Yajie Xu
Ph.D. Candidate, Economics
M.S. Student, Artificial Intelligence
University of South Florida
Tampa, FL 33620

To
Emily Breza, Davide Viviano, and Arun Chandrasekhar
Workshop on Development Economics and Networks for Early-Career Researchers
National Bureau of Economic Research
1050 Massachusetts Avenue
Cambridge, MA 02138

September 15, 2026

Dear Professors Breza, Viviano, and Chandrasekhar,

I am writing to apply to the Workshop on Development Economics and Networks for Early-Career Researchers on December 5, 2026. I am a Ph.D. candidate in Economics at the University of South Florida, concurrently completing an M.S. in Artificial Intelligence, and my research develops methods for causal inference in endogenously formed networks.

My working paper, "Bayesian Estimation of Endogenous Network Effects in SAR Models: Application to R&D Collaboration," develops a Bayesian MCMC framework to identify spillover effects in R&D collaboration networks while correcting for network endogeneity — addressing the reflection problem more robustly than standard instrumental-variable approaches. A companion paper, "Super-Connectors: Immigrant Executives and the Architecture of R&D Networks," builds a structural model of network formation grounded in Cournot competition to study how immigrant executives' human capital and social preferences shape the topology of these networks. Together, this agenda combines structural network formation theory with Bayesian estimation strategies to move beyond reduced-form treatments of network endogeneity.

I am applying to this workshop because I am at a stage where engaging directly with researchers at the frontier of empirical network methods would meaningfully sharpen my identification strategy and open new directions for extending it — particularly toward development applications, where endogenous network formation is central to understanding diffusion, cooperation, and market outcomes. I would also welcome the chance to learn from other early-career researchers working on related problems.

A recommendation, if needed, can be provided by Professor Murat Munkin (mmunkin@usf.edu).

Thank you for considering my application. I look forward to the possibility of joining this workshop.

Sincerely,

Yajie Xu

Yajie Xu

Yajie Xu

Updated September 2026

Email: yajie@usf.edu Office: CMC 205E, University of South Florida Web: yajiexu.com

Research interests

Econometrics of networks, Bayesian methods, causal inference
Innovation and R&D collaboration, behavioral economics, human–AI interaction

Education

University of South Florida August 2021 – May 2027 (Expected)
PhD in Economics
Committee: Andrei Barbos (Co-chair), Murat Munkin (Co-chair), William Greene, Wing Yin Chung, Michael König, and Vincent Boucher
Admitted to candidacy July 2024. GPA 4.0/4.0.

MS in Artificial Intelligence (concurrent) 2027 (Expected)

Hubei University September 2016 – June 2020
BS in Economics

PhD coursework

Completed: Microeconomics I–II, Macroeconomics I–II, Econometrics I–IV, Mathematical Economics I–II, Bayesian Data Analysis, Social Network Analysis, Behavioral Economics, Labor Economics I–II, Industrial Organization I, Economic Policy Analysis.

Current (Fall 2026): Dissertation research; Natural Language Processing.

Graduate AI coursework (MS): Introduction to AI, Ethical Issues in AI, Math for AI, Data Structures, Algorithms, Discrete Structures, Object-Oriented Programming, Computing Essentials.

Working papers

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

Develops a Bayesian MCMC framework that identifies endogenous spillovers under network endogeneity, addressing the reflection problem more robustly than IV approaches and with higher efficiency than comparable Bayesian structural models. Preparing for journal submission.

"Super-Connectors: Immigrant Executives and the Architecture of R&D Networks" with Kohei Takakura

Structural model of R&D link formation grounded in Cournot competition and network formation games; finds that immigrant executives densify collaboration networks and accelerate knowledge diffusion. Built on a novel hand-collected dataset linking executive backgrounds to inter-firm alliances.

"CRetor: A Mandarin Counter-Speech Dataset Annotated for Rhetorical Strategy" with Michael Bennie et al.

Under review at ACL Rolling Review.

Work in progress

"Scalable Bayesian Estimation of Heterogeneous Spillovers in Large, Sparse Networks"

Develops an estimation toolkit that exploits network sparsity to make posterior computation feasible for large networks, allows spillover effects to vary across node and link characteristics, and remains well-identified in low-density networks.

"AI-Supported Intellectual Humility: An Experimental Study" with Raiyan Abdul Baten

IRB-approved lab experiment; data collection fall 2026. Future work examines how intellectual humility shapes belief updating and collective decision-making in networked settings.

"Timeline: An AI-Assisted Cooperative Reading Game" with Glenn Smith et al.

Multiplayer reading game grounded in situation-model theory. Accepted at AECT 2026; pilot data collection fall 2026.

Pre-doctoral publications

"The Impact of Executive Compensation Incentives on Firm Size: An Empirical Study Based on Innovative Companies"

China Journal of Commerce, 2020 (in Chinese).

"The Study of Style Rotation in Chinese Stock Market"

Proceedings of the 3rd International Conference on Business and Information Management, 2019.

Research experience

Research Assistant, University of South Florida 2025 – present
Glenn Smith (AI-assisted educational games); Raiyan Abdul Baten (human–AI interaction)

Research Assistant, Peking University 2019
Shun He (quantitative investment; style rotation in Chinese equity markets)

Presentations

2026: NSF/CEME Seminar in Bayesian Inference in Econometrics and Statistics (Seattle); Eleventh Annual Conference on Network Science and Economics (Miami)

2025: Southern Economic Association Annual Conference (Tampa); Workshop in Applied and Theoretical Economics (University of Florida); Tenth Annual Conference on Network Science and Economics (Stanford, poster)

2024: Conference on Complex Systems and Complexity in Economics & Finance Workshop (Exeter, UK); NetSci (Québec City, poster); Conference on Digital Experimentation at MIT and ACM Collective Intelligence (attendee)

2019: International Conference on Business and Information Management (Paris)

Advanced training

Econometrics of Networks Summer School, VU Amsterdam 2023
Network formation modeling and peer-effects estimation. Ranked top 0.6% of cohort (10.0/10.0).

MITx MicroMasters in Statistics and Data Science 2024
Probability, Statistics, Machine Learning, Data Analysis. Distinction (90%+) in all courses and capstone exam.

Teaching experience

Principles of Macroeconomics, Instructor of Record, USF — 2023 – 2024
Two semesters (Fall 2023, Spring 2024).

Graduate Teaching Assistant, USF — 2021 – present
Time Series Analysis (graduate); Game Theory, Econometrics, Mathematical Economics, Economic Growth, Intermediate Macroeconomics, Intermediate Price Theory, Analytical Tools for Economists.

Honors and scholarships

Vernon Grigg Scholars Award, University of South Florida — 2025
Applegate & WLP Scholar, Women in Leadership & Philanthropy — 2024–26
1st Place, HackUSF (Google Developer Student Club) — 2025
1st Place, MIT Policy Hackathon, Health Challenge (MIT IDSS) — 2024
Outstanding Thesis Award, Hubei University — 2020

Service

Medical Translation Specialist, Wuhan Sixth Hospital (COVID-19) — 2021
Volunteer English Tutor, Ministry of Education of China — 2020
Community Development Assistant, State Council of China — 2018

Skills

Languages: Mandarin Chinese (native), English (fluent)
Statistical software: Python (NetworkX, NumPy, Pandas, scikit-learn, TensorFlow), R, MATLAB, Stata, Gephi
Engineering: Git, AWS, Firebase, Vercel, Empirica, LLM APIs

Statement of Motivation and Research Interests

Yajie Xu, University of South Florida

My research develops econometric methods for identifying spillover effects in endogenously formed networks, with applications to inter-firm R&D collaboration. My job-market paper, joint with Murat Munkin, develops a Bayesian MCMC framework for spatial autoregressive models in which the network itself is endogenous; by modeling link formation jointly with outcomes, it addresses the reflection problem without the exclusion restrictions that make instrumental-variable strategies fragile, and achieves higher estimation efficiency than comparable Bayesian structural models. A companion paper, with Kohei Takakura, builds a structural model of R&D link formation grounded in Cournot competition and network formation games, and uses a hand-collected dataset of executive backgrounds to show that immigrant executives act as "super-connectors" who densify collaboration networks and accelerate knowledge diffusion.

