Connected Trade Flows: A Spatial Econometric Approach

How Trade Networks Endogenize Trade Costs and Amplify Pair-Specific Heterogeneity

Hanbat Jeong[*] and Jieun Lee[†]

November 5, 2025

Abstract

This paper proposes a microfoundation-based econometric specification and estimation for a gravity equation of *connected* trade flows, grounded in the network-leveraging behavior of countries. By modeling trade costs as a function of network structure, we endogenize trade costs, where the conventional iceberg cost emerges as a special case when countries do not leverage their network connections. This network-leveraging behavior induces interdependence among trade flows and amplifies pair-specific heterogeneity, which we capture using a spatial autoregressive specification. For estimation, we extend the Poisson Pseudo Maximum Likelihood Estimator to account for network dependence and propose robust inference procedures that allow heteroskedasticity and arbitrary correlation in the error structure. Monte Carlo simulations demonstrate the consistency and reasonable nominal coverage of the proposed estimator.

In the empirical application, we uncover significant and evolving network effects across four key phases of global trade: Phase 1 (1986, trade liberalization), Phase 2 (1997, active NAFTA implementation), Phase 3 (2007, emergence of the China trade shock), and Phase 4 (2016, expansion of global supply chains). The results show that trade networks became increasingly interconnected, with competition dominating in early liberalization, complementarity rising under NAFTA, renewed substitution emerging during the China trade shock, and balanced interdependence under mature global value chains. A counterfactual analysis quantifies these dynamics, revealing that network effects amplified trade flows and welfare gains most strongly during NAFTA, moderated during the China trade shock, and stabilized with persistent heterogeneity under global supply chains. These findings highlight the central role of trade networks in shaping the magnitude and distribution of globalization's benefits.

Keywords: Origin-destination flows, international trade, gravity equation, network leveraging behavior, endogenous trade costs, pair-specific heterogeneity, spatial autoregressive model, Poisson pseudo-maximum likelihood estimation, spatial heteroskedasticity- and autocorrelation-consistent standard errors.
JEL codes: C13, C31, F1.

1 Introduction

Gravity equations are the workhorse for explaining bilateral trade flows. In the structural gravity framework, how we specify trade costs is crucial because it shapes both the composition of trading partners and the quantities exchanged. The iceberg-cost specification—under which a fraction of the shipped good "melts" in transit—has become

[*]Department of Economics, Macquarie Business School, Macquarie University. E-mail: *hanbat.jeong@mq.edu.au*
[†]Department of Economics, Emory University. E-mail: *jieun.lee@emory.edu.*

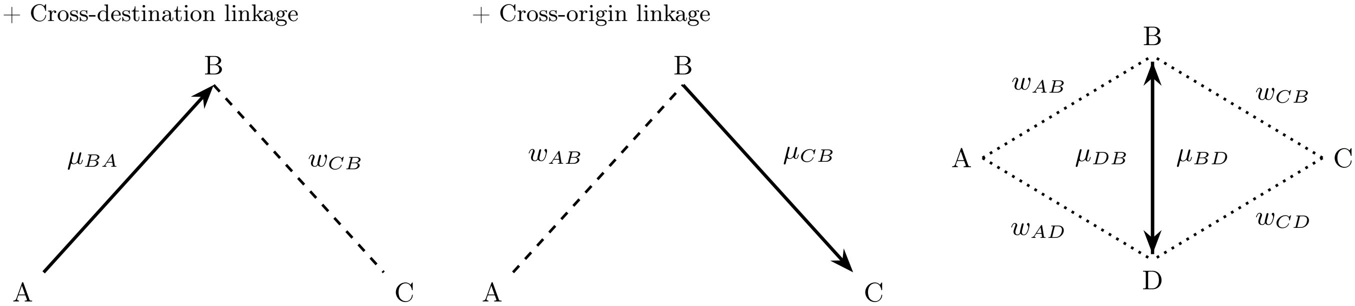
the canonical choice: it is microfounded across standard models, enters trade shares multiplicatively, and maps cleanly to identification and counterfactual analyses (e.g., multilateral resistance and hat algebra). In empirical work, distance, borders, and tariffs are used as proxies for this ad-valorem–type cost. This conventional formulation treats trade costs as largely exogenous and unavoidable, emphasizing geographic distance as the dominant proxy. Moreover, these costs are assumed to form independently across country pairs, even when some pairs (e.g., A–C or B–D) are geographically or logistically connected.

Figure 1: Three configurations that can lower the effective cost of trading from A to C

(a) Common origin + Cross-destination linkage

(b) Common destination + Cross-origin linkage

(c) Cross-origin/-destination linkages



Note: μ denotes an expected trade flow; w denotes a proximity/connectivity measure. Panels (a)–(c) illustrate three network-based channels that can reduce the effective A→C cost.

This paper revisits that conventional view by asking to what extent countries manage bilateral trade costs by exploiting information about expected third-party flows—for instance, by routing or consolidating shipments. Figure 1 sketches the core mechanisms. Consider three agents, A, B, and C. If exports from A to B are expected to be large and B is strongly linked to C (panel a), routing via B allows sharing of schedules and fixed logistics, reducing the effective A→C cost. Suppose instead that A and B are closely connected and a large B→C flow is expected (panel b); then A can bundle A→B shipments with B→C services (e.g., consolidation or backhaul), again lowering the effective cost to C. Third, expected flows among third parties (e.g., B↔D), combined with strong proximities (A–D and B–C), can make multi-leg routes (A→D→B→C or A→B→D→C) cost-effective (panel c). Conversely, if shipping capacity is constrained, large expected flows on adjacent routes (e.g., A→B, B→C, B↔D) may raise the effective cost of A→C. Hence, bilateral trade costs depend on expectations about third-party flows, generating cross-origin and cross-destination linkages that a purely pairwise iceberg view misses. These considerations motivate the development of a theoretical and econometric framework that endogenizes trade costs through network-based linkages across origins and destinations.[1]

We develop a gravity model with microfoundations in which trade costs are endogenized. To capture network interactions, we adopt the structure of a spatial autoregressive (SAR) operator. The core hypothesis is that expected trade flows with countries' connectivity shape bilateral costs, rather than costs being driven solely by geographic distance. Let y_{ij} denote the observed flow from j to i and μ_{ij} its systematic component. The model proceeds in three

[1] Hub ports such as Singapore or Dubai illustrate how third-party flows create network economies: by consolidating shipments across multiple origins and destinations, they reduce effective costs even for country pairs that are not directly linked. Similarly, in container shipping between Asia and North America, eastbound flows (Asia→US) are typically much larger than westbound flows, leading to costly empty-container repositioning. When westbound exports expand, carriers exploit backhaul opportunities, thereby lowering average eastbound costs. Analogous mechanisms appear in air cargo between Asia and Europe and in long-haul trucking, where return-leg demand reduces unit costs by sharing fixed logistics and scheduling across directions.

stages. Stage 1 characterizes the country connectivity matrix $W = (w_{ij})$ as given and long-run: each w_{ij} is a row-normalized connectivity weight indicating how strongly i is linked to j in trade-relevant proximity (e.g., geography, logistics, historical ties). Stage 2 specifies ad-valorem–type trade costs as the product of a network-driven component (a function of expected flows and W) and an exogenous component (distance measures and standard bilateral covariates). Stage 3 determines optimal trade flows via a CES demand system à la Anderson and van Wincoop (2003).

Network-driven cost component. The endogenous part of the cost function is multiplicative (ad-valorem–type) and depends on geometric averages of expected flows weighted by connectivity $\{w_{ij}\}$. Three terms capture distinct network channels: (i) outflows from a common origin (A$\to\cdot$) induce cross-destination linkages (routing, backhaul); (ii) inflows to a common destination ($\cdot\to$C) induce cross-origin linkages (consolidation toward the same market); and (iii) third-party flows (e.g., via B and D) capture hub-and-spoke and multi-leg routing economies. Given these costs from Stages 1–2, Stage 3 yields the equilibrium flows and, hence, the systematic component μ_{ij}.

We build an econometric model grounded in the equilibrium trade flows derived from our theoretical framework: μ_{ij} specifies the conditional expectation of trade flows. The econometric model aims to identify (i) network interaction parameters, which measure how countries leverage trade networks to manage costs, and (ii) elasticities of bilateral and unilateral characteristics that affect trade costs. While these mechanisms capture network-based efficiency gains that reduce effective trade costs, the same interdependence can also generate congestion effects when nearby routes or hubs operate at capacity. Accordingly, the sign of the network elasticities determines whether the underlying network amplifies efficiency or transmits capacity constraints. When the network interaction parameters are zero, the model collapses to the conventional iceberg-cost case. To show this, we first characterize equilibrium trade flows. The resulting equilibrium follows a spatial autoregressive (SAR) structure embedded in an exponential functional form. This representation is a semi-reduced form, as the fixed-effect components are implicit functions of the systematic part of trade flows—they include multilateral resistance terms, which are aggregations of countries' GDP shares weighted by endogenized trade costs. We construct the systematic component of trade flows based on this semi-reduced form, treating the implicit functions as fixed-effect terms in the econometric model.

To ensure that the econometric model is well defined, we establish conditions for the existence and uniqueness of equilibrium trade flows. The existence of the semi-reduced form follows from the invertibility of the network SAR operator, which holds when the spectral radius of W is less than one. We characterize this condition as a function of a single network statistic, the minimum eigenvalue of W. Under this condition, equilibrium trade flows can be expressed as network-weighted aggregates of exogenous bilateral characteristics and fixed-effect components, where each weight corresponds to an element of the network multiplier matrix, the inverse of the SAR operator. Because the fixed-effect components are implicit functions of expected trade flows, we impose a slightly stronger regularity condition to guarantee uniqueness. Intuitively, this condition ensures that changes in expected trade flows do not cause unstable feedback loops through the multilateral resistance terms. The uniqueness of equilibrium is essential for identifying the additional parameters associated with unilateral (origin or destination) characteristics embedded in the fixed effects. Based on the uniquely derived expected trade flows, we characterize the structure of y_{ij} as a combination of the expected trade flow μ_{ij} and error.

