the own-firm characteristics, the coefficients on *LongTermDebt*, *CostOfRevenue*, and *R&D* remain positive and significant across specifications, indicating that capital structure and R&D intensity are key drivers of output growth. In contrast, the coefficient on *Productivity* is small and statistically weak, suggesting diminishing marginal effects once interfirm dependencies are accounted for. In terms of contextual coefficients, only *LongTermDebt (Contextual)* is statistically significant, with a posterior mean of -0.078 and a 95% credible interval of $[-0.117, -0.039]$. This negative sign suggests that higher long-term debt levels among a firm's R&D partners are associated with a reduction in the focal firm's output, holding other factors constant. One interpretation is that when partner firms face higher leverage or financial constraints, their capacity to sustain cooperative innovation and share resources diminishes, which in turn negatively affects the joint productivity of the network. This finding underscores the importance of partners' financial stability in determining the overall effectiveness of R&D collaborations.

Following LeSage and Pace (2009) and Hsieh and Lee (2016), the estimated coefficients can be further decomposed into *direct*, *indirect*, and *total* effects. The direct effects reflect the marginal impact of a firm's own covariate on its output, whereas the indirect effects (or network spillovers) measure how that covariate propagates through connected peers via $(I - \lambda W)^{-1}$. The total effect is the sum of both, capturing the overall equilibrium response of the network to an exogenous change in a given covariate. For instance, the posterior mean of the own-firm long-term debt coefficient implies that a 1% increase in a firm's own long-term debt raises its output directly by about 0.11%, while the indirect (network-mediated) effect through connected collaborators reduces output by approximately 0.01%, yielding a total marginal effect of roughly 0.10%. These magnitudes indicate that although own-firm characteristics remain the primary determinants of performance, interfirm financial linkages exert a measurable, and sometimes offsetting, influence on overall productivity through the collaboration network. Taken together, the positive direct effect and the negative contextual spillover effect of long-term debt reveal a nuanced dynamic in which financially leveraged firms may benefit individually but transmit adverse externalities to their partners.

For instance, the posterior mean of the long-term debt coefficient implies that a 1% increase in a firm's own long-term debt raises its output directly by about 0.11%, while the indirect (network-mediated) effect through connected collaborators minus approximately 0.01%, yielding a total marginal effect of roughly 0.1%. These magnitudes underscore that although own-firm characteristics dominate in explaining performance, interfirm linkages play a meaningful—albeit secondary—role in adjusting the effects through the collaboration network.

Table 6: Standard and Bayesian Probit Estimates (without contextual effects)

	MLE		Bayesian				
Variable	**Estimate**	**SD**	**Mean**	**SD**	**Ratio**	**ACF20**	**CI95**
δ (Endogeneity)	–	–	-0.168	0.023	-7.224	0.065	−0.212, −0.120
Constant	-5.047	0.123	-5.351	0.117	-45.791	0.680	−5.573, −5.119
ImmigrantExecutive	0.212	0.034	0.213	0.033	6.391	0.003	0.147, 0.280
SameCountryOfBirth	-0.002	0.037	-0.013	0.037	-0.362	-0.001	−0.086, 0.057
OldFriends	1.693	0.032	1.703	0.029	59.367	-0.010	1.648, 1.760
CommonFriends	-0.123	0.040	-0.109	0.038	-2.850	0.024	−0.182, −0.034
SameCity	-0.158	0.180	-0.185	0.168	-1.103	-0.004	−0.522, 0.135
AveOutput	0.489	0.035	0.629	0.038	16.572	0.491	0.549, 0.701
AveDebt	-0.051	0.011	-0.071	0.012	-6.049	0.053	−0.093, −0.047
AveCost	-0.194	0.025	-0.262	0.026	-10.097	0.297	−0.312, −0.209
AveR&D	0.096	0.017	0.066	0.016	4.057	0.091	0.035, 0.099
AveProductivity	-0.118	0.012	-0.123	0.013	-9.464	0.102	−0.147, −0.097
DifOutput	0.145	0.018	0.182	0.017	10.749	0.391	0.147, 0.214
DifDebt	-0.004	0.006	-0.009	0.006	-1.444	0.004	−0.021, 0.004
DifCost	-0.039	0.013	-0.052	0.012	-4.227	0.206	−0.076, −0.028
DifR&D	0.024	0.011	0.013	0.011	1.252	0.068	−0.008, 0.034
DifProductivity	-0.045	0.009	-0.044	0.009	-4.718	0.089	−0.063, −0.026
Observations							254,910

Notes: MLE (OLS) columns report probit maximum-likelihood estimates and robust standard errors. Bayesian columns report posterior means, posterior standard deviations (SD), mean-to-SD ratios (Ratio), lag-20 autocorrelation of the MCMC chain (ACF20), and 95% equal-tailed credible intervals (CI95), presented as lower and upper bounds separated by commas. This specification excludes contextual effects in the SAR model.