I am applying to this workshop for three reasons, all tied to where my research is heading.

First, I am designing an estimation toolkit for network effects that goes beyond my current application, and I need a clearer picture of the settings it should serve. My work so far has been on firm networks, but the method is general: it is built to scale to large networks, to allow spillovers to vary with node and link characteristics, and to remain well-identified when networks are sparse. Whether those design choices are the right ones depends on the empirical problems applied researchers actually face—the data structures, the sources of endogeneity, the policy questions. Development economics has arguably the richest empirical network literature in the discipline, and I want to understand what questions in that literature a better estimator could unlock. Hearing what problems participants and the organizers are working on would directly shape what I build.

Second, I want to learn the frontier of empirical network research from the people defining it. Much of my methodological training has come from the econometrics side—network formation models, peer-effects identification, Bayesian estimation—and I have less exposure to how these tools are being pushed in field settings, in network experiments, and in policy targeting. This workshop is the most direct way to close that gap.

Third, I am working through an open question that I would value discussing with this community: whether the machine learning methods I have been trained in can strengthen causal inference in social networks. Concurrently with my Ph.D., I am completing an M.S. in Artificial Intelligence, and I have begun thinking about where ML fits—in learning flexible link-formation models that serve as first stages, in estimating heterogeneous spillovers without imposing parametric structure, and in handling the high-dimensional node and edge characteristics that large networks generate. I do not yet know which of these directions is most promising or where the identification pitfalls lie, and I would welcome the perspective of researchers who have thought carefully about what credible network inference requires.

Beyond the networks agenda, I run lab experiments on belief updating and coordination in groups, using AI-supported platforms; I see this as the behavioral microfoundation of how networks form and how information moves through them, and I expect the two strands of my work to converge.

I would be grateful for the chance to present my work, to learn what the empirical frontier looks like from the people working on it, and to find out what a network economist working on methods can contribute to the questions that matter in development.

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

September 15, 2026

Abstract

We develop a Bayesian panel SAR model with endogenous network formation and an MCMC estimator. An application to pharmaceutical R&D alliances reveals positive revenue spillovers and negative dependence between outcome shocks and alliance formation. Immigrant leadership is positively associated with own and collaborators' revenues and alliance formation. Counterfactuals at posterior mean parameters, allowing network rewiring, show that changing the immigrant-led share of U.S. firms from 51.9% to approximately 20% or 80% changes average U.S.-firm log revenue by -0.077 or 0.048, respectively. Effects extend to firms whose leadership status is unchanged, highlighting the importance of jointly modeling outcomes and networks.

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 model 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 generating the endogenous weight matrix is assumed to have a probit structure instead of the usual logit choice adopted by the existing literature. That 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 which simplifies the MCMC algorithm. Since the underlying distributions are normal the posterior distribution can be augmented with the corresponding latent variables the distributions of which are also normal. Finally, instead of generating endogeneity by introducing several unobserved variables into the network model representing unobservable human traits, such as tastes and ability, as in (Hsieh and Lee, 2016) we do so by placing a single error term from the outcome equation into the latent variables of the network equations. In addition to the aimed relative computational simplicity, this is also done to avoid multiple potential identification issues 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 emphasized by Manski (1993) as the inability of distinguishing between 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 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, health behavior (Hsieh and Lee, 2016; Boucher et al., 2014; Hsieh and Lin, 2017; Chung, 2020; Han et al., 2021) and 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), shocks 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 according to the simple latent-variable method 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 which could be easily 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 commonly used MH steps including those 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) also observes identification difficulties of finite mixture models even when the outcome variable is an overdispersed count. The problem is likely to persist or even get worse in the case of latent utility generating the 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 performance. Then the estimation method is applied to 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 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 misleading conclusion 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 presents 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. Influence runs the other way: row i of W_t collects the firms whose outcomes enter i's own, so it is j that influences i. Sender and receiver therefore describe who nominates whom, not who influences whom. 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 $Y_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ be an $n \times 1$ vector of outcomes (company's revenue in our application) at time t, $X_t = (x_{1t}, x_{2t}, \ldots, x_{nt})'$ 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:

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

where λ represents the endogenous peer effect, which is 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 $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)'$ is an $n \times 1$ vector capturing time-invariant individual fixed effect. Also $\tau = (\tau_1, \tau_2, \ldots, \tau_t)'$ 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 sector 1 and period 1 as the baseline (Han et al. 2021; Greene 2018)[1] Finally, $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{nt})'$ is a vector of stochastic error terms, where $\varepsilon_{it} \overset{i.i.d.}{\sim} N(0, \sigma_\varepsilon^2)$. Equation (1) in the vector form for each i $(i = 1, ..., n)$ and t $(t = 1, ..., T)$ can be written as

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

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 that case, replace α_i in Equation (2) by $\alpha_{s(i)}$, where $s(i) \in \{1, \ldots, S\}$ denotes the sector to which firm i belongs and α_s is the corresponding sector effect, with the normalization $\alpha_1 = 0$ retained. This substitution can be accommodated with a very small change to the MCMC algorithm described in Section 3.

[This line can be deleted: We reserve the symbol ζ exclusively for the composite error of the network formation equation introduced in Section 2.1.]

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

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

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

$$\Delta u_{ijt} = \psi_{ijt} + \zeta_{ijt} \tag{4}$$

[1]These 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 $(s = 2, \ldots, S)$ measure sector-specific deviations relative to sector 1, and the remaining time effects τ_t $(t = 2, \ldots, T)$ measure period-specific deviations relative to period 1.

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}. However, 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

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

where 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 distibution of Δu_{ijt} given ε_{it} is fixed for identification of the Probit part. Then

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

and the observability condition is specified as

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

Thus endogeneity of the spatial choice matrix W_t operates entirely through δ: the same shock ε_{it} that moves firm i's outcome in (2) also moves the latent utilities governing which firms i nominates.

Side notes: The unit variance in (6) is a normalization rather than a restriction on magnitudes. Only the sign of Δu_{ijt} is observed, so multiplying (6) through by any constant $c > 0$ leaves every w_{ijt} unchanged; one restriction is needed to pin down the scale. Setting $\mathrm{Var}(\eta_{ijt}) = 1$ leaves δ unrestricted. Another way is to normalize the composite variance σ_ζ^2 to unity, but it would require $1 - \delta^2\sigma_\varepsilon^2 > 0$ and so confine δ to $|\delta| < \sigma_\varepsilon^{-1}$, a constraint that would

have to respect at every draw in MCMC simulation, which can increase the complexity and decrease the algorithmic efficiency. From (5),

$$\sigma_\zeta^2 \equiv \mathrm{Var}\left(\zeta_{ijt}\right) = 1 + \delta^2\sigma_\varepsilon^2, \qquad \sigma_{\varepsilon\zeta} \equiv \mathrm{Cov}\left(\varepsilon_{it}, \zeta_{ijt}\right) = \delta\sigma_\varepsilon^2, \tag{8}$$

equivalently $\delta = \sigma_{\varepsilon\zeta}\sigma_\varepsilon^{-2}$ and $\sigma_\zeta^2 = 1 + \sigma_{\varepsilon\zeta}^2\sigma_\varepsilon^{-2}$; the coefficient δ in (5) is thus the population regression slope of ζ_{ijt} on ε_{it}. The free parameters of the error structure are $(\sigma_\varepsilon^2, \delta)$. Being linear combinations of the independent normal pair $(\varepsilon_{it}, \eta_{ijt})$, the two errors are jointly normal. 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 (6) where

$$\left(\varepsilon_{it}, \zeta_{ijt}\right)' \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\varepsilon^2 & \sigma_{\varepsilon\zeta} \\ \sigma_{\varepsilon\zeta} & \sigma_\zeta^2 \end{pmatrix}\right), \tag{9}$$

with conditional distribution $\zeta_{ijt} \mid \varepsilon_{it} \sim N(\delta\varepsilon_{it}, 1)$.[2]