We consider the Poisson pseudo maximum likelihood (PPML) estimator for estimating the first stage parameters, network interaction parameters, elasticity parameters for bilateral characteristics, and fixed-effect parameters, since this method only requires a correctly specified conditional mean of y_{ij}. That is, we allow arbitrary heteroskedasticity and correlations in the error terms. Next, we characterize the identification conditions for the first-stage parameters. In contrast to the conventional PPML estimation, our model requires additional restrictions on the parameter space originating from network influences. Since we can identify/estimate the second-stage parameters (parameters for unilateral characteristics) using the identified fixed-effect components, we also provide a detailed inference on the fixed-effect parameters.

For inference, we derive the asymptotic distribution of the PPML estimator for the first-stage parameters first. We further a method for adjusting the standard errors to account for both heteroscedasticity and arbitrary correlations in the error terms. This adjustment is necessary because we only specify the conditional first moment of the error term as being one, without imposing further structures on the variance-covariance matrix but weak correlations. Hence, we establish a method to obtain heteroscedasticity-autocorrelation-consistent (HAC) standard errors. One nontrivial feature here is that we need to obtain proximities among pairs from given information about proximities among cross-section units. We utilize the maximum distance measure from origin- and destination- distances. Next, we characterize the asymptotic distribution of the estimator for the second-stage parameters. Due to less variability of univariate characteristics, the convergence rate of the estimator for the second-stage parameters is rather slower than that for the first-stage main parameters.

In our simulations, we first obtain an admissible parameter range for the network interaction parameters satisfying the equilibrium uniqueness conditions. Second, we conduct Monte Carlo simulations to know whether the PPML estimator for the first-stage parameters and the NLS estimator for the second-stage parameters perform well in finite samples. We observe that our estimators perform reasonably well. Last, we evaluate the finite sample performance of our variance estimator. Our adjusted standard errors also give reasonable nominal coverage in finite samples.

In our empirical analysis, we interpret average total trade flows over recent years as a proxy for the underlying connection structure among countries—a composite outcome shaped, in an agnostic manner, possibly by geographic proximity, economic interdependence, political relationships, etc. We utilize world trade flow data from the Center for International Data at UC Davis. For each phase, we use the following historical trade flow periods: Phase 1 (1986) uses data from 1984–1985; Phase 2 (1997) uses data from 1993–1996; Phase 3 (2007) uses data from 2000–2006; and Phase 4 (2016) uses data from 2010–2015. Based on data availability, our analysis includes a sample of 136, 142, 146, and 147 countries by phase, respectively.[2] The lists of dominant units by phase were determined based on the magnitude of dominance as defined in Pesaran and Yang (2020). The results reveal a significant presence of spillover effects in trade flows, offering evidence not only of interconnected trade relationships but, more specifically, of countries' strategic behavior in leveraging their trade networks as resources to reduce trade costs—highlighting the endogeneity of trade costs. This interconnectedness likely arises from shared sectoral dependencies and supply chain complementarities, where countries coordinate their trade strategies to achieve more efficient trade cost structures.

This network can be constructed by historical trade flows (if applicable) since historical trade flows are regarded as exogenous network connectivities across countries.

1.1 Our Contribution and Related Literature

This subsection discusses how our work relates to and contributes to the existing literature, highlighting four main aspects.

First, we contribute to the endogenization of trade costs by introducing the network-leveraging behavior of cross-sectional units, in which they utilize trade networks as a resource that shapes and governs the cost of trading across borders. This approach contrasts with traditional gravity-based trade models, which typically treat trade costs as

[2] The full list of the countries is provided in Table A.1 and A.2.

exogenous. They are often represented by the iceberg cost assumption, in which part of a good "melts away" during transport and the triangle inequality is assumed to hold. This framework has been foundational in the international trade literature, as exemplified by the gravity models of Tinbergen (1962), Anderson (1979), Helpman and Krugman (1985), Eaton and Kortum (2002), Anderson and van Wincoop (2003), and Melitz (2003). The iceberg cost specification, originating in Samuelson (1952, 1954) and later adopted by Krugman (1995) and Eaton and Kortum (2002), offers a tractable analytical framework but relies on a key abstraction: trade costs are assumed to be predetermined and symmetric.

This assumption overlooks the role of trade networks and the ways in which trade costs can emerge endogenously from their structure. In practice, the cost of trading between two countries depends not only on bilateral characteristics but also on interdependencies within the broader network of trade relationships. We move away from the iceberg-cost assumption and instead characterize trade costs as network-dependent and mutually determined. This direction is consistent with, though independent of, the perspective in Lind and Ramondo (2023), which also departs from the traditional iceberg-cost framework. Our approach enables the modeling of interdependent trade flows that evolve endogenously from the network structure rather than being imposed externally. By treating the trade network as an operational resource that shapes effective trade costs, our framework captures how connectivity, competition, and spillovers jointly influence trade frictions.

Second, our framework for interdependent trade flows builds on the spatial econometrics literature, which models interdependence across space or networks through spatial lag structures. Our specification adopts the spatial autoregressive (SAR) model (Cliff and Ord (1995); Ord (1975); Lee (2004, 2007)), where outcomes for one unit depend on those of neighboring units. Recent extensions of SAR models to origin–destination (OD) flows (Jeong et al. (2023); Jeong and Lee (2024)) and international trade applications (Behrens et al. (2012); Jin et al. (2023)) have advanced the modeling of bilateral interdependence.

However, existing SAR models for OD flows cannot directly address zero trade flows because they rely on log-linearized specifications. As noted by Santos Silva and Tenreyro (2006), log transformations can lead to biased inference in the presence of zeros, which are often treated in an ad hoc manner.[3] Consequently, these models cannot properly accommodate zero trade flows. Existing SAR models for OD flows also lack a microfoundation that explains the underlying mechanisms through which such interdependence arises.

Our model provides a microfoundation for this framework by combining an endogenized trade cost system with a consumer utility maximization problem, allowing trade costs to evolve as endogenous outcomes of the trade network itself. Because the model is derived from microfoundations, it also clarifies the structure of fixed effects, which are implicitly endogenous—that is, functions of the outcome—whereas the existing literature typically treats them as exogenous and does not specify the conditions under which equilibrium is unique. Building on this microfounded structure, we formulate the model at the original level, thereby avoiding the log-transformation issue associated with zero trade flows. Our framework thus overcomes these limitations by accommodating zero outcomes in their original form and by formally establishing the conditions for a unique equilibrium.

We also contribute to the analysis of network properties by examining the eigenvalues of the trade network matrix. Specifically, we analyze the eigenvalue spectrum to characterize the structure of network connections and use these quantities to derive the condition for the invertibility of the network multiplier matrix. Our condition requires that the spectral radius of the matrix governing the network multiplier be less than one. This requirement is both more flexible and less restrictive than the conventional condition—common in the spatial econometrics literature—that the absolute

[3]Further details are provided in Supplement Sections 1.3.1 and 1.3.2.

sum of network parameters must be less than one. We further relate this result to various network structures, such as the linear-in-means model (Manski, 1993) and bipartite networks, and discuss how the corresponding eigenvalue spectrum delineates the admissible parameter space.

Third, we contribute to introducing a new form of heterogeneity—one that is generated by the interdependence among trade flows, thereby amplifying the individual heterogeneity of cross-sectional units. The presence of spillover effects adds an additional dimension of heterogeneity, which has become a focal point in contemporary trade theory. Earlier studies have modeled heterogeneity in various forms, including productivity differences across countries (2002), variation in trade costs (2003), and firm-level productivity heterogeneity (Chaney, 2008; Helpman et al., 2008; Melitz and Ottaviano, 2008). These studies extended the classical gravity equations (Isard, 1954; Tinbergen, 1962) by introducing unit-specific heterogeneities, typically absorbed by fixed effects.

Importantly, the interdependence of trade flows gives rise to pair-specific heterogeneity, as interdependent trade flows influence the multilateral resistance components traditionally captured by fixed effects. The multilateral resistance terms themselves reflect trade frictions across all trading partners and thus constitute an integral part of individual heterogeneity. While the trade literature typically treats these multilateral resistance terms as fixed effects that control for country-specific characteristics, we argue that they are equilibrium outcomes that are also shaped by the structure of trade networks. Consequently, multilateral resistance terms are inherently interdependent. The resulting spillover-induced heterogeneity is therefore relational, manifesting at the pair level and requiring a framework capable of capturing such interdependencies beyond conventional fixed-effect specifications.

Finally, we contribute to advancing the econometric foundations of constant elasticity models. For estimation, we propose a nonlinear network interaction model for nonnegative OD flows that operates in the original level. Our model extends the Poisson Pseudo Maximum Likelihood Estimation (PPMLE) framework (Gourieroux et al. (1984)) to explicitly incorporate interdependence among trade flows and unit-specific effects, thereby capturing both spillover effects and pair-specific heterogeneity. Our estimation strategy is consistent with the gravity equation estimation literature (Santos Silva and Tenreyro (2006, 2022); Head and Mayer (2014); Nagengast and Yotov (2025); Kwon et al. (2025)) but extends it by incorporating interdependence among trade flows and the implicit fixed effects.

For inference, we incorporate recent econometric advances in high-dimensional fixed effects and spatial dependence (Kapetanios et al. (2021); Weidner and Zylkin (2021); Fernandez-Val and Weidner (2016, 2018); Chen et al. (2021)). We adopt heteroskedasticity and autocorrelation-consistent (HAC) methods (Newey and West, 1987; Andrews, 1991) based on kernel-based truncation that leverages proximity information among cross-sectional units. In particular, we employ heteroskedasticity and autocorrelation-robust (HAR) methods for spatial data (Kelejian and Prucha (2007); Kim and Sun (2011); Conley et al. (2023)), which ensure consistency and valid inference under network-induced dependence.