In the estimation of the latent-utility (probit) model, the results from the Bayesian and standard MLE approaches are largely consistent in both magnitude and statistical significance. The parameter of primary interest is the endogeneity term δ, which is negative and statistically significant even under the 95% credible interval in both specifications. This finding 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 lowering the probability of alliance formation tend to be associated with higher unobserved productivity shocks in the SAR equation, implying a negative correlation between the residuals of the two latent processes.

The large and negative constant term (posterior mean ≈ -5.35) is consistent with

Table 7: Standard and Bayesian Probit Estimates (with contextual effects)

	MLE		Bayesian				
Variable	**Estimate**	**SD**	**Mean**	**SD**	**Ratio**	**ACF20**	**CI95**
δ (Endogeneity)	–	–	-0.166	0.024	-6.918	0.169	−0.212, −0.118
Constant	-5.047	0.123	-5.340	0.144	-37.210	0.795	−5.615, −5.071
ImmigrantExecutive	0.212	0.034	0.212	0.034	6.279	0.021	0.146, 0.279
SameCountryOfBirth	-0.002	0.037	-0.014	0.037	-0.377	0.045	−0.086, 0.059
OldFriends	1.693	0.032	1.704	0.029	58.768	-0.018	1.648, 1.761
CommonFriends	-0.123	0.040	-0.112	0.039	-2.832	0.002	−0.189, −0.036
SameCity	-0.158	0.180	-0.188	0.171	-1.101	0.020	−0.534, 0.130
AveOutput	0.489	0.035	0.625	0.044	14.131	0.634	0.538, 0.714
AveDebt	-0.051	0.011	-0.070	0.012	-5.816	0.087	−0.093, −0.046
AveCost	-0.194	0.025	-0.261	0.028	-9.472	0.342	−0.315, −0.207
AveR&D	0.096	0.017	0.065	0.018	3.712	0.159	0.031, 0.100
AveProductivity	-0.118	0.012	-0.121	0.013	-9.585	0.049	−0.146, −0.096
DifOutput	0.145	0.018	0.181	0.019	9.388	0.503	0.142, 0.219
DifDebt	-0.004	0.006	-0.009	0.006	-1.466	0.068	−0.022, 0.003
DifCost	-0.039	0.013	-0.052	0.013	-4.016	0.260	−0.077, −0.027
DifR&D	0.024	0.011	0.015	0.011	1.369	0.064	−0.007, 0.035
DifProductivity	-0.045	0.009	-0.045	0.009	-4.746	0.072	−0.063, −0.026
Observations							254,910

Notes: MLE (OLS) columns report probit maximum-likelihood estimates and robust standard errors. Bayesian columns report posterior means, posterior standard deviations (SD), mean-to-SD ratios (Ratio), lag-20 autocorrelation of the MCMC chain (ACF20), and 95% equal-tailed credible intervals (CI95), presented as lower and upper bounds separated by commas. The SAR model includes contextual effects.

the sparsity of the observed network, reflecting the empirical difficulty of establishing successful R&D strategic alliances among firms. Among firm-pair characteristics, the positive and statistically significant coefficient on *ImmigrantExecutive* highlights the role of immigrant executives in facilitating cross-firm cooperation, potentially through broader international networks and cultural openness. By contrast, the coefficient on *SameCountryOfBirth* is negative but statistically insignificant, suggesting that shared nationality alone does not drive alliance formation—a result that may reflect an open and cosmopolitan orientation among collaborating firms. The positive and highly significant coefficient on *OldFriends* underscores the importance of relational trust and accumulated experience in sustaining interfirm collaboration. In contrast, the negative coefficient on *CommonFriends* suggests that pharmaceutical firms may not rely heavily on indirect acquaintances or third-party recommendations when seeking new R&D partners, possibly reflecting the sector's emphasis on confidentiality and specialized expertise. Meanwhile, the negative but insignificant coefficient on *SameCity* indicates that geographic proximity plays a limited role once social and strategic linkages are controlled for.