The sign of δ has an economic interpretation. Since ε_{it} enters both the outcome equation and the linking index, $\delta < 0$ implies that a firm experiencing an unfavorable revenue shock derives higher latent utility from collaborating partners, and therefore nominates more of them. The specification is a selection model in the sense of Heckman (1979): the network is observed only as the outcome of a choice that the unobservable ε_{it} helps determine, so that conditioning on an observed link 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

$$Y_t \;=\; (I_n - \lambda W_t)^{-1}(\beta_0\mathbf{1}_n + X_t\beta_1 + W_tX_t\beta_2 + \alpha + \tau_t\mathbf{1}_n) + (I_n - \lambda W_t)^{-1}\varepsilon_t, \tag{10}$$

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

[2]Relative to earlier drafts, the second diagonal entry and the off-diagonal entry of (9) are now written σ_ζ^2 and $\sigma_{\varepsilon\zeta}$ rather than σ_η^2 and $\sigma_{\varepsilon\eta}$. The matrix is the covariance matrix of $(\varepsilon_{it}, \zeta_{ijt})$, and the quantities in question are those in (8); neither refers to η_{ijt}, for which $\mathrm{Var}(\eta_{ijt}) = 1$ and $\mathrm{Cov}(\varepsilon_{it}, \eta_{ijt}) = 0$.

$Z_t = (\mathbf{1}_n, X_t, W_t X_t)\,, Z_{it} = \left(1, x_{it}, \sum_{j=1}^n w_{ijt} x_{jt}\right), C_{ijt} = (c_{it}, c_{jt}, c_{ijt})$, $Q_{ijt} = (C_{ijt}, \varepsilon_{it})$, and parameters $\Gamma = (\gamma_1, \gamma_2, \gamma_3)'$, $\theta' = (\Gamma', \delta)$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$, $\Theta = \{\lambda, \beta, \alpha, \tau, \sigma_\varepsilon, \theta\}$. Based on Equation (10) and (6), the joint likelihood can be written as

$$
\begin{aligned}
& P(Y, \Delta u \mid W, \Theta) \qquad (11)\\
\propto\ & \Big\{(2\pi\sigma_\varepsilon^2)^{-\frac{Tn}{2}} |I - \lambda W_t|^T \\
& \times \exp\left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^T \sum_{i=1}^n [y_{it} - \lambda \sum_{j=1}^n w_{ijt} y_{jt} - \alpha_i - \tau_t - Z_{it}\boldsymbol{\beta}]^2\right) \\
& \times \exp\left[-\frac{1}{2} \sum_{t=1}^T \sum_{i=1}^n \sum_{j\neq i}^n \left(\Delta u_{ijt} - C_{ijt}\Gamma - \delta[y_{it} - \lambda \sum_{j=1}^n w_{ijt} y_{jt} - \alpha_i - \tau_t - Z_{it}\boldsymbol{\beta}]\right)^2\right]\Big\}.
\end{aligned}
$$

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:

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

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

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

4. For each individual $i = 1, \ldots,$n, we specify a prior $\alpha_{i} \sim N\left(\underline{\alpha}_{i},\ \underline{H}_{\alpha_{i}}^{-1}\right)$, with α_{1} fixed at zero. Then the full conditional distribution of α_{i} is

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

where

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

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

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

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 (11). 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).

Updated Version:

S1.1: 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 condition[3], the proposal is rejected and set $\lambda^{(q)} = \lambda^{(q-1)}$.

S1.2: 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, $\lambda^{(q)} = \lambda^{(q-1)}$.

Old Version:

S1.1: Propose $\tilde{\lambda} \sim N(\lambda^{(q-1)}, c_\lambda)$, where $\lambda^{(q-1)}$ is the current state of the chain and variance c_λ is adjusted to achieve an acceptance rate between 30% and 40%. Check whether $\tilde{\lambda}$ satisfies the stability condition implied by the prior. If not, redraw $\tilde{\lambda}$ until it satisfies that condition.

S1.2: Accept the drawn value with the acceptance probability

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

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

[3]The 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.

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, setting 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 $s \in \{2, \ldots, 5\}$ and $\tau_t \sim \mathcal{N}(0, 1)$ for $t \in \{2, \ldots, 20\}$.

The directed network adjacency matrix W_t is constructed by first generating latent utilities and evaluating their differences Δu_{ijt}. Next we describe how the utility levels were constructed. We generate two-dimensional firm-level covariates, $c_{it} \sim \mathcal{N}(0, I_2)$, representing node-specific attributes. To conform to the dyadic space, these vectors are expanded to capture both sender (i) and receiver (j) characteristics for all $N(N-1)$ pairs in each period. We also draw three-dimensional dyad-specific covariates, $c_{ijt} \sim \mathcal{N}(0, I_3)$, to capture homophily or synergistic effects. The structural error term ε_{it} is drawn from $\mathcal{N}(0, 1)$ and enters the network formation process to introduce endogeneity, then it is mapped to the dyadic space in the same way as c_{it}. Finally, the idiosyncratic edge-specific shock is drawn as $\eta_{ijt} \sim \mathcal{N}(0, 1)$. The network formation parameters are set 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 $\mathcal{U}_d\{0.1, 0.2, \ldots, 0.9\}$; and the coefficient capturing the endogenous selection effect of the structural error is set to $\delta = -0.5$. Network indicators are set to 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 economic outcome vector Y_t. The matrix of exogenous covariates X_t consists of an intercept and four explanatory variables independently drawn 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 $\lambda = 0.3$. Combining the exogenous covariates, the unobserved sector (α_s) and time (τ_t) fixed effects, the simulated spatial weight matrix W_t (row-normalized), and the

stochastic error term ε_{it}, the final equilibrium outcome vector Y_t is constructed according to the reduced form solution of Equation (2).

Table 1: Monte Carlo experiment results

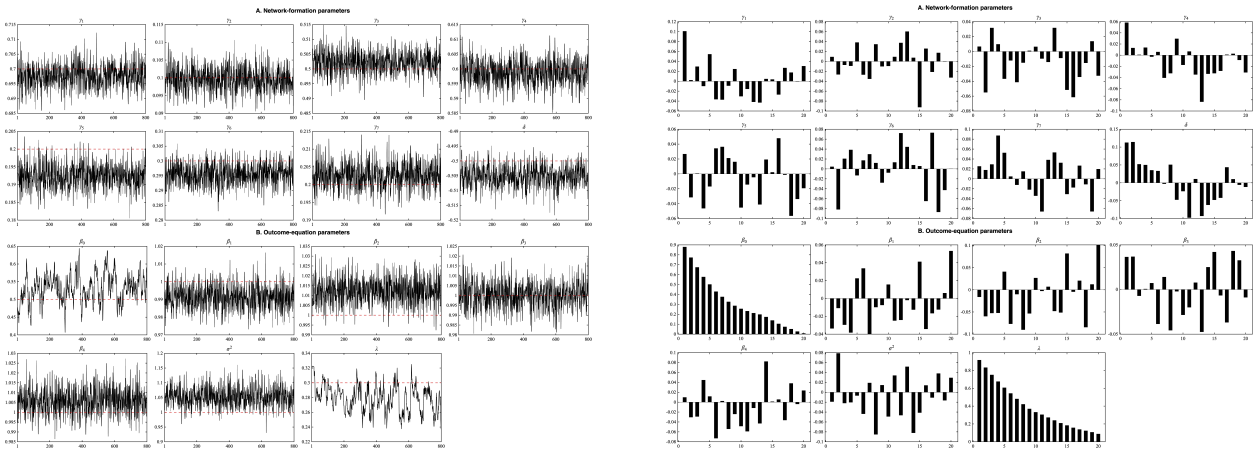
Parameter	True	Mean	SD	Ratio	ACF20	CI95
Spatial Autoregressive (SAR) Outcome Model						
λ	0.300	0.289	0.008	37.305	0.096	0.274, 0.304
β_0	0.500	0.559	0.058	9.679	0.715	0.430, 0.661
β_1	1.000	1.014	0.011	90.091	0.026	0.992, 1.035
β_2	1.000	1.004	0.011	91.255	−0.021	0.982, 1.026
β_3	1.000	1.015	0.011	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
Network Formation Model						
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.