The rest of the paper is organized as follows. In Section 2, we present the microfoundations of our model alongside the econometric motivations. We then introduce an econometric specification and estimation strategy tailored for nonlinear models, addressing key challenges and offering methodological insights. Section 3 provides the statistical analysis, including the asymptotic properties of our estimator and simulation evidence, illustrating how partially specified models can lead to invalid inferences. In Section 4, we apply our framework to empirical data on world trade flows from the Center for International Data at UC Davis. Section 5 concludes.

2 Model

Our main interest is to explain the spillover structure of a spatial origin-destination (OD) flow y_{ij} for $i, j = 1, \cdots, n$, which denotes a directed outcome generated from two locations: i (destination) and j (origin). Suppose we have $N = n^2$ flow observations, where we represent this data structure as an $n \times n$ matrix Y or $\mathbf{y} = \text{vec}(Y)$.[4] For its indexes, we define a pair of cross-sectional units $ij \equiv (i, j)$ to indicate a case originating from j and destined for i. The assumption below introduces the location setting for cross-section units outlined by Jenish and Prucha (2009, 2012).

Assumption 2.1. Each $i \in \{1, \cdots, n\}$ is located in a d-dimensional space $\mathcal{D}_n \subset \mathcal{D}$, where $\mathcal{D}$ denotes a set of all potential locations in $\mathbb{R}^d$. We assume $\lim_{n\to\infty} \#(\mathcal{D}_n) = \infty$ and $\min_{i\neq j} d(l(i), l(j)) \geq 1$, where $\#(\mathcal{D}_n)$ is the cardinality of $\mathcal{D}_n$, $l : i \mapsto l(i) \in \mathcal{D}$ stands for an injective location function, and $d(l(i), l(j))$ is a distance between i and j.

Assumption 2.1 means increasing domain asymptotic widely used in the spatial econometric literature (Qu and Lee, 2015; Xu and Lee, 2015a,b, 2018; Jeong and Lee, 2024). The first feature of Assumption 2.1 is that the space of locations in the sample ($\mathcal{D}_n$) expands as n grows. Beyond the geographic space, the concept of $\mathcal{D}_n$ and $\mathcal{D}$ can be extended to a characteristic space accommodating regions' economic and political locations. Second, the minimum distance assumption is intended to avoid extremely densely located cross-section units.

Traditional spatial autoregressive (SAR) models (Cliff and Ord, 1995; Ord, 1975; Lee, 2004, 2007) treat observations (y_i, x_i) collected across space (Assumption 2.1), where y_i denotes an i's outcome and x_i is a vector of regressors. The SAR model specifies how y_i's are interrelated: for $i = 1, \cdots, n$,

$$y_i = \lambda \sum_{j=1}^{n} w_{ij} y_j + x_i' \beta + v_i, \tag{1}$$

where λ and β are the model's parameters, w_{ij} represents a strength of i being influenced by j, and v_i is an error. Then, an $n \times n$ spatial weighting matrix $W = (w_{ij})$ characterizes the relationships among n units. When $S \equiv I_n - \lambda W$ is invertible, (1) can be represented by

$$y_i = \sum_{j=1}^{n} (S^{-1})_{ij} (x_j \beta + v_j) = \sum_{j=1}^{n} (I_n + \lambda W + \lambda^2 W^2 + \cdots)_{ij} (x_j \beta + v_j). \tag{2}$$

Hence, $\mathbb{E}(y_i|x_1, \cdots, x_n) = \sum_{j=1}^{n} (S^{-1})_{ij} x_j \beta$ and $\mathbb{V}(y_i|x_1, \cdots, x_n) = \sigma^2 \left(S^{-1} S^{-1\prime}\right)_{ii}$ if we assume $\mathbb{E}(v_i|x_1, \cdots, x_n) = 0$ and $\mathbb{V}(v_i|x_1, \cdots, x_n) = \sigma^2 > 0$ for all $i = 1, \cdots, n$. Since equation (2) is a unique solution to the contraction mapping (1), the SAR model represents how outcomes y_is are formed by consensus. In consequence, this model captures regional heterogeneities and spillovers stemming from their interconnectivities.

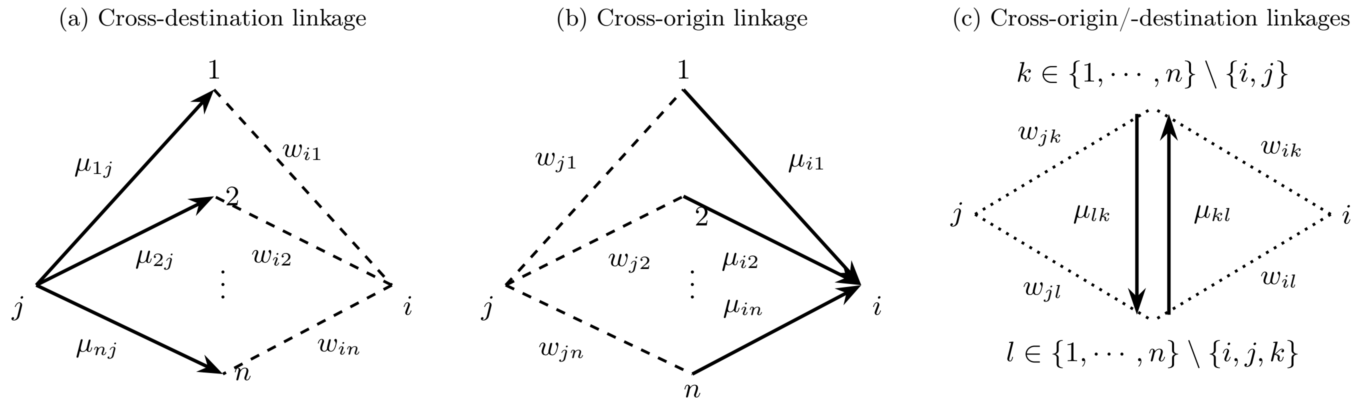
2.1 Theoretical foundation

Building on Anderson and van Wincoop (2003), we provide microfoundations for a new model of spatial origin–destination (OD) flows. Our goal is to specify μ_{ij}—the systematic component of y_{ij}—by endogenizing trade costs. We assume $\mu_{ij} > 0$ for all $i, j \in \{1, \ldots, n\}$.

[4]This notation scheme is called the *destination centric ordering* (see LeSage and Pace (2008)). We use this scheme since this is consistent with (i) a matrix-based notation (i.e., y_{ij} is the (i, j)-element of Y) and (ii) spatial/network econometric interpretations. According to spatial/network econometric literature, Y can be interpreted as a directed weighted network, and each y_{ij} denotes a signal from j to i. In the trade literature, Eaton and Kortum (2002) and Head and Mayer (2014) utilize this scheme.

We now formalize the model in three stages.

Figure 2: Formation of key trade partners



Stage 1: When country j ships its products to i, for each trade opportunity $\nu = 1, \cdots, M$ (i.e., M opportunities), country $k \in \{1, \cdots, n\} \setminus \{i\}$ is selected as a partner of destination country i with probability $w_{ik}^d(\nu)$ and country $l \in \{1, \cdots, n\} \setminus \{j\}$ is chosen as a partner of origin country j with probability $w_{jl}^o(\nu)$. In the first stage equilibrium,

$$w_{ik}^d = \lim_{M \to \infty} \frac{\sum_{\nu=1}^M \mathbb{I}\{k \text{ is chosen as } i\text{'s partner for opportunity } \nu\}}{M} \text{ for } k \in \{1, \cdots, n\} \setminus \{i\},$$
$$w_{jl}^o = \lim_{M \to \infty} \frac{\sum_{\nu=1}^M \mathbb{I}\{l \text{ is chosen as } j\text{'s partner for opportunity } \nu\}}{M} \text{ for } l \in \{1, \cdots, n\} \setminus \{j\},$$

and

$$w_{ij} = w_{ij}^d = w_{ij}^o \text{ for all } i \neq j. \tag{3}$$

Assumption 2.2. Let $W = (w_{ij})$ be an $n \times n$ country connectivity matrix. For all $i = 1, \cdots, n$, $\sum_{j=1}^n w_{ij} = 1$ and $w_{ii} = 0$. Additionally, we assume that W is constructed from symmetric geographic/logistical relationships (while W itself can be asymmetric). That is, $W = \text{Diag}^{\text{sum}}(\widetilde{W})^{-1}\widetilde{W}$ with $\widetilde{W}' = \widetilde{W}$, where $\widetilde{w}_{ij}$ is the (i, j)-element of $\widetilde{W}$ and $\text{Diag}^{\text{sum}}(\widetilde{W}) = \text{diag}\left(\sum_{j=1}^n \widetilde{w}_{1j}, \cdots, \sum_{j=1}^n \widetilde{w}_{nj}\right)$.

Stage 2: Let π_{ij} be a measure of a trade cost from j to i. We posit the following cost function that endogenizes trade costs.