The coefficients on the average variables (*AveDebt*, *AveCost*, *AveOutput*, and *AveR&D*) highlight the importance of joint financial soundness and research capacity in facilitating alliance formation—firms prefer partners that are both financially stable and technologically capable. Finally, the results for the difference variables (*DifOutput*, *DifR&D*, etc.) provide evidence consistent with homophily in partnership formation: firms tend to collaborate with others that are similar in performance and innovation intensity. The positive and significant effects of *DifOutput* and *DifR&D* indicate a pattern of strategic complementarity—suggesting that firms also seek partners whose strengths can compensate for their own weaknesses, reflecting a balance between similarity and complementarity in R&D cooperation.

6 Conclusion

This paper develops and estimates a structural model that extends the Spatial Autoregressive (SAR) framework to account for endogenous network formation. When the spatial weight matrix is endogenously determined—meaning it is correlated with the disturbances of the SAR model—standard estimators become inconsistent, leading to endogeneity and selection bias in the outcome equation.

While the literature has proposed powerful solutions, such as control function approaches using high-dimensional latent variables or composite likelihood methods , these can be computationally demanding and rely on specific assumptions about the unobserved drivers of network formation. This paper, therefore, proposes a more direct and computationally tractable joint modeling approach. Instead of introducing latent variables, we explicitly model the endogeneity by incorporating a function of the error term from the outcome equation directly into the network formation specification.

We develop a Bayesian MCMC algorithm for estimation and examine the finite-sample performance of our estimator through Monte Carlo experiments before applying it to a network of inter-firm R&D collaborations. Our empirical results confirm the presence of significant endogeneity. In contrast to a standard SAR model, our proposed estimator robustly identifies a strong, positive spillover effect, even in specifications that include contextual effects, demonstrating its utility in overcoming the identification challenges posed by endogenous network formation.

7 Acknowledgment

The authors are grateful for the advice received from Andrei Barbos, Vincent Boucher, Bobby W. Chung, William Greene, and Michael D. König, as well as for the support from the Department of Economics. Appreciation is also given for the data collection efforts of Rezuanul Huq Rafi, Mohaimin Chowdhury, Safayet Hossain Tushar, Albab Amin Beky, Fatema Sobnam Moumita, Md Ifteker Chowdhury, MD. Borhan Uddin, Md. Nafiz Imtiaz, MD. Sadik Rahman, Rahul Barua, and Shahida Aktar.

References

Arp, F., Hutchings, K., and A. Smith, W. (2013). Foreign executives in local organisations: An exploration of differences to other types of expatriates. *Journal of Global Mobility*, 1(3):312–335.

Bloom, N., Schankerman, M., and Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4):1347–1393.

Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of econometrics*, 150(1):41–55.

Flahaux, M.-L. and De Haas, H. (2016). African migration: trends, patterns, drivers. *Comparative migration studies*, 4:1–25.

Goldsmith-Pinkham, P. and Imbens, G. W. (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics*, 31(3):253–264.

Greene, W. H. (2018). *Econometric Analysis*. Pearson, 8 edition.

Han, X., Hsieh, C.-S., and and, S. I. M. K. (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.

Hsieh, C.-S., König, M. D., and Liu, X. (2025). Endogenous technology spillovers in r&d collaboration networks. *The RAND Journal of Economics*, n/a(n/a).

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.

König, M. D., Liu, X., and Zenou, Y. (2019). R&d networks: Theory, empirics, and policy implications. *Review of Economics and Statistics*, 101(3):476–491.

Koop, G., Poirier, D. J., and Tobias, J. L. (2007). *Bayesian Econometric Methods*. Econometric Exercises. Cambridge University Press.

LeSage, J. and Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

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. (2001). Europe and the immigration of highly skilled labour. *International Migration*, 39(5):27–43.

Mahroum, S. and Ansari, R. (2017). What the data tells us about immigrant executives in the us. *Harvard Business Review Digital Article, November*, 29.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542.