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

Overall, the estimation method estimates the true spatial and structural parameters with high accuracy. The posterior mean of the endogenous peer effect is $\hat{\lambda} = 0.289$ with a 95% credible interval of $[0.274, 0.304]$, with the true value of 0.3 right inside of it. Similarly, the estimated endogeneity parameter $\hat{\delta} = -0.514$ (95% CI $[-0.531, -0.498]$) is precisely estimated around the true value of -0.5. This shows that the MCMC method correctly captures the structural correlation between the unobserved network formation process and the spatial autoregressive disturbances even in the extreme case of low network density. The coefficients for the exogenous characteristics (β_1 through β_4) are also well estimated near their true value of 1.0, exhibiting high precision with mean-to-standard-deviation ratios approaching 90.

Figure 1 displays the trace plots and autocorrelation functions (ACF) for the thinned



(a) MCMC Trace Plots (b) Autocorrelation Plots

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

Notes: Diagnostics are computed from 20,000 MCMC draws after a 2,000-draw burn-in, retaining every tenth draw. Both panels use the same thinned posterior sample. Panels are grouped by equation: 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 parameter symbols reported in Table 1.

posterior draws. The MCMC algorithm achieves an acceptance rate of 40.07% of the MH step for the spatial parameter λ, within the optimal target 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. While moderate autocorrelation persists for the baseline intercept (γ_0) and the global outcome intercept (β_0), this is an expected outcome for anMCMC algorithm involving extensive fixed effects. Overall, the simulation results provides strong evidence that the proposed methodology yields reliable, unbiased 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 of results for the d.g.p. consistent with the application.

5 Empirical Application

The structural model developed in this paper is applied 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. To alleviate the concern of reverse causality, all financial controls are lagged one year: 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. 2003 observations are used to construct the first lags, so the estimation sample covers 2004–2022. Table 2 presents the variables used by the outcome equation and their descriptive statistics.

ImmigrantExecutive (Imm) is a time-varying dummy variable that equals one when at least one senior executive was identified as an immigrant in that year. Information about

[4] 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 (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

The outcome is contemporaneous; the five financial regressors are lagged one 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. *ImmigrantExecutive* is the time-varying U.S.-firm indicator defined in the text and equals zero for non-U.S. firms. *ImmigrantExecutive (Contextual)* is the row-normalized share of a firm's current collaborators that are immigrant-led; it is zero for isolates. Statistics cover 3,080 firm-years from 172 SIC-283 firms over 2004–2022.

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 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 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).

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 the binary variable representing the 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 the pair and zero otherwise. In particular, *IM* equals one when at least one firm in the pair is immigrant-led. *OldFriends* is the cumulative number of the pair's prior alliance initiations; its observed range is zero to four. Variables prefixed *Dif* are absolute differences between the 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. Statistics are computed over the 248,226 dyad-years of the estimation sample, 2004–2022, of which 1,272 are linked.

The dyadic variables, reported in Table 3, capture the relational, demographic and financial characteristics of each firm pair. *ImmigrantExecutive (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 the two firms was born in the same country. *OldFriends* counts the 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 the two firms are located in the same city, and *SameIndustry* equals one if they share the same four-digit SIC code. Together these indicators characterise the social, geographic and technological proximity that may influence the formation of interfirm R&D collaborations.

The annual network outcome, denoted *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 rather than being discarded.

Homophily, the tendency to link with similar partners, is among the most important factors in deciding whom to partner with (Mele, 2022; Hsieh and Lin, 2017). We therefore enter all five lagged financial attributes in the network-formation equation as absolute differences between the two firms in a pair. These are the variables prefixed *Dif*: *DifAssets*, *DifLeverage*, *DifReturnOnAssets*, *DifR&D* and *DifCost*, measuring the gap between partners in size, leverage, profitability, R&D expenditure and cost of revenue, respectively. This specification lets the data determine the direction of sorting along each dimension: a negative coefficient indicates homophily, since a wider gap then lowers the propensity to link, whereas a positive coefficient indicates complementarity, with firms favouring 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 followed by the discarded 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 both tables 4, 5 and plots 2 (Geweke, 1992; Raftery and Lewis, 1992; Heidelberger and Welch, 1983).

For comparison, the tables also report estimates obtained by treating the two equations separately: OLS for the outcome equation and maximum-likelihood Probit for the network equation. The Bayesian columns instead report the joint estimator developed in Section 3,

which estimates the outcome and network equations simultaneously and allows for endogenous network formation through δ.

The coefficient of main interest is the endogenous spillover parameter λ, representing the strength of peer or network interactions in its effect on the output among firms. 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 posterior mean is approximately 0.046, with zero being outside the 95% credible intervals (0.027, 0.064), indicating statistically strong spillovers. Compared to the structural models, standard OLS underestimates the spillover effect (0.034) when network endogeneity is unaddressed, consistent with evidence from König et al. (2019).

For a connected firm in 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, through feedback among connected firms. The estimates thus support 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.

In terms of immigrant-related variables, *ImmigrantExecutive* has a posterior mean of 0.113, with a 95% credible interval that excludes zero, while the posterior mean for *US* is -0.372. Together these two coefficients partition the sample into three groups. The reference group is non-U.S. firms, for which $US = 0$ and $ImmigrantExecutive = 0$. The second group is U.S. firms without an identified immigrant executive, for which $US = 1$ and $ImmigrantExecutive = 0$; the *US* coefficient implies that their expected log revenue is lower by 0.372 log points, corresponding to revenue about $e^{-0.372} - 1 \approx 31.1\%$ lower than the reference group. The third group is U.S. firms with an identified immigrant executive, for which both indicators equal one; relative to the second group, their expected log revenue is higher by 0.113 log points, or revenue roughly $e^{0.113} - 1 \approx 12.0\%$ higher. Measured against 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 revenue. 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.

Meanwhile, the contextual regressor $W_t ImmigrantExecutive$ has a posterior mean of 0.265, with a 95% credible interval of [0.053, 0.468]. Holding the network fixed, a 10-percentage-point increase in the immigrant-executive share among a firm's collaborators therefore leads

to an increase of about $e^{0.1\times0.265}-1 \approx 2.7\%$ in expected revenue. Beyond this direct impact, a change in *ImmigrantExecutive* also propagates through the feedback loop in the network. The corresponding impact matrix is $S(W_t) = (I-\lambda W_t)^{-1}(\beta_{\text{Imm}} I + \beta_{W\text{Imm}} W_t)$, whose average row sum gives the total effect (LeSage and Pace, 2009). As an illustrative back-of-the-envelope calculation, if W_t is row-normalized and every firm in a connected network 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. This scalar benchmark applies to connected firms, holding the network fixed. Section 5.3 therefore provides a further analysis that applies feasible changes to U.S. firms and allows the collaboration network to adjust endogenously.

Among the lagged own-firm characteristics, *Assets* and *CostOfRevenue* dominate 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] that lie well away from zero. Because both regressors and the 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 are less significant. *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 contain zero. *ReturnOnAssets* is the most suggestive of the three, entering negatively at −0.078 with interval [−0.164, 0.006] and a posterior probability of 0.967 of being negative, so a one-standard-deviation increase in lagged profitability corresponds to about 7.3% lower current revenue, but the interval still covers zero, and we do not treat it as decisive.

The estimated coefficients of the probit equations show that the results for the same parameters from the Bayesian and standard MLE approaches are largely consistent in both magnitude and statistical significance. The parameter of main interest is the endogeneity term δ, which is negative and statistically very strong and available only in the model that allows endogeneity. The 95% credible intervals in both specifications, with and without contextual effects, do not include zero. 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 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 possibly 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 on *ImmigrantExecutive*, with posterior mean 0.259 and a 95% credible interval of [0.196, 0.322] that excludes zero, 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 insignificant: *SameCountryOfBirth* (0.003), *SameIndustry*(0.000), *SameCity* (−0.048) and *BothUS* (−0.092) each have intervals containing zero, 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—a result that may reflect the dominance of global principles among collaborating firms.