Assumption 2.3. (i) We assume

$$\pi_{ij}(\boldsymbol{\mu}) = \pi_{ij}^e(\boldsymbol{\mu}) \cdot \pi_{ij}^+, \tag{4}$$

where

$$\pi_{ij}^e(\boldsymbol{\mu}) = \left(\overline{\mu}_{\cdot j}^i\right)^{-\tilde{\lambda}_d} \left(\overline{\mu}_{i\cdot}^j\right)^{-\tilde{\lambda}_o} \left(\overline{\mu}_{\cdot\cdot}^{ij}\right)^{-\tilde{\lambda}_w}, \text{ and}$$
$$\pi_{ij}^+ = D_{ij,1}^{\tilde{\beta}_1} \cdots D_{ij,K}^{\tilde{\beta}_K} \cdot E_{i,1}^{\tilde{\gamma}_1^d} \cdots E_{i,L}^{\tilde{\gamma}_L^d} \cdot E_{j,1}^{\tilde{\gamma}_1^o} \cdots E_{j,L}^{\tilde{\gamma}_L^o}$$

with $\boldsymbol{\mu} = (\mu_{11}, \mu_{21}, \cdots, \mu_{n1}, \cdots, \mu_{1n}, \mu_{2n}, \cdots, \mu_{nn})'$. Here,

- $\overline{\mu}^{i}_{\cdot j} = \prod_{k=1}^{n} \mu_{kj}^{w_{ik}}$ is the (geometric) average of outflows from j; $\overline{\mu}^{j}_{i\cdot} = \prod_{l=1}^{n} \mu_{il}^{w_{jl}}$ denotes the average of inflows to i; and $\overline{\mu}^{ij}_{\cdot\cdot} = \prod_{k,l=1}^{n} \mu_{kl}^{w_{ik} w_{jl}}$ represents the average of flows among third-party units with coefficients $\tilde{\lambda}_d$, $\tilde{\lambda}_o$ and $\tilde{\lambda}_w$; and
- $D_{ij,k}$ $(k = 1, \cdots, K)$ represents a bilateral characteristic affecting π_{ij}, and $E_{i,l}$ (and $E_{j,l}$) is a destination-specific (origin-specific) characteristic. $\tilde{\beta}_1, \cdots, \tilde{\beta}_K, \tilde{\gamma}^d_1, \cdots, \tilde{\gamma}^d_L, \tilde{\gamma}^o_1, \cdots$, and $\tilde{\gamma}^o_L$ are the elasticity parameters.

(ii) If $\tilde{\lambda}_d = \tilde{\lambda}_o = \tilde{\lambda}_w = 0$, π_{ij} satisfies the triangle inequality: for arbitrary three countries i, j, and k, $\pi_{ij} \leq \pi_{ik} \cdot \pi_{kj}$.

(iii) All stages—**Stage 1**, **Stage 2**, and **Stage 3**—operate under no informational frictions, meaning that the expected flows in **Stage 2** are the same as the optimal flows in **Stage 3**, $\boldsymbol{\mu}^* = (\mu^*_{11}, \mu^*_{21}, \cdots, \mu^*_{n1}, \cdots, \mu^*_{1n}, \mu^*_{2n}, \cdots, \mu^*_{nn})'$.

Under these settings, **Stage 2** forms $\pi_{ij}(\boldsymbol{\mu}^*)$ for $i, j = 1, \cdots, n$.

Assumption 2.2 illustrates properties of a resulting country connectivity matrix from **Stage 1**. Each element w_{ij} can be interpreted as the choice probability of j from i's aspect. This is consistent with standard assumptions on spatial weighting matrices in spatial econometrics. The symmetry of $\widetilde{W}$ is adopted for computational convenience and can be relaxed. Throughout, W is row-normalized and allows for directional (asymmetric) linkages even when the underlying proximity matrix $\widetilde{W}$ is symmetric. Assumption 2.3 describes how trade costs are endogenously shaped by information in trade networks. The three panels in Figure 2 correspond directly to the three network-based terms in equation (4). Unlike conventional models with exogenous π_{ij}, here trade costs are themselves functions of the network of expected flows, capturing spillovers across origins and destinations. This provides a new channel of spillovers not captured in conventional frameworks such as Anderson and van Wincoop (2003). Given $\pi_{ij}(\boldsymbol{\mu})$, **Stage 3** specifies how trade flows are determined.

The component $\pi^e_{ij}(\boldsymbol{\mu})$ within $\pi_{ij}(\boldsymbol{\mu})$ captures the network-based part of trade costs:

1. Outflows from j ($\overline{\mu}^{i}_{\cdot j}$) summarize the common-origin with cross-destination linkages;
2. Inflows to i ($\overline{\mu}^{j}_{i\cdot}$) summarize the common-destination with cross-origin linkages; and
3. Third-party flows ($\overline{\mu}^{ij}_{\cdot\cdot}$), which do not share the same origin or destination, influence costs through cross-origin and cross-destination linkages.[5]

Note that $w_{i1}, \ldots, w_{in}$ from **Stage 1** in $\overline{\mu}^{i}_{\cdot j}$ capture cross-destination linkages; $w_{j1}, \ldots, w_{jn}$ in $\overline{\mu}^{j}_{i\cdot}$ capture cross-origin linkages; and both sets of weights appear in $\overline{\mu}^{ij}_{\cdot\cdot}$ to capture combined cross-destination and cross-origin linkages. $\tilde{\lambda}_d > 0$ implies network-based efficiency gains, reducing effective trade costs by the cross-destination linkages. On the other hand, $\tilde{\lambda}_d < 0$ means congestion effects from the cross-destination linkages if nearby routes or hubs operate at capacities. Analogously, $\tilde{\lambda}_o$ governs cross-origin linkages (e.g., consolidation vs. destination-side bottlenecks), and $\tilde{\lambda}_w$ captures third-party routing (hub-and-spoke economies vs. congestion on multi-leg paths).

Our specification nests the conventional iceberg cost specification since $\pi_{ij}(\boldsymbol{\mu}) = \pi^+_{ij}$ when $\tilde{\lambda}_d = \tilde{\lambda}_o = \tilde{\lambda}_w = 0$. The conventional iceberg assumption treats trade costs as purely wasteful losses that scale mechanically with geography. In this case, cross-border arbitrage satisfies the triangle inequality, highlighting the effectiveness of geographic barriers (e.g., Eaton and Kortum (2002)). By contrast, in our general case, the triangle inequality need not hold, implying that countries can achieve effective cost reductions based on the expected trade network.

[5] An important element of this third category is the reverse flow μ_{ji}, weighted by w_{ij} and w_{ji}.

Stage 3: Given π_{ij} from **Stage 2**, a country i's consumer chooses $\{c_{i1}, \cdots, c_{in}\}$ by the following problem:

$$\max_{\{c_{ij}\}_{j=1}^n} U_i = \left(\sum_{j=1}^n \chi_j^{\frac{1}{\varrho}} \cdot c_{ij}^{\frac{\varrho-1}{\varrho}}\right)^{\frac{\varrho}{\varrho-1}} \quad \text{subject to} \quad \sum_{j=1}^n \pi_{ij} c_{ij} = G_i, \tag{5}$$

where χ_j denotes a preference parameter, and G_i stands for the exogenously given country i's nominal expenditure (GDP). Here, (5) is a standard CES aggregator and elasticity $\varrho > 1$.[6]

The following conditions characterize the uniqueness of the nominal value of the optimal trade flow $\mu_{ij}^* = \pi_{ij}(\boldsymbol{\mu}^*)c_{ij}^*$, where c_{ij}^* $(j = 1, \cdots, n)$ denotes the solution of (5). For this, we define $\mathbf{S} = I_N - \mathbf{A}$ where $\mathbf{A} = \lambda_d(I_n \otimes W) + \lambda_o(W \otimes I_n) + \lambda_w(W \otimes W)$ with $\lambda_d = (\varrho-1)\tilde{\lambda}_d$, $\lambda_o = (\varrho-1)\tilde{\lambda}_o$, $\lambda_w = (\varrho-1)\tilde{\lambda}_w$. Here, $I_n \otimes W$, $W \otimes I_n$, and $W \otimes W$ characterize respectively the cross-destination, cross-origin, and cross-origin and destination linkages.

Assumption 2.4. (i) We assume

$$\max\{\lambda_d + \lambda_o + \lambda_w, \lambda_d\varphi_{\min} + \lambda_o + \lambda_w\varphi_{\min}, \lambda_d + \lambda_o\varphi_{\min} + \lambda_w\varphi_{\min}, \lambda_d\varphi_{\min} + \lambda_o\varphi_{\min} + \lambda_w\varphi_{\min}^2\} < 1, \tag{6}$$

where $\varphi_{\min}$ is the minimum eigenvalue of W. Then, $\mathbf{S}$ is invertible.

(ii) $\boldsymbol{\mu}^*$ satisfies the following condition:

$$\sup_{i,j} \sum_{k,l=1}^n \left| \sum_{p,q=1}^n s_{ij,pq} \left(\frac{\partial\left(\alpha_q(\boldsymbol{\mu}) + \eta_p(\boldsymbol{\mu})\right)}{\partial \ln(\mu_{kl})} \right) \right| < 1,$$

where $s_{ij,kl}$ denotes the $((j-1)n + i, (l-1)n + k)$-element of $\mathbf{S}^{-1}$. Further,

$$\begin{aligned} \alpha_q(\boldsymbol{\mu}) &= -\frac{1}{2}\ln(G^W) + \ln(G_q) + \ln\left(\Pi_q^{\varrho-1}(\boldsymbol{\mu})\right) + x_q^u\gamma_o \text{ for } q = 1, \cdots, n \text{ and} \\ \eta_p(\boldsymbol{\mu}) &= -\frac{1}{2}\ln(G^W) + \ln(G_p) + \ln\left(P_p^{\varrho-1}(\boldsymbol{\mu})\right) + x_p^u\gamma_d, \text{ for } p = 1, \cdots, n, \end{aligned} \tag{7}$$

where $\Pi_q(\boldsymbol{\mu}) = \left(\sum_{r=1}^n \left(\frac{G_r}{G^W}\right)\left(\frac{\pi_{rq}(\boldsymbol{\mu})}{P_r(\boldsymbol{\mu})}\right)^{1-\varrho}\right)^{\frac{1}{1-\varrho}}$ $(q = 1, \cdots, n)$, $P_p(\boldsymbol{\mu}) = \left(\sum_{s=1}^n \left(\frac{G_s}{G^W}\right)\left(\frac{\pi_{ps}(\boldsymbol{\mu})}{\Pi_s(\boldsymbol{\mu})}\right)^{1-\varrho}\right)^{\frac{1}{1-\varrho}}$ $(p = 1, \cdots, n)$, $G^W = \sum_{s=1}^n G_s$, $x_q^u = (\ln(E_{q,1}), \cdots, \ln(E_{q,L}))'$, $\gamma_o = (\gamma_{o,1}, \cdots, \gamma_{o,L})'$ and $\gamma_d = (\gamma_{d,1}, \cdots, \gamma_{d,L})'$ with $\gamma_{o,l} = (1-\varrho)\tilde{\gamma}_{o,l}$ and $\gamma_{d,l} = (1-\varrho)\tilde{\gamma}_{d,l}$ for $l = 1, \cdots, L$.