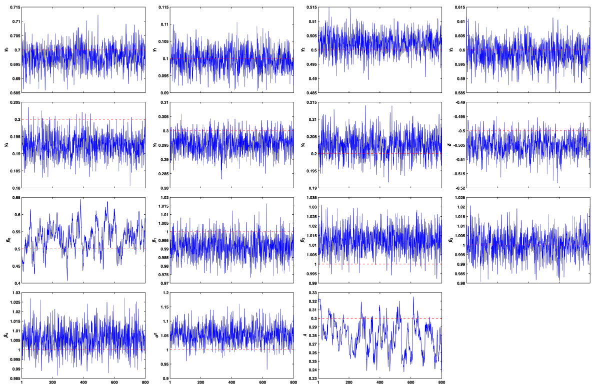
Platonova, A. and Urso, G. (2013). Asian immigration to the european union, united states and canada: an initial comparison. *Journal of Global Policy and Governance*, 1(2):143–156.

Prato, M. (2024). The global race for talent: Brain drain, knowledge transfer, and growth. *The Quarterly Journal of Economics*, 140(1):165–238.

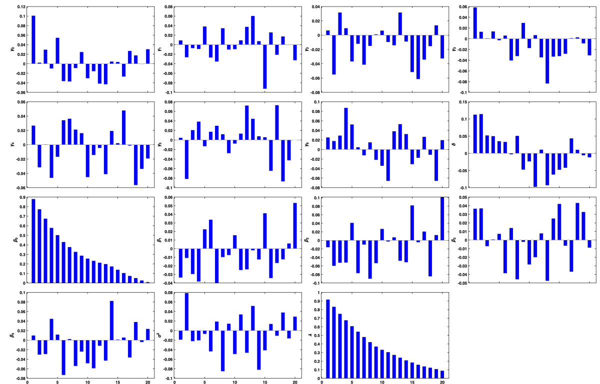
Schilling, M. A. (2009). Understanding the alliance data. *Strategic Management Journal*, 30(3):233–260.

8 Appendix

8.1 Figures

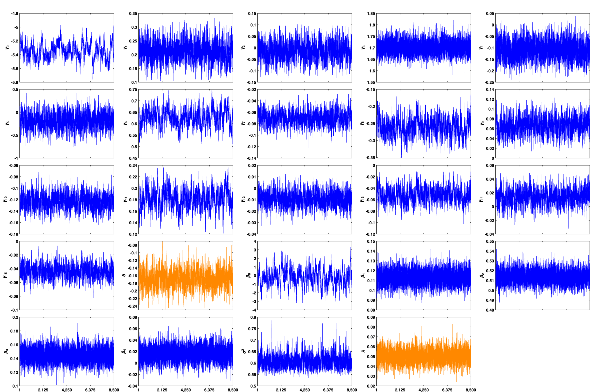


(a) MCMC Trace Plots

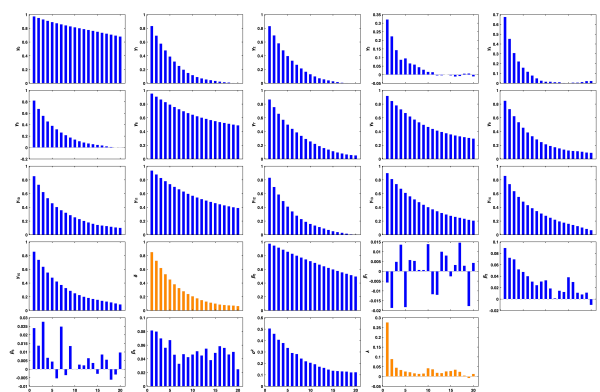


(b) Autocorrelation Plots

Figure 1: MCMC diagnostics based on 10,000 draws with a 2,000 burn-in and thinning interval of 10. Both plots use the same posterior samples.