The five dyadic financial regressors enter purely in difference form, so each reveals whether similarity or dissimilarity along that dimension attracts partners. The results point in both directions. *DifReturnOnAssets* is negative at −0.193 with interval [−0.259, −0.133], the strongest 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. *Dif*R&D at 0.031 [0.016, 0.046] and *DifCost* at 0.030 [0.015, 0.043] are smaller but have intervals above zero, and 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 effectively zero with intervals spanning 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: separate OLS versus joint Bayesian

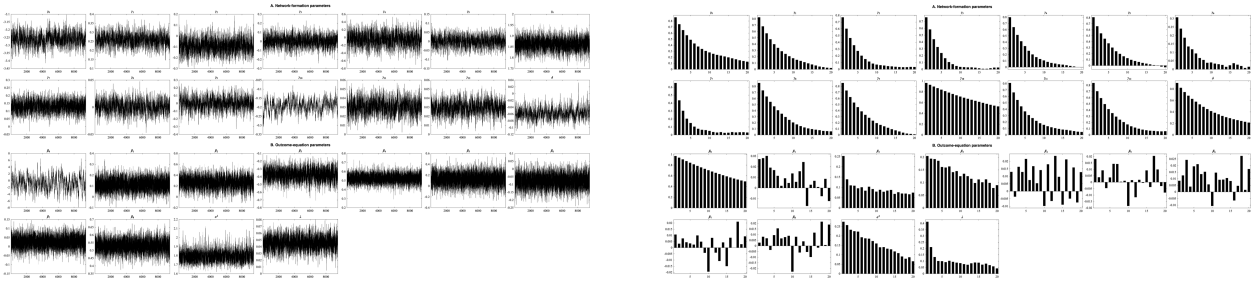
		OLS		Bayesian			
Variable	**Parameter**	**Estimate**	**SE**	**Mean**	**SD**	**ACF20**	**CI95**
Spillover effect	λ	0.034	0.008	0.046	0.009	0.042	$0.027, 0.064$
Constant	β_0	-0.325	0.334	-0.723	2.010	0.496	$-4.616, 3.165$
Imm	β_1	0.139	0.071	0.113	0.055	-0.012	$0.007, 0.224$
Imm (Contextual)	β_2	0.201	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. *ImmigrantExecutive (Contextual)* is W_t*ImmigrantExecutive*, 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 estimates: separate Probit MLE versus joint Bayesian

		MLE		Bayesian			
Variable	**Parameter**	**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.



(a) MCMC Trace Plots (b) Autocorrelation Plots

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

Notes: Diagnostics are computed from 100,000 MCMC draws after a 5,000-draw burn-in, 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 estimates in Section 5.2 imply that immigrant executives can affect firm performance through a firm's own status (Imm), its collaborators' immigrant-executive composition ($W_t Imm$), and alliance formation (IM). No single coefficient captures the combined effect of these channels. Following Han et al. (2021), we therefore evaluate a counterfactual change in the share of immigrant-led U.S. firms, allowing the intervention to affect both the outcome and network-formation equations and propagate through the spatial multiplier.

The counterfactual is evaluated for 2022, the last year in our dataset, when the estimation sample contains 156 firms. Seventy-nine are U.S. firms, of which 41 are immigrant-led at baseline; the observed immigrant-led share among U.S. firms is consequently $41/79 = 51.9\%$. We examine two distinct national policies. The contraction policy, targeting 20%, reduces the share to 16 of 79 U.S. firms (20.3%) by changing the status of 25 baseline immigrant-led firms from one to zero. The expansion policy, targeting 80%, raises the share to 63 of 79 firms (79.7%) by changing 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. Thus, the simulations average over outcome and link shocks; they do not repeatedly redefine who receives the policy.

Following the discrete-policy exercise in Appendix F of Han et al. (2021), we evaluate the policy effects using posterior mean parameter estimates presented in Tables 4 and 5. Let $\widehat{\Theta}$ denote this vector of posterior means, using the parameter vector Θ defined in Section 3, and let $a \in \{0, p\}$ index the observed immigrant assignment and the assignment under a target

share $p \in \{0.2, 0.8\}$; the superscript 0 denotes the observed baseline, not a zero immigrant share. Let Y_i^a be firm i's log revenue obtained from Equation (1) under assignment a, with the sector fixed effects used in the empirical specification. Under endogenous rewiring, the network is generated by the estimated network-formation model in Section 2. The assignment changes Imm and the dyadic regressor IM, and $W_t Imm$ is recomputed for each generated network. All other covariates and the sector and year effects remain at their 2022 values. This is a static counterfactual for 2022, rather than a forecast for the following year.

The firm-level expected revenue effect under endogenous rewiring is

$$ME_i^{\text{rewired}}(p) = \mathbb{E}\left[Y_i^p \mid \widehat{\Theta}\right] - \mathbb{E}\left[Y_i^0 \mid \widehat{\Theta}\right]. \tag{12}$$

Both expectations integrate over the outcome and link shocks, which jointly determine the network and revenue, conditional on the fixed assignments and 2022 covariates described above. They do not integrate over posterior parameter uncertainty. We approximate this difference using paired simulations:

$$\widehat{ME}_i^{\text{rewired}}(p) = \frac{1}{R}\sum_{r=1}^{R}\left(Y_i^{p,r} - Y_i^{0,r}\right), \qquad R = 1{,}000, \tag{13}$$

where $Y_i^{a,r}$ is the outcome under assignment a in replication r. Within each replication, we use identical draws of the firm-level outcome shocks and dyad-level link shocks in the baseline and policy scenarios to reduce simulation noise in their difference. Expected degree effects are calculated analogously by replacing log revenue with the number of collaborators. For comparison, $\widehat{ME}_i^{\text{fixed}}(p)$ holds the observed 2022 network fixed under both assignments and rebuilds $W_t Imm$ under the policy assignment. With the same network in both terms, the outcome shock and unchanged fixed effects cancel from the revenue difference, so this fixed-network effect is calculated directly without simulation.

Figures 3 and 4 follow the presentation in Han et al. (2021). Each observation in a histogram is one firm's expected effect after averaging over the 1,000 simulated networks. The figures therefore describe heterogeneity across firms. Their horizontal spread is not a confidence interval, a posterior credible interval, or the distribution of the 1,000 raw simulations. The first row reports firms whose immigrant-executive status changes (switchers), the second reports all U.S. firms whose status does not change (stayers), and the third reports all 79 U.S. firms. The third row intentionally combines the firms shown in the first two rows.

Figure 3 shows that the change in alliance formation is strongest among switchers, but is also present among non-switchers. Under the contraction, U.S. firms lose an average of 0.170 links: switchers lose 0.457 links on average, while non-immigrant-led stayers lose 0.052.