Under conditions (i) and (ii) in Assumption 2.4, there is a unique $\boldsymbol{\mu}^*$ satisfying

$$\mu_{ij}^* = \exp\left(\sum_{k,l=1}^n s_{ij,kl}\left(x_{kl}'\beta + \alpha_l(\boldsymbol{\mu}^*) + \eta_k(\boldsymbol{\mu}^*)\right)\right), \text{ for } i, j = 1, \cdots, n, \tag{8}$$

where $x_{kl} = (\ln(D_{kl,1}), \cdots, \ln(D_{kl,K}))'$ and $\beta = (\beta_1, \cdots, \beta_K)'$ with $\beta_k = (1-\varrho)\tilde{\beta}_k$. Under the no information friction setting, the expected flows coincide with the optimized flows, i.e., $\boldsymbol{\mu}^*$. Equation (8) is a semi-reduced form and establishes the main econometric equation by regarding $\alpha_l(\boldsymbol{\mu}^*)$ and $\eta_k(\boldsymbol{\mu}^*)$ as fixed-effect components. Note that

[6] We adopt a demand-side focus to highlight how network-leveraged trade costs shape flows, taking the production side as exogenous. Microfounding network dependence on the production side (e.g., extending Eaton and Kortum (2002)) would require an alternative specification (e.g., Lind and Ramondo (2023)); we leave this for future research. See Appendix 1.3.3 for related discussion.

equation (8) is not a full reduced form since the fixed-effect components ($\alpha_l(\boldsymbol{\mu})$ and $\eta_k(\boldsymbol{\mu})$) still rely on $\boldsymbol{\mu}$. See Appendix 1.3.2 for details.

Assumption 2.4 (i) ensures the well-definedness of $\{s_{ij,kl}\}$, and then $\ln(\mu^*_{ij})$ is a weighted aggregation of $x'_{kl}\beta + \alpha_l(\boldsymbol{\mu}^*)+\eta_k(\boldsymbol{\mu}^*)$. This condition is equivalent that $\rho_{\text{spec}}(\mathbf{A}) < 1$, where $\rho_{\text{spec}}(\mathbf{A})$ denotes the spectral radius of $\mathbf{A}$.[7] Note that the minimum eigenvalue of W ($\varphi_{\min}$) is a key network statistic showing bipartiteness (if $\varphi_{\min} \to -1$) and averaging rate (if $\varphi_{\min} \to 0$) (see Chung (1997); Bramoullé et al. (2014) for more details). This structure originates from the SAR framework (see LeSage and Pace (2008); Section C.3.2 of LeSage and Fischer (2010); Jeong et al. (2023); and Jeong and Lee (2024)).[8] However, while our formulation borrows from SAR models, its interpretation differs fundamentally: the expected trade flows here influence costs, rather than directly determining flows. The importance of $\varphi_{\min}$ is also highlighted in our framework, even though it is not trivial since $\mathbf{A}$ consists of multiple Kronecker product matrices with parameters as an additive form (see Section **??** of the supplement file for derivation). From the network SAR operator $\mathbf{S} = I_N - \mathbf{A}$, $\mathbf{S}^{-1}$ (network multiplier matrix) characterizes how network spillovers propagate across regions. Each element of $\mathbf{S}^{-1}$, $s_{ij,kl}$, represents the influence from pair kl to ij: the diagonal elements, $s_{ij,ij}$, might be larger than 1 (in detail, $s_{ij,ij} - 1$ presents the magnitude of the feedback effect), while an off-diagonal element ($s_{ij,kl}$ for $i \neq k$ or $j \neq l$) characterizes the effect involving third-party countries.[9] When $\lambda_d = \lambda_o = \lambda_w = 0$, (8) becomes the conventional gravity equation (e.g., (Tinbergen, 1962; McCallum, 1995; Eaton and Kortum, 2002; Anderson and van Wincoop, 2003)).

By Assumption 2.4 (ii), equation (8) becomes a contraction mapping leading to the unique forms of $\alpha_l(\boldsymbol{\mu}^*)$ ($l = 1, \cdots, n$) and $\eta_k(\boldsymbol{\mu}^*)$ ($k = 1, \cdots, n$). The main implication of Assumption 2.4 (ii) is that the cumulative network influence should not change the fixed-effect components too much. In detail, this condition ensures that cumulative network effects do not excessively perturb the multilateral resistance terms.[10] Consequently, the resulting μ^*_{ij} is a unique function of the exogenous factors $\{x_{kl}\}_{k,l=1}^n$ and $\{x_k^u\}_{k=1}^n$.

2.2 Econometric model specification

This subsection introduces our econometric model based on the theoretical foundation (Sec. 2.1). By formulating the conditional expectation function (CEF) of y_{ij}, the true data-generating process (DGP) can be specified by

$$y_{ij} = \mu^0_{ij} \times \xi_{ij}, \text{ where } \mu^0_{ij} = \exp\left(\sum_{k,l=1}^n s_{ij,kl}\left(x'_{kl}\beta^0 + \alpha^0_l + \eta^0_k\right)\right). \tag{9}$$

Here, $\mu^0_{ij} = \mathbb{E}(y_{ij}|\mathbf{z})$; $\mathbf{z} = (x'_{11}, \cdots, x'_{n1}, \cdots, x'_{1n}, \cdots, x'_{nn}, x_1^{u\prime}, \cdots, x_n^{u\prime})'$ stands for a vector of exogenous characteristics; ξ_{ij} is a multiplicative CEF disturbance satisfying $\mathbb{E}(\xi_{ij}|\mathbf{z}) = 1$, $\lambda^0 = \left(\lambda^0_d, \lambda^0_o, \lambda^0_w\right)'$ denotes a vector of the true spatial influence parameters, $\beta^0 = \left(\beta^0_1, \cdots, \beta^0_K\right)'$ is the true parameter for x_{kl}, α^0_j and η^0_i represent the true origin- and destination- fixed effects, respectively. Importantly, the advantage of this specification is that a practitioner only needs

[7] $\rho_{\text{spec}}(\mathbf{A}) < 1$ is more general condition for invertibility of $\mathbf{S}$ when W has complex eigenvalues. Since we consider a row-normalized W from a symmetric relationship, all eigenvalues are real.

[8] In detail, a linear SAR model specifies the log-transformed equation based on (8) by replacing μ^*_{ij} by y_{ij}.

[9] The detailed structures of $s_{ij,kl}$ and their interpretations can be found in Section **??** of the supplement file.

[10] To see this, note that Assumption 2.4 (ii) can also be represented by

$$\sup_{i,j} \sum_{k=1}^n \sum_{l=1}^n \left| \sum_{p=1}^n \sum_{q=1}^n s_{ij,pq} \left(\frac{\partial \Pi_q^{\varrho-1}(\boldsymbol{\mu})}{\partial \mu_{kl}} \frac{\mu_{kl}}{\Pi_q^{\varrho-1}(\boldsymbol{\mu})} + \frac{\partial P_p^{\varrho-1}(\boldsymbol{\mu})}{\partial \mu_{kl}} \frac{\mu_{kl}}{P_p^{\varrho-1}(\boldsymbol{\mu})} \right) \right| < 1.$$

$\mathbb{E}(\xi_{ij}|\mathbf{z}) = 1$ for estimation. For analytic simplicity, we can define the additive error $u_{ij} = \mu_{ij}^0 (\xi_{ij} - 1)$ for each ij to have $y_{ij} = \mu_{ij}^0 + u_{ij}$ and $\mathbb{E}(u_{ij}|\mathbf{z}) = 0$.

Equation (9) addresses three essential issues. First, equation (9) correctly specifies $\mathbb{E}(y_{ij}|\mathbf{z})$. The logarithmic transformation approach, as LeSage and Pace (2008), encounters several significant issues. As Santos Silva and Tenreyro (2006) emphasize, applying a logarithmic transformation in constant elasticity models can lead to inconsistent estimates. This inconsistency arises because the transformation specifies $\mathbb{E}(\ln(y_{ij})|\mathbf{z})$ instead of $\mathbb{E}(y_{ij}|\mathbf{z})$, and due to Jensen's inequality, $\mathbb{E}(\ln(y_{ij})|\mathbf{z}) \neq \ln(\mathbb{E}(y_{ij}|\mathbf{z}))$. The gap, $\mathbb{E}(\ln(y_{ij})|\mathbf{z}) - \ln(\mathbb{E}(y_{ij}|\mathbf{z})) = \mathbb{E}(\ln(\xi_{ij})|\mathbf{z})$, characterizing the bias from the log-transformed model increases when (i) some y_{ij}s take huge positive values or (ii) many zero OD flows are contained in a sample. Even though the original purpose of the model (9) is to estimate the constant elasticities, the bias from the logarithmic transformation model is highly dependent on the unit of the outcome. Detailed analysis can be found in Sec.1.1.1 of the supplement file.[11]

Second, ad hoc transformations of y_{ij} become necessary because a logarithmic transformed equation only accommodates strictly positive outcomes. Considering the undefined nature of $\ln(0)$, practitioners often use $\ln(y_{ij} + c)$ where $c > 0$, commonly setting $c = 1$. However, as Chen and Roth (2024) and Mullahy and Norton (2024) emphasize, this choice of c lacks a theoretical basis and leads to variations in the specification of $\mathbb{E}(\ln(y_{ij} + c)|\mathbf{z})$, which can result in biases and misleading interpretations of a covariate's marginal effect. We verify that the magnitude of the bias from employing $\ln(y_{ij} + c)$ becomes larger when the frequency of values close to zero grows. Similar issues arise with other transformations, such as the inverse hyperbolic sine and power transformations.