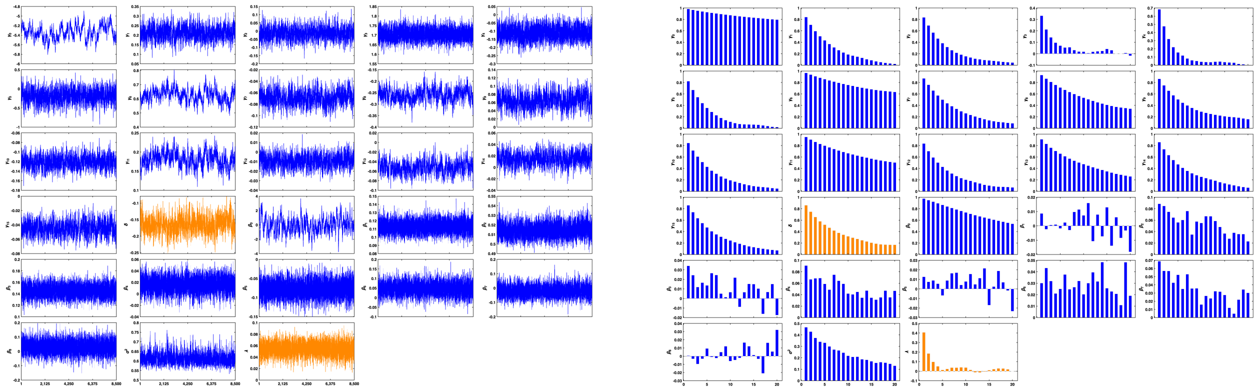


(a) MCMC Trace Plots



(b) Autocorrelation Plots

Figure 2: MCMC diagnostics for the model without contextual effects, based on 100,000 total draws with a 15,000 burn-in period. Both panels correspond to the same thinned posterior samples, retaining every 10th iteration.



(a) MCMC Trace Plots (b) Autocorrelation Plots

Figure 3: MCMC diagnostics for the model including contextual effects, based on 100,000 total draws with a 15,000 burn-in period. Both panels correspond to the same thinned posterior samples, retaining every 10th iteration.

8.2 Posterior Distribution for $\boldsymbol{\theta}$

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:**

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

- **Link likelihood Δu_{ijt}:**

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

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

  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 $\boldsymbol{\theta}$

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

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

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

8.3 Posterior Distribution of $\boldsymbol{\beta}$

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

- **Prior:** $\boldsymbol{\beta} \sim \mathcal{N}(\underline{\boldsymbol{\beta}}, \underline{H}_{\boldsymbol{\beta}}^{-1})$
- **Outcome likelihood y_{it}:** for every individual $i = 1, \ldots, n$ and every period $t = 1, \ldots, T$, because $\boldsymbol{\beta}$ enters the mean through the regressor block $Z_{it}\boldsymbol{\beta}$.
- **Dyadic likelihood Δu_{ijt}:** for every senderâ€“receiver pair (i, j) with $j \neq i$ across the same T periods, since the senderâ€s error term $\varepsilon_{it} = r_{it} - Z_{it}\boldsymbol{\beta}$ appears in the latent utility equation whenever i is the sender.

1. Log-Prior for $\boldsymbol{\beta}$

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

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

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

$$\varepsilon_{it} = r_{it} - Z_{it}\boldsymbol{\beta}$$

$$\log p(\mathbf{y} \mid \cdot) = -\frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^{n} \sum_{t=1}^{T} (r_{it} - Z_{it}\boldsymbol{\beta})^2 + \text{const}$$

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

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

$$\Delta u_{ijt} = s_{ijt} + \delta \varepsilon_{it} + \eta_{ijt}, \quad \varepsilon_{it} = r_{it} - Z_{it}\boldsymbol{\beta}$$

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

4. Combined Log-Posterior

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

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}:

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

8.4 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 $t = 1, \ldots, T$, 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 $\varepsilon_{it} = y_{it} - \lambda \sum_{j=1}^{n} w_{ijt} y_{jt} - \cdots - \alpha_i$ 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:

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

4. Combined Log-Posterior:

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

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:

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

where:

$$\overline{H}_{\alpha_i} = H_{\alpha_i} + T\left(\frac{1}{\sigma_\varepsilon^2} + \delta^2(n-1)\right),$$

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

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

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

8.5 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 $i = 1, \ldots, n$ 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 senderâ€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 $\varepsilon_{it} = r_{it} - \tau_t$:

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

4. **Combined Log-Posterior:**

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

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:

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

where:

$$\overline{H}_{\tau_t} = H_{\tau_t} + \frac{n}{\sigma_\varepsilon^2} + \delta^2 n(n-1),$$

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

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

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

8.6 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:

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

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:

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

8.7 Unbalanced Panel

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

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

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):

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

The joint kernel of the observed data

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

Substituting ε_{it}:

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

Posterior of $\boldsymbol{\beta}$:

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

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

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

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

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

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

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

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

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

Full conditional density of λ

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

This can be reduced to

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

Log of density:

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

Quadratic expansion in λ

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

Since

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

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

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