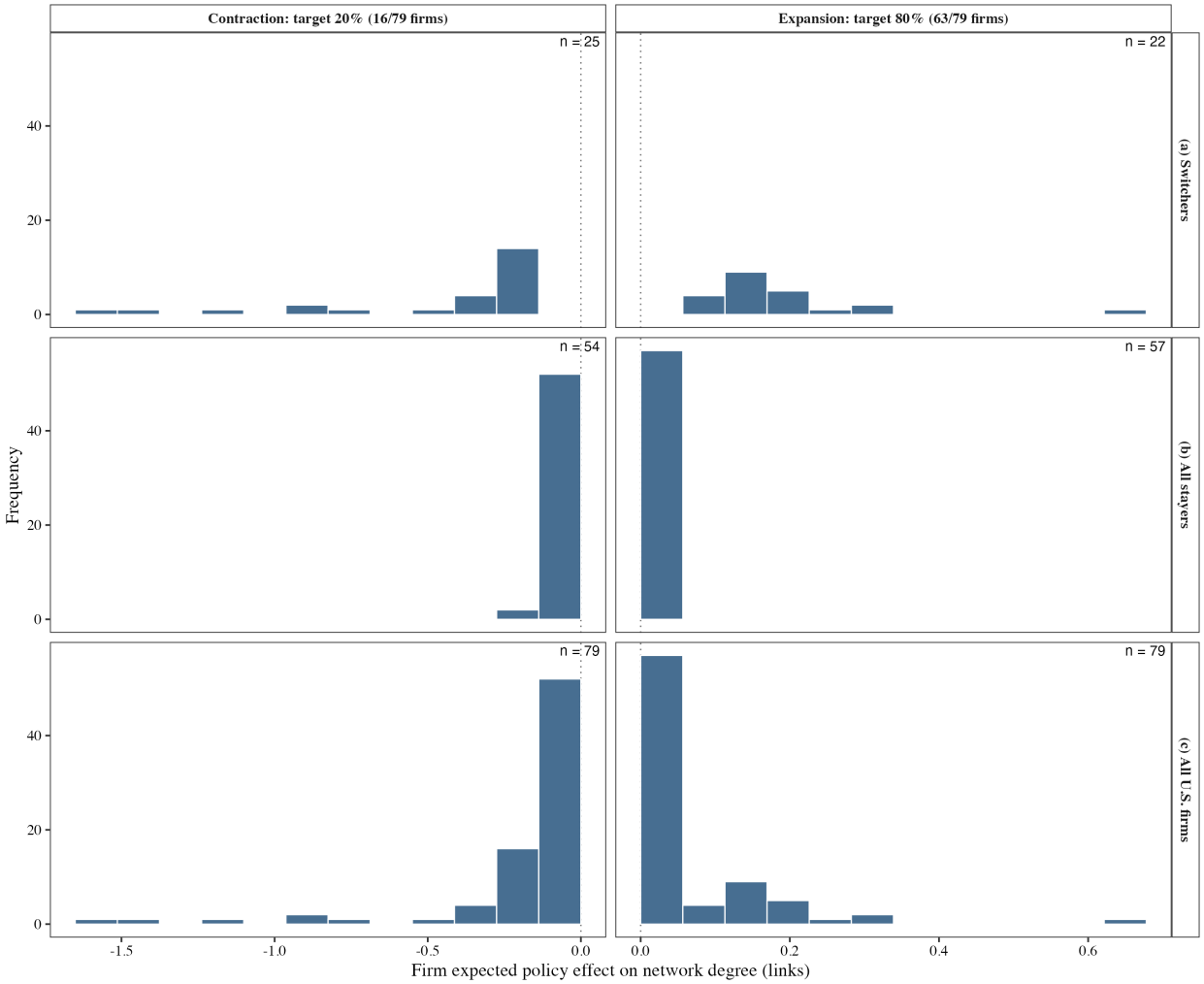


Figure 3: Distributions of expected network-degree effects under contraction and expansion

Notes: Frequency histograms across U.S. firms. The columns compare a contraction to 16/79 immigrant-led U.S. firms and an expansion to 63/79 with the observed baseline of 41/79. Each observation is a firm's effect averaged over 1,000 paired network simulations at fixed posterior-mean parameters. Degree effects are changes in links per firm. Fixed-network degree effects are identically zero and are not plotted. Recipient assignments are fixed across simulations.

Under the expansion, the average U.S.-firm gain is 0.056 links, including 0.192 for switchers. The asymmetry is partly mechanical because contraction changes 25 firms whereas expansion changes 22, and it may also reflect the nonlinear threshold rule governing link formation. It should therefore not be read as an equal-dose comparison of policy directions.

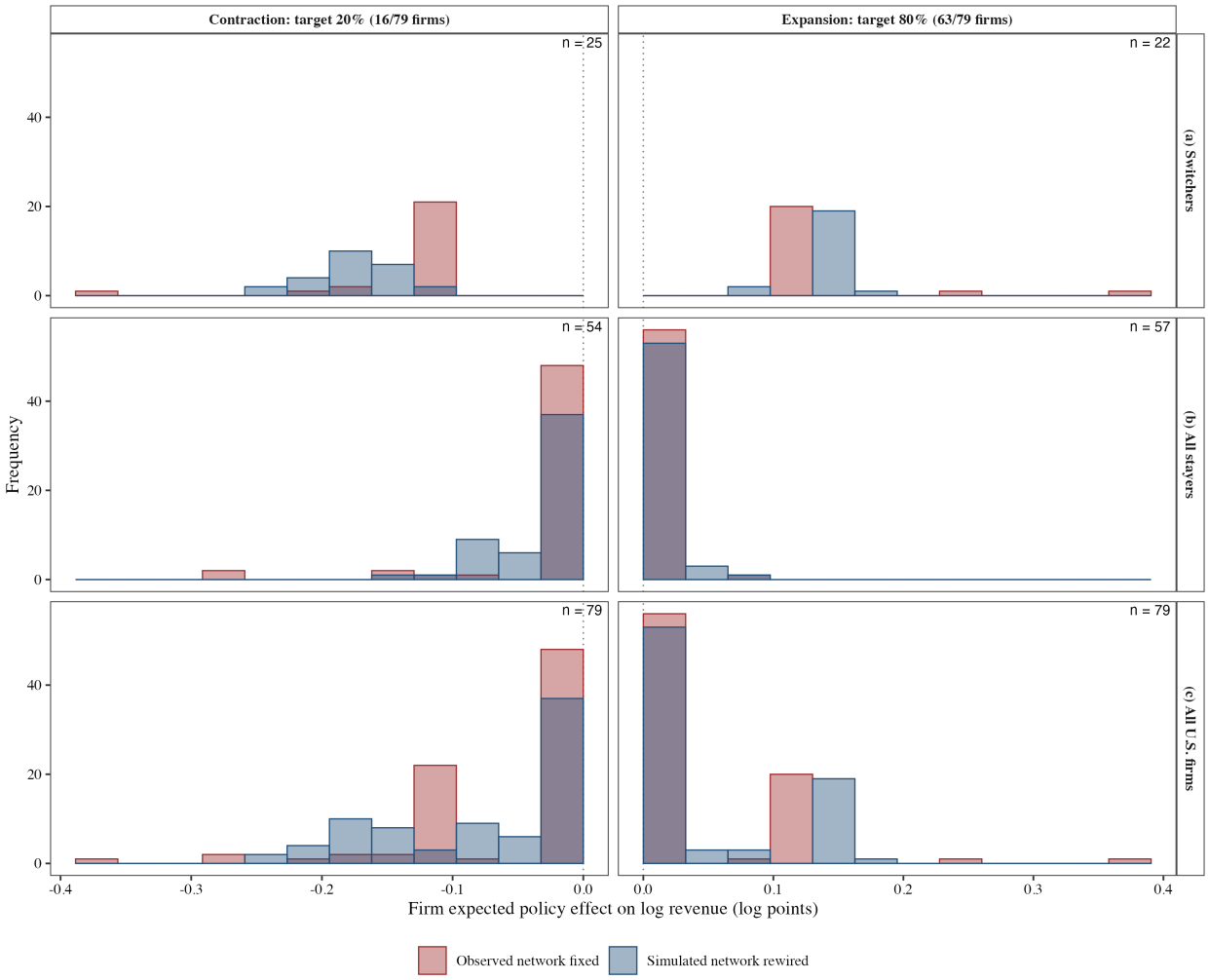


Figure 4: Distributions of expected revenue effects under contraction and expansion

Notes: Frequency histograms across U.S. firms of expected effects on log revenue. Transparent red bars hold the observed 2022 network fixed; blue bars simulate both baseline and policy networks. Common bin boundaries are used for the two scenarios within each panel. Each rewired effect is averaged over 1,000 paired simulations at fixed posterior-mean parameters. Histogram spread measures cross-firm heterogeneity, not statistical uncertainty. Financial covariates and fixed effects remain at their 2022 values.