The third issue arises from the specific feature of spatial econometric models. Let v_{ij} be the additive error term in the logarithmic transformed model (LeSage and Pace (2008)). To illustrate how a distributional assumption on $\{v_{ij}\}$ affects $\mathbb{E}(y_{ij}|\mathbf{z})$, consider two scenarios based on the distribution of v_{ij}. By assuming $y_{ij} > 0$ for all ij, the LeSage and Pace's (2008) model can be rewritten as

$$y_{ij} = \exp\left(\sum_{k,l=1}^{n} s_{ij,kl} x_{kl}' \beta\right) \prod_{k,l=1}^{n} \exp(v_{kl})^{s_{ij,kl}}.$$

1. If $v_{ij}|\mathbf{z} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$ as LeSage and Pace (2008),

$$\mathbb{E}(y_{ij}|\mathbf{z}) = \exp\left(\sum_{k,l=1}^{n} s_{ij,kl} x_{kl}' \beta\right) \exp\left(\frac{\sigma^2}{2} \sum_{k,l=1}^{n} s_{ij,kl}^2\right) \tag{10}$$

   since $\mathbb{E}(\exp(v_{kl})^{s_{ij,kl}}|\mathbf{z}) = \exp(\frac{\sigma^2 s_{ij,kl}^2}{2})$,

2. If $v_{ij}|\mathbf{z} \overset{i.i.d.}{\sim} \text{logGamma}(\theta_{\text{shape}}, \theta_{\text{rate}})$ with $\frac{\theta_{\text{shape}}}{\theta_{\text{rate}}} = 1$,

$$\mathbb{E}(y_{ij}|\mathbf{z}) = \exp\left(\sum_{k,l=1}^{n} s_{ij,kl} x_{kl}' \beta\right) \frac{\theta_{\text{shape}}^{\sum_{k,l=1}^{n} s_{ij,kl}} \prod_{k,l=1}^{n} \Gamma(\theta_{\text{shape}} + s_{ij,kl})}{\Gamma(\theta_{\text{shape}})^{n^2}} \tag{11}$$

[11] In addition to the issue $\mathbb{E}(\ln(\xi_{ij})|\mathbf{z}) \neq 0$, various dependence structures between ξ_{ij} and a component in $\mathbf{z}$ can also affect the magnitude of the bias from the log-transformed model. Since $\ln(\xi_{ij})$ can be expressed by Maclaurin series expansion, $\mathbb{E}(\ln(\xi_{ij})|\mathbf{z})$ is a function of the infinite-order of the moments, $h_p(\mathbf{z}) \equiv \mathbb{E}\left((\xi_{ij}^-)^p|\mathbf{z}\right)$ for $p = 2, 3, \cdots$ where $\xi_{ij}^- = \xi_{ij} - 1$. Then, a possible moment for estimating the log-transformed model $\mathbb{E}(x_{ij} \ln(\xi_{ij}))$ depends on the correlations between x_{ij} and $h_p(\mathbf{z})$ for $p = 2, 3, \cdots$.

since $\mathbb{E}(\exp(v_{kl})^{s_{ij,kl}}|\mathbf{z}) = \frac{\Gamma(\theta_{\text{shape}}+s_{ij,kl})}{\Gamma(\theta_{\text{shape}})\theta_{\text{rate}}^{s_{ij,kl}}} = \frac{\Gamma(\theta_{\text{shape}}+s_{ij,kl})\theta_{\text{shape}}^{s_{ij,kl}}}{\Gamma(\theta_{\text{shape}})}$.

Hence, $\mathbb{E}(y_{ij}|\mathbf{z})$ of the LeSage and Pace's (2008) model involves $\mathbb{E}(\exp(v_{kl})^{s_{ij,kl}}|\mathbf{z})$ for every kl, which are relevant to covariance structures of $\{v_{kl}\}$ even if there is no theoretical background on the distributional assumption about $\{v_{kl}\}$.

2.3 Estimation

This subsection describes an estimation method for the parameters in (9), $\boldsymbol{\theta}^0 = (\theta^{0\prime}, \boldsymbol{\phi}^{0\prime})'$, $\theta^0 = (\lambda^{0\prime}, \beta^{0\prime})'$, $\boldsymbol{\phi}^0 = (\boldsymbol{\alpha}^{0\prime}, \boldsymbol{\eta}^{0\prime})'$, where $\boldsymbol{\alpha}^0 = (\alpha_1^0, \cdots, \alpha_n^0)'$ and $\boldsymbol{\eta}^0 = (\eta_1^0, \cdots, \eta_n^0)'$. For possible values of the parameters, we denote $\boldsymbol{\theta} = (\theta', \boldsymbol{\phi}')'$, $\theta = (\lambda', \beta')'$ and $\boldsymbol{\phi} = (\boldsymbol{\alpha}', \boldsymbol{\eta}')'$, where $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_n)'$ and $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_n)'$.

Assumption 2.5. Let Λ be the parameter space of λ. For each $\lambda \in \Lambda$, we define

$$\mathbf{A}(\lambda) = \lambda_d(I_n \otimes W) + \lambda_o(W \otimes I_n) + \lambda_w(W \otimes W) \text{ and } \mathbf{A} = \mathbf{A}(\lambda^0).$$

We assume that $\rho_{\text{spec}}(\mathbf{A}) < 1$, where $\rho_{\text{spec}}(\mathbf{A}) = \max\{b(1,1), b(1,\varphi_{\min}), b(\varphi_{\min},1), b(\varphi_{\min},\varphi_{\min})\}$ with $b(1,1) = \lambda_d + \lambda_o + \lambda_w$, $b(1,\varphi_{\min}) = \lambda_d\varphi_{\min} + \lambda_o + \lambda_w\varphi_{\min}$, $b(\varphi_{\min},1) = \lambda_d + \lambda_o\varphi_{\min} + \lambda_w\varphi_{\min}$, and $b(\varphi_{\min},\varphi_{\min}) = \lambda_d\varphi_{\min} + \lambda_o\varphi_{\min} + \lambda_w\varphi_{\min}^2$.

Assumption 2.5 guarantees the existence of the semi-reduced form at each $\lambda \in \Lambda$.[12] Assumption 2.5 also restricts the aggregated influences to ij since $\|\mathbf{S}^{-1}(\lambda)\|_\infty \leq \frac{1}{1-\sup_n \sup_{\lambda\in\Lambda} \|\mathbf{A}(\lambda)\|_\infty} < \infty$ where $\mathbf{S}(\lambda) = I_N - \mathbf{A}(\lambda)$ for each $\lambda \in \Lambda$.

The Poisson pseudo maximum likelihood (PPML) estimation method. The Poisson pseudo maximum likelihood (PPML) estimator (Gourieroux et al., 1984; Santos Silva and Tenreyro, 2006) focuses on correctly specifying $\mathbb{E}(y_{ij}|\mathbf{z})$ without transforming the model.[13] As the most notable advantage, no additional assumptions are required on ξ_{ij} in estimation except $\mathbb{E}(\xi_{ij}|\mathbf{z}) = 1$. The log-likelihood function for $(\theta, \boldsymbol{\phi})$ is specified by following:

$$\ell_N(\theta, \boldsymbol{\phi}) = \sum_{i,j=1}^n \ell_{ij}(\theta, \boldsymbol{\phi}) - \frac{1}{2}\left(\sum_{j=1}^n \alpha_j - \sum_{i=1}^n \eta_i\right)^2, \tag{12}$$

where $\ell_{ij}(\theta, \boldsymbol{\phi}) = -\mu_{ij}(\theta, \boldsymbol{\phi}) + y_{ij}\ln(\mu_{ij}(\theta, \boldsymbol{\phi})) - \ln(y_{ij}!)$ and the second term in (12) denotes a penalty term for the normalization (see Sec.2 in Fernandez-Val and Weidner (2016)).[14] In contrast to traditional PPML estimation for a gravity equation, a notable feature of (12) is that each $\ell_{ij}(\theta, \boldsymbol{\phi})$ possibly includes all fixed-effect components $\{\alpha_j\}_{j=1}^n$ and $\{\eta_i\}_{i=1}^n$.

Then, the PPML estimator (PPMLE) can be obtained by

$$\left(\widehat{\theta}, \widehat{\boldsymbol{\phi}}\right) = \underset{\theta\in\Theta_\theta, \boldsymbol{\phi}\in\mathbb{R}^{2n}}{\arg\max}\ \ell_N(\theta, \boldsymbol{\phi}),$$

[12] This condition is slightly stronger than condition 6 in Assumption 2.4 (i). The purpose of this assumption is for asymptotic analysis.

[13] Another approach for estimating $\mathbb{E}(y_{ij}|\mathbf{z})$ without transforming the model is employing non-linear least squares (NLS). However, this method tends to give disproportionate weight to noise observations, leading to inefficient estimations. This inefficiency arises because the method heavily depends on a relatively small number of observations (Silva and Tenreyro (2006, Sec. III A). For details, refer to Section 2.1 in the supplement.