The revenue distributions in Figure 4 show both own-firm responses and propagation to firms whose status is unchanged. Contraction shifts the effects to the left: average U.S.-firm log revenue changes by -0.056 when the observed network is fixed and by -0.077 when networks are allowed to rewire. Expansion produces corresponding effects of 0.038 and 0.048. Stayers have no direct change in Imm, yet their rewired effects are generally nonzero. This

is the clearest evidence in this experiment that the intervention reaches beyond recipient firms through contextual exposure, spatial outcome interactions, and endogenous changes in collaboration. The difference between the fixed and rewired scenarios is informative about network adjustment, but it is descriptive: the two scenarios use different network baselines and are not an exact additive decomposition into mutually exclusive causal channels.

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

		Log revenue			Degree	
Group	N	Fixed	Rewired	SD	Rewired	SD
Contraction: target 20% (16/79 firms)						
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
Expansion: target 80% (63/79 firms)						
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

Parameters are fixed at posterior means from 9,500 retained C8 draws. Each firm's rewired effect averages 1,000 paired network simulations. Effects are policy minus baseline: log points for revenue and links per firm for degree.
SD is the standard deviation across firm-level expected effects after averaging over simulations. It measures cross-firm heterogeneity, not a standard error or posterior uncertainty. Fixed-network degree effects are zero.
Groups are policy-specific. Non-U.S. firms are never switched.

Table 6 reports the detailed police effects. Under contraction, the relatively large loss among the 25 switchers combines with smaller negative effects among both stayer groups to produce the -0.077 U.S.-firm average. Under expansion, switchers gain 0.143 log points, while both stayer groups gain about 0.011, yielding an aggregate effect of 0.048. The SD column quantifies how much firm-level expected effects differ across firms; it does not test whether the group average differs from zero.

Table 7 compares the switcher effect with a simple no-propagation benchmark. If only the own Imm coefficient were in effect, a switcher's effect would be -0.113 under contraction and 0.113 under expansion. The rewired effects are -0.173 and 0.143, respectively, corresponding to ratios of 1.527 and 1.262. These ratios demonstrate the amplification effect relative to the own-status coefficient; they do not separately identify the contributions of $W_t Imm$, network rewiring, and the spatial multiplier. The very small MCSE values show that 1,000 network replications estimate the simulation averages precisely. They are not standard errors for the economic effects and should not be used for significance claims. The posterior-draw policy

Table 7: Own-status benchmarks and policy propagation

Group	Benchmark	Effect	SD	MCSE	Ratio
Contraction: target 20%					
Switchers	-0.113	-0.173	0.034	0.00095	1.527
Immigrant-led stayers	0.000	-0.036	0.032	0.00052	—
Non-immigrant-led stayers	0.000	-0.031	0.031	0.00052	—
Expansion: target 80%					
Switchers	0.113	0.143	0.021	0.00081	1.262
Immigrant-led stayers	0.000	0.011	0.011	0.00019	—
Non-immigrant-led stayers	0.000	0.011	0.021	0.00035	—

Effects use simulated rewired networks. For switchers, the benchmark is minus or plus the posterior-mean own-status coefficient, $\widehat{\beta}_{Imm} = 0.113$, under contraction or expansion. A stayer's benchmark is zero because its own status does not change.

MCSE is the standard deviation of the group-average effect across the 1,000 network simulations, divided by $\sqrt{1{,}000}$. Its numerator is the simulation-to-simulation dispersion of the group average, not the cross-firm SD reported in the preceding column. It measures numerical precision only and excludes parameter uncertainty. Ratio is the switcher effect divided by its signed benchmark; ratios relative to a zero benchmark are undefined.

spectrum, which additionally propagates parameter uncertainty, is reported in Appendix 6.1 as a robustness analysis.

6 Conclusion

This paper develops a Bayesian panel SAR model that jointly accounts for firm outcomes and endogenous network formation. The outcome disturbance enters the latent utility of the Probit link equation, allowing unobserved determinants of performance to influence collaboration decisions. This structure yields a tractable MCMC algorithm with Gibbs updates for all but the spatial autoregressive parameter. The simulation study demonstrates parameter recovery in a sparse network resembling the empirical application.

The application to pharmaceutical R&D alliances provides evidence of positive revenue spillovers and endogenous collaboration. The posterior mean of λ is 0.046, while that of the dependence parameter δ is -0.058; both have 95% credible intervals excluding zero. The negative δ indicates that positive unobserved revenue shocks are associated with lower latent incentives to form alliances, conditional on the model's covariates and fixed effects. Immigrant-executive status enters positively in the revenue equation, as does exposure to immigrant-led collaborators through $W_t Imm$. The positive coefficient on the dyadic variable IM also links immigrant leadership to alliance formation. Together, these findings motivate evaluating immigrant leadership through the outcome and network equations jointly.

The 2022 counterfactuals illustrate the combined implications of these channels. Starting from an immigrant-led share of 51.9% among the 79 U.S. firms, contraction toward 20% reduces average U.S.-firm log revenue by 0.077 under endogenous rewiring, while expansion toward 80% raises it by 0.048. The corresponding changes with the observed network held fixed are -0.056 and 0.038 log points. Under rewiring, average network degree changes by -0.170 and 0.056 links per U.S. firm, respectively. Revenue effects extend beyond firms whose immigrant-executive status changes to firms whose status remains unchanged. The comparison therefore illustrates how contextual exposure, network adjustment, and spatial feedback jointly shape the effects of an intervention.

These counterfactuals are conditional on the estimated model, fixed recipient assignments, and 2022 covariates. The main exercise averages over outcome and link shocks at posterior mean parameters; its distributions describe heterogeneity across firms rather than posterior uncertainty. The appendix additionally examines parameter uncertainty using posterior draws. Contraction and expansion change different numbers of firms, and the fixed and rewired scenarios use different network baselines, so their comparison does not isolate a single causal channel. Within these limits, the results show why assessing changes in immigrant leadership requires accounting for both firms' own responses and the collaboration networks through which effects propagate.

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 Posterior-Draw Immigrant-Executive Policy Spectrum

The main-text policy exercise follows Han et al. (2021) by fixing the model parameters at their posterior means and averaging over simulated networks. However, this approach does not incorporate parameter uncertainty. As a robustness analysis, for each of 1,000 selected posterior draws, the procedure simulates 200 paired baseline and policy networks, averages the network replications within that draw, and then summarizes the resulting distribution across posterior draws. Consequently, the shaded bands below are 95% posterior credible intervals, unlike the cross-firm histograms in the main text.

The horizontal axis gives the target immigrant-led share among the 79 U.S. firms. 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 the 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 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 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 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 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 whose status 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 experience no change in their own Imm indicator, their response reflects propagation through contextual exposure, endogenous network adjustment, and spatial outcome interaction.

At the 75% target, the switcher effect is 1.26 times the posterior-mean own-status co-

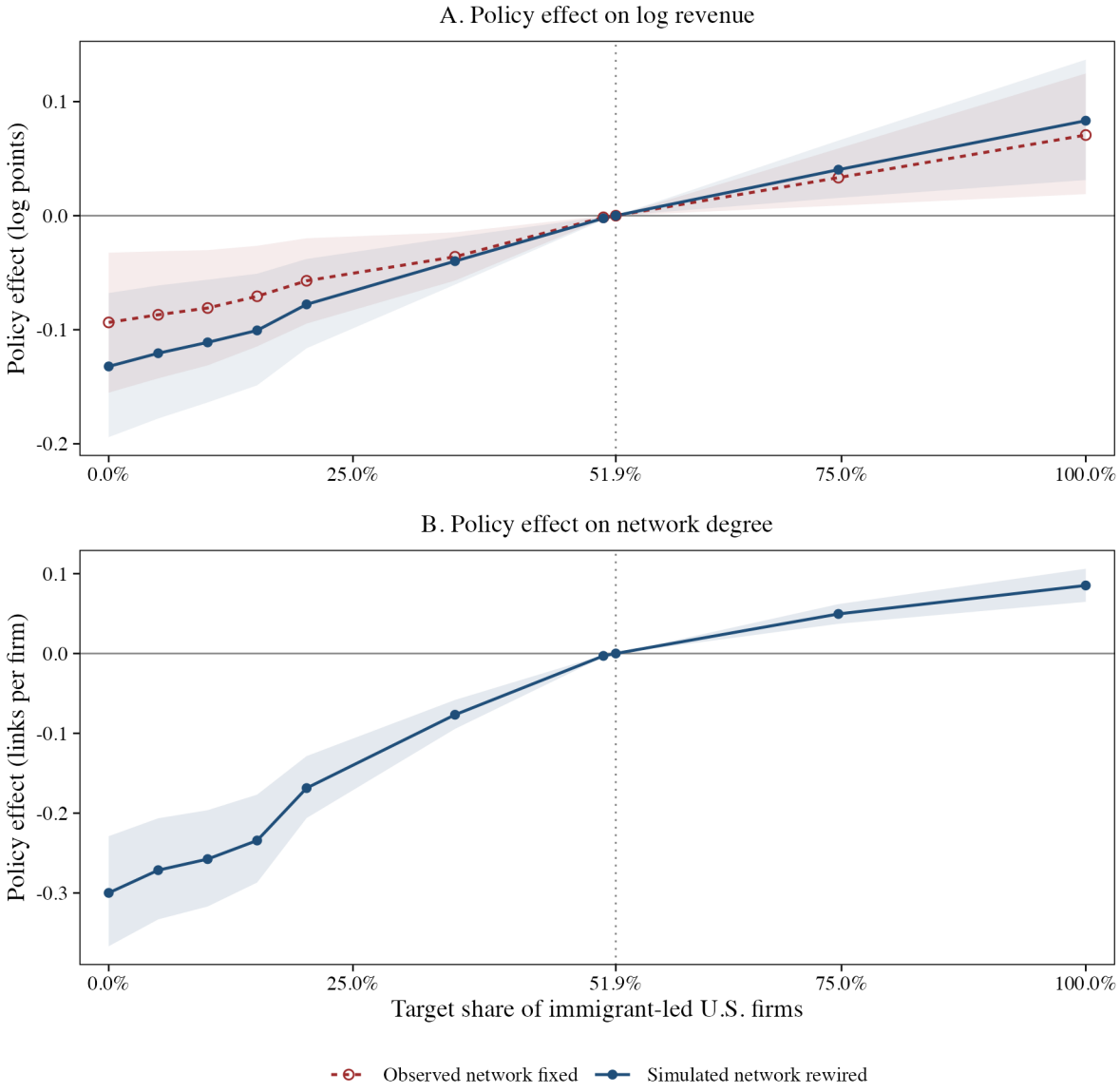


Figure 5: Posterior-draw policy spectrum under fixed observed and simulated rewired networks

Notes: Panel (a) reports average effects on U.S.-firm log revenue and Panel (b) reports effects on 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 and rebuilds $W_t Imm$ under the counterfactual assignment. 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%
Panel A. Log revenue						
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)	
Panel B. Network degree (links per firm)						
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 whose immigrant-executive status changes at that target share; they gain one above the baseline share and lose one below it. *Stayers* are the U.S. firms whose status 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.

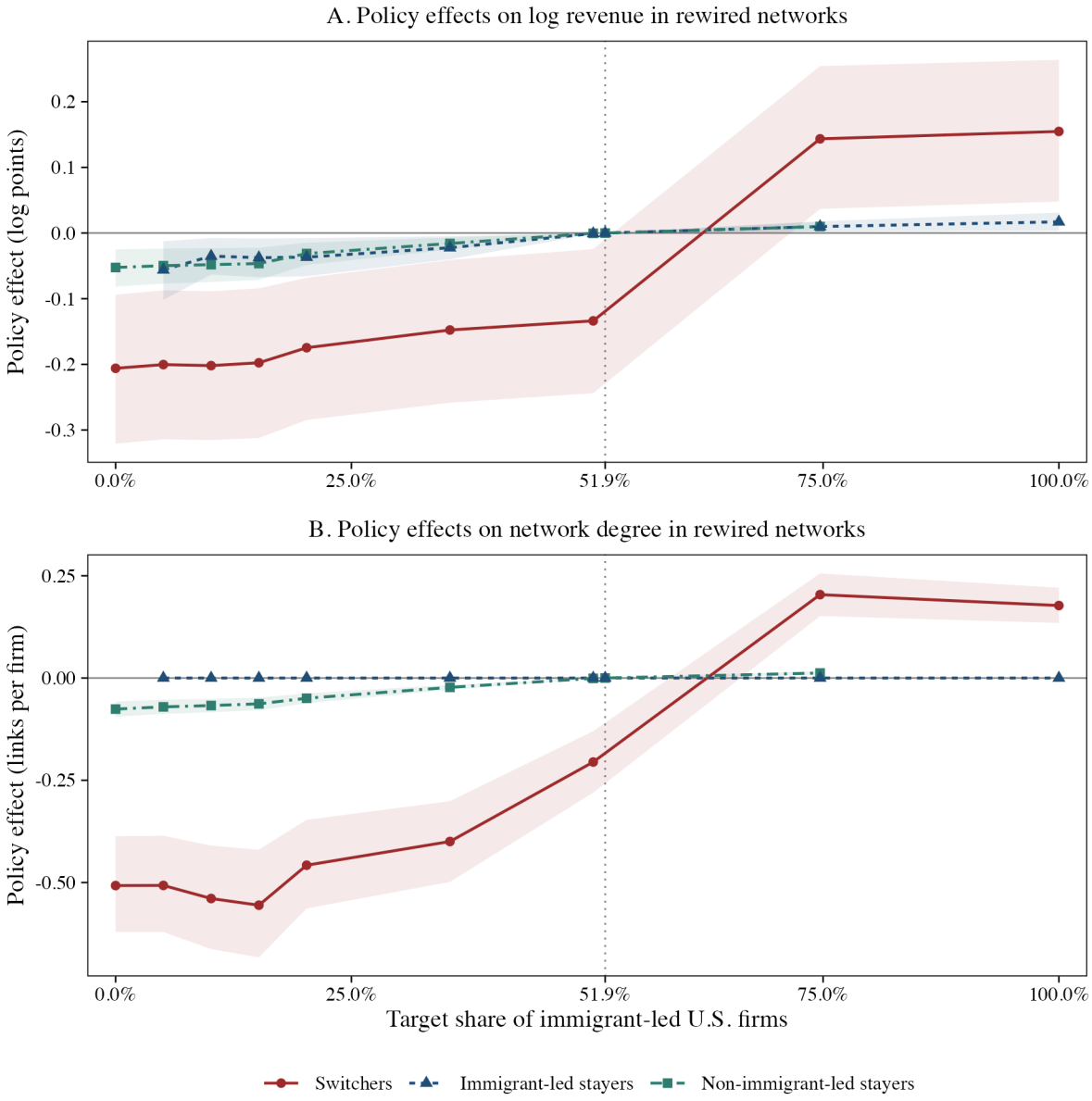


Figure 6: Posterior-draw group decomposition under simulated network rewiring

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 therefore equals the group-size-weighted average of the three group effects for every posterior draw. Credible-interval endpoints should not themselves be combined using those weights.

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.

efficient of 0.113. This ratio is a compact amplification measure, not a decomposition of the separate mechanisms. The posterior-draw results are qualitatively consistent with the posterior-mean analysis in the main text, while the latter adheres more closely to Han et al. (2021) use of fixed parameter estimates for policy simulation.

6.2 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

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

6.3 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 sender–receiver pair (i, j) with $j \neq i$ across the same T periods, since the sender's error term $\varepsilon_{it} = r_{it} - Z_{it}\boldsymbol{\beta}$ 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

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

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}:

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

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 $t = 1, \ldots, T$, 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:

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

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}:

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

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

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

4. **Combined Log-Posterior:**

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

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}:

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

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:

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

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:

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

6.7 Unbalanced Panel

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

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

For each individual i, let

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

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

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

Substituting ε_{it}:

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

Posterior of $\boldsymbol{\beta}$:

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

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

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

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

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

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

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

This can be reduced to

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

Log of density:

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

Quadratic expansion in λ

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

Since

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

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

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