[14] The key insight of (12) is that we can work with an unconstrained optimization problem. Indeed, (12) imposes a single linear constraint, $v'\boldsymbol{\phi} = 0$ where $v = (l_n', -l_n')'$, to eliminate the identification issue originated from the fixed effects' additive feature: $\alpha_j + \eta_i = \alpha_j^* + \eta_i^*$ where $\alpha_j^* = \alpha_j + c$ and $\eta_i^* = \eta_i - c$ for any c. Note that a normalization restriction for this issue is not unique: Fernandez-Val and Weidner (2016) additionally mention the possibility of $\alpha_1 = 0$, while Lee and Yu (2010) employs $\sum_{i=1}^n \eta_i = 0$.

where Θ_θ denotes a parameter space of θ. For the asymptotic analysis, let $\widehat{\boldsymbol{\phi}}(\theta) = \arg\max_{\boldsymbol{\phi}\in\mathbb{R}^{2n}} \ell_N(\theta,\boldsymbol{\phi})$ for each $\theta \in \Theta_\theta$ and $\ell_N^c(\theta) = \ell_N(\theta,\widehat{\boldsymbol{\phi}}(\theta))$ denote the concentrated penalized log-likelihood function. For each element in $\widehat{\boldsymbol{\phi}}(\theta)$, we define $\widehat{\boldsymbol{\alpha}}(\theta) = (\widehat{\alpha}_1(\theta),\cdots,\widehat{\alpha}_n(\theta))'$ and $\widehat{\boldsymbol{\eta}}(\theta) = (\widehat{\eta}_1(\theta),\cdots,\widehat{\eta}_n(\theta))'$ for each $\theta \in \Theta_\theta$.

Identification condition. Next, we study the identification of $\boldsymbol{\theta}^0$. We define the following notations for this purpose: Let $\mathbf{H}^{\theta\theta}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{H}^{\lambda\lambda}(\boldsymbol{\theta}) & \mathbf{H}^{\beta\lambda\prime}(\boldsymbol{\theta}) \\ \mathbf{H}^{\beta\lambda}(\boldsymbol{\theta}) & \mathbf{H}^{\beta\beta}(\boldsymbol{\theta}) \end{bmatrix}$ with

$$\mathbf{H}^{\lambda\lambda}(\boldsymbol{\theta}) = \begin{bmatrix} \left(2\mathbf{W}_d^2\mathbf{S}^{-3}(\lambda)\mathbf{Z}(\boldsymbol{\theta})\right)'\mathbf{u}(\boldsymbol{\theta}) & \left(2\mathbf{W}_d\mathbf{W}_o\mathbf{S}^{-3}(\lambda)\mathbf{Z}(\boldsymbol{\theta})\right)'\mathbf{u}(\boldsymbol{\theta}) & \left(2\mathbf{W}_d\mathbf{W}_w\mathbf{S}^{-3}(\lambda)\mathbf{Z}(\boldsymbol{\theta})\right)'\mathbf{u}(\boldsymbol{\theta}) \\ * & \left(2\mathbf{W}_o^2\mathbf{S}^{-3}(\lambda)\mathbf{Z}(\boldsymbol{\theta})\right)'\mathbf{u}(\boldsymbol{\theta}) & \left(2\mathbf{W}_o\mathbf{W}_w\mathbf{S}^{-3}(\lambda)\mathbf{Z}(\boldsymbol{\theta})\right)'\mathbf{u}(\boldsymbol{\theta}) \\ * & * & \left(2\mathbf{W}_w^2\mathbf{S}^{-3}(\lambda)\mathbf{Z}(\boldsymbol{\theta})\right)'\mathbf{u}(\boldsymbol{\theta}) \end{bmatrix},$$

$$\mathbf{H}^{\beta\lambda}(\boldsymbol{\theta}) = \left[\left(\mathbf{W}_d\mathbf{S}^{-2}(\lambda)\mathbf{X}\right)'\mathbf{u}(\boldsymbol{\theta}) \quad \left(\mathbf{W}_o\mathbf{S}^{-2}(\lambda)\mathbf{X}\right)'\mathbf{u}(\boldsymbol{\theta}) \quad \left(\mathbf{W}_w\mathbf{S}^{-2}(\lambda)\mathbf{X}\right)'\mathbf{u}(\boldsymbol{\theta})\right],\ \mathbf{H}^{\beta\beta}(\boldsymbol{\theta}) = \mathbf{0}_{K\times K},$$

$$\mathbf{H}^{\phi\theta}(\boldsymbol{\theta}) = \left[\left(\mathbf{W}_d\mathbf{S}^{-2}(\lambda)\mathbf{D}\right)'\mathbf{u}(\boldsymbol{\theta}) \quad \left(\mathbf{W}_o\mathbf{S}^{-2}(\lambda)\mathbf{D}\right)'\mathbf{u}(\boldsymbol{\theta}) \quad \left(\mathbf{W}_w\mathbf{S}^{-2}(\lambda)\mathbf{D}\right)'\mathbf{u}(\boldsymbol{\theta}) \quad \mathbf{0}_{2n\times K}\right] \text{ and } \mathbf{H}^{\phi\phi} = -\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \otimes l_n l_n'.$$

Here,

- $\mathbf{G}(\boldsymbol{\theta}) = \left[\mathbf{W}_d\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta}), \mathbf{W}_o\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta}), \mathbf{W}_w\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta}), \mathbf{X}\right]$, and $\mathbf{G} = \mathbf{G}(\boldsymbol{\theta}^0)$, where $\mathbf{X}$ denotes an $N \times K$ matrix which has $x_{ij,k}$ as the $((j-1)n+i,k)$- element of $\mathbf{X}$, $\mathbf{Z}(\boldsymbol{\theta}) = \mathbf{X}\beta + \boldsymbol{\alpha}\otimes l_n + l_n \otimes \boldsymbol{\eta}$ with $\mathbf{Z} = \mathbf{Z}(\boldsymbol{\theta}^0)$, $\mathbf{W}_d = I_n \otimes W$, $\mathbf{W}_o = W \otimes I_n$, and $\mathbf{W}_w = W \otimes W$;
- $\mathbf{D} = [\mathbf{I}_n \otimes l_n, l_n \otimes \mathbf{I}_n]$ is an $N \times 2n$ matrix for dummy variables;
- $\boldsymbol{\mu}(\boldsymbol{\theta}) = \left(\exp\left(\tilde{\mu}_{11}(\boldsymbol{\theta})\right),\cdots,\exp\left(\tilde{\mu}_{n1}(\boldsymbol{\theta})\right),\cdots,\exp\left(\tilde{\mu}_{1n}(\boldsymbol{\theta})\right),\cdots,\exp\left(\tilde{\mu}_{nn}(\boldsymbol{\theta})\right)\right)$ with $\boldsymbol{\mu}^0 = \boldsymbol{\mu}(\boldsymbol{\theta}^0)$ and $\tilde{\mu}_{ij}(\boldsymbol{\theta}) = \sum_{k,l=1}^n s_{ij,kl}(\lambda)\left(x_{kl}'\beta + \alpha_l + \eta_k\right)$; and $\mathbf{u}(\boldsymbol{\theta}) = \mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\theta})$ with $\mathbf{u} = \mathbf{y} - \boldsymbol{\mu}^0$.

Assumption 2.6 (Identification)**.** Let $\boldsymbol{\Theta} = \Theta_\theta \times \boldsymbol{\Phi}$ be the parameter space of $\boldsymbol{\theta}$, where Θ_θ denotes a compact parameter space of θ and $\boldsymbol{\Phi}$ represents a parameter space of $\boldsymbol{\phi}$. Here, $\boldsymbol{\Phi}$ is a subset of $[-C,C]^{2n}$ for some finite constant $C > 0$.

(i) Assume $\liminf_{n\to\infty}\inf_{\boldsymbol{\phi}\in\boldsymbol{\Phi}}\varphi_{\min}\left(\frac{1}{n}\mathbf{D}'\mathbf{S}^{-1\prime}(\lambda)\mathrm{Diag}\left(\boldsymbol{\mu}(\boldsymbol{\theta})\right)\mathbf{S}^{-1}(\lambda)\mathbf{D} - \frac{1}{n}\mathbf{H}^{\phi\phi}\right) > 0$ for each $\theta \in \Theta_\theta$. Then, $\widehat{\boldsymbol{\phi}}(\theta) = \mathrm{argmax}_{\boldsymbol{\phi}\in\boldsymbol{\Phi}}\ell_N(\theta,\boldsymbol{\phi})$ is unique for each $\theta \in \Theta_\theta$ and for a sufficiently large n.

(ii) For each $\theta \in \Theta_\theta$, let

$$\begin{aligned}\widehat{\mathbf{H}}(\theta) =& \frac{1}{N}\widehat{\mathbf{G}}'(\theta)\mathbf{S}^{-1\prime}(\lambda)\mathrm{Diag}\left(\widehat{\boldsymbol{\mu}}(\theta)\right)\mathbf{S}^{-1}(\lambda)\widehat{\mathbf{G}}(\theta) - \frac{1}{N}\widehat{\mathbf{H}}^{\theta\theta}(\theta) - \frac{1}{N}\left(\widehat{\mathbf{G}}'(\theta)\mathbf{S}^{-1\prime}(\lambda)\mathrm{Diag}\left(\widehat{\boldsymbol{\mu}}(\theta)\right)\mathbf{S}^{-1}(\lambda)\mathbf{D} - \widehat{\mathbf{H}}^{\phi\theta\prime}(\theta)\right) \\ &\cdot\left(\mathbf{D}\mathbf{S}^{-1\prime}(\lambda)\mathrm{Diag}\left(\widehat{\boldsymbol{\mu}}(\theta)\right)\mathbf{S}^{-1}(\lambda)\mathbf{D} - \mathbf{H}^{\phi\phi}\right)^{-1}\cdot\left(\mathbf{D}'\mathbf{S}^{-1\prime}(\lambda)\mathrm{Diag}\left(\widehat{\boldsymbol{\mu}}(\theta)\right)\mathbf{S}^{-1}(\lambda)\widehat{\mathbf{G}}(\theta) - \widehat{\mathbf{H}}^{\phi\theta}(\theta)\right),\end{aligned}$$

where $\widehat{\mathbf{G}}(\theta) = \mathbf{G}\left(\theta,\widehat{\boldsymbol{\phi}}(\theta)\right)$, $\widehat{\boldsymbol{\mu}}(\theta) = \boldsymbol{\mu}(\theta,\widehat{\boldsymbol{\phi}}(\theta))$, $\widehat{\mathbf{H}}^{\theta\theta}(\theta) = \mathbf{H}(\theta,\widehat{\boldsymbol{\phi}}(\theta))$, and $\widehat{\mathbf{H}}^{\phi\theta}(\theta) = \mathbf{H}^{\phi\theta}(\theta,\widehat{\boldsymbol{\phi}}(\theta))$ for each $\theta \in \Theta_\theta$.

Assume $\liminf_{n\to\infty}\inf_{\theta\in\Theta_\theta}\varphi_{\min}\left(\widehat{\mathbf{H}}(\theta)\right) > 0$. Then, $\hat{\theta} = \mathrm{argmax}_{\theta\in\Theta_\theta}\ell_N^c(\theta)$ is unique for a sufficiently large n.

Assumption 2.6 establishes the identification conditions for the first-stage parameters $\boldsymbol{\theta}^0$. Assumption 2.6 (i) implies the uniqueness of $\boldsymbol{\phi}(\theta) = \mathrm{argmax}_{\boldsymbol{\phi}\in\boldsymbol{\Theta}_\phi}\ell_\infty(\theta,\boldsymbol{\phi})$ for each $\theta \in \Theta_\theta$, where $\ell_\infty(\boldsymbol{\theta}) \equiv \mathrm{plim}_{n\to\infty}\frac{1}{N}\ell_N(\boldsymbol{\theta})$ for each $\boldsymbol{\theta}$.

Assumption 2.6 leads to the uniqueness of $\theta^0 = \text{argmax}_{\theta \in \Theta_\theta} \ell_\infty(\theta, \boldsymbol{\phi}(\theta))$, and consequently, $\boldsymbol{\phi}^0 = \boldsymbol{\phi}(\theta^0)$. Matrices, $\mathbf{H}^{\theta\theta}(\boldsymbol{\theta})$, $\mathbf{H}^{\phi\theta}(\boldsymbol{\theta})$, and $\mathbf{H}^{\phi\phi}(\boldsymbol{\theta})$, are the minor components of Hessian, which are newly introduced by the model's network influences. Contrary to the traditional gravity equation estimation, which guarantees the global maximum, identifying the parameters in our specification is not trivial. Intuitively, the major part of $\widehat{\mathbf{H}}(\theta)$ is likely positive definite when $\mathbf{S}^{-1}(\lambda)$ is well-defined, Hence, if the minor part's influence is less than that of the major part, the conditions in Assumption 2.6 are satisfied. From this identification condition, we suggest estimating the conventional gravity equation first and obtaining $\widetilde{\beta}$, $\widetilde{\boldsymbol{\alpha}}$, and $\widetilde{\boldsymbol{\eta}}$. One then set the initial parameter to maximize (12) as $\widehat{\boldsymbol{\theta}}^{(0)} = \left(0, 0, 0, \widetilde{\beta}, \widetilde{\boldsymbol{\alpha}}', \widetilde{\boldsymbol{\eta}}'\right)'$ since $\mathbf{S}^{-1}(\lambda) = I_N$ when $\lambda = \mathbf{0}$. Details can be found in Lemmas 2.4 and 2.5 of Section 2.3 of the supplement file.

Remarks. Our estimation method with equation (9) is generally applicable to other applications such as migration flows, traffic flows, commuting flows, and so on, when one wants to specify spatial externalities in OD flows. The main distinct feature in the case of trade flows is the structure of the fixed-effect components in (7). To identify γ_d and γ_o in (7), we need to utilize the estimated fixed effects. Since λ^0 and β^0 are identifiable from the semi-reduced form (9), an iteration procedure is not required. For details, refer to Section 2.4 in the supplement file.

3 Statistical Analysis

This section derives the asymptotic distribution of the PPMLE, presents relevant statistical inference, and presents simulation results for finite samples.

3.1 Asymptotic distribution of the PPMLE

To derive the asymptotic distribution of $\widehat{\theta}$ and $\widehat{\boldsymbol{\phi}}$, here is the regularity assumption for Theorems 3.1 and 3.2. Details can be found in Section 2 of the supplement file.

Assumption 3.1. (i) Assume $\{x_{ij}\}$, $\{\eta_i^0\}$, and $\{\alpha_j^0\}$ are random fields satisfying $\max_k \sup_{i,j,n} |x_{ij,k}| < C$, $\sup_{i,n} |\eta_i^0| < C$, and $\sup_{j,n} |\alpha_j^0| < C$, where $C > 0$ denotes a generic finite constant.

(ii) Assume $\{\xi_{ij}\}$ is a random field satisfying $\sup_{i,j,n} \mathbb{E}|\xi_{ij}|^{2+c} < C$ for some $c > 0$.

(iii) Assume $\mathbb{E}(\xi_{ij}|\mathbf{z}) = 1$ for all $i, j = 1, \cdots, n$.

Assumption 3.1 describes the properties of the components in $\{x_{ij}\}$, $\{\eta_i^0\}$ and $\{\alpha_j^0\}$, and the errors $\{\xi_{ij}\}$ for a simple asymptotic analysis. Under Assumption 3.1, we can induce the same implication on the additive error u_{ij}: $\mathbb{E}(u_{ij}|\mathbf{z}) = 0$ and $\sup_{n,i,j} \mathbb{E}|u_{ij}|^{2+c} < C$. The theorems below state the asymptotic properties of the PPMLE. The asymptotic properties of the PPMLE for the fixed-effect parameters are utilized for statistical inference on the estimate of γ_d and γ_o. Details can be found in 2.3 of the supplement.

Theorem 3.1. Suppose that Assumptions 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 3.1, and 3.2 hold. Let

$$\boldsymbol{\Sigma}_{\theta,N} = \frac{1}{N}\mathbf{G}'\mathbf{S}^{-1\prime}\text{Diag}(\boldsymbol{\mu})\mathbf{M_D}\mathbf{S}^{-1}\mathbf{G}, \text{ and } \boldsymbol{\Omega}_{\theta,N} = \frac{1}{N}\mathbf{G}'\mathbf{S}^{-1\prime}\mathbf{M_D}'\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{M_D}\mathbf{S}^{-1}\mathbf{G},$$

where $\mathbf{M_D} = I_N - \mathbf{P_D}\text{Diag}(\boldsymbol{\mu})$ with $\mathbf{P_D} = \mathbf{S}^{-1}\mathbf{D}(\widetilde{\mathbf{D}'\mathbf{D}})^{-1}\mathbf{D}'\mathbf{S}^{-1\prime}$ and $\widetilde{\mathbf{D}'\mathbf{D}} = \mathbf{D}'\mathbf{S}^{-1\prime}\,\text{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{D} - \mathbf{H}^{\phi\phi}$.

For each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, let $\boldsymbol{\Sigma}_{\theta,N}(\boldsymbol{\theta})$ and $\boldsymbol{\Omega}_{\theta,N}(\boldsymbol{\theta})$ denote respectively $\boldsymbol{\Sigma}_{\theta,N}$ and $\boldsymbol{\Omega}_{\theta,N}$ evaluated at $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. If we assume $\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \varphi_{\min}\left(\boldsymbol{\Sigma}_{\theta,N}(\boldsymbol{\theta})\right) > 0$ and $\liminf_{n\to\infty} \inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \varphi_{\min}\left(\boldsymbol{\Omega}_{\theta,N}(\boldsymbol{\theta})\right) > 0$, we then have

$$\sqrt{N}\left(\hat{\theta} - \theta^0\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_\theta^{-1}\boldsymbol{\Omega}_\theta\boldsymbol{\Sigma}_\theta^{-1}\right) \text{ as } n \to \infty, \tag{13}$$

where $\boldsymbol{\Sigma}_\theta = \text{plim}_{n\to\infty} \boldsymbol{\Sigma}_{\theta,N}$, $\boldsymbol{\Omega}_\theta = \text{plim}_{n\to\infty} \boldsymbol{\Omega}_{\theta,N}$.

Theorem 3.2. Suppose that Assumptions 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 3.1 and 3.2 hold. Let

$$\mathbf{V}_{\boldsymbol{\phi},N} = n\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1} \mathbf{D}'\mathbf{S}^{-1\prime}\mathbf{M}_{\boldsymbol{\phi}}'\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{M}_{\boldsymbol{\phi}}\mathbf{S}^{-1}\mathbf{D}\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1},$$

where

$$\mathbf{M}_{\boldsymbol{\phi}} = I_N - \mathbf{M}_{\mathbf{D}}\mathbf{S}^{-1}\mathbf{G}\left(\mathbf{G}'\mathbf{S}^{-1\prime}\text{Diag}(\boldsymbol{\mu})\mathbf{M}_{\mathbf{D}}\mathbf{S}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{S}^{-1\prime}\text{Diag}(\boldsymbol{\mu}).$$

Then,

$$\sqrt{n}\left(\widehat{\alpha}_j - \alpha_j^0\right) \xrightarrow{d} \mathcal{N}(0, \lim_{n\to\infty} e_{2n,j}'\mathbf{V}_{\boldsymbol{\phi},N}e_{2n,j}), \text{ and}$$
$$\sqrt{n}\left(\widehat{\eta}_i - \eta_i^0\right) \xrightarrow{d} \mathcal{N}(0, \lim_{n\to\infty} e_{2n,n+i}'\mathbf{V}_{\boldsymbol{\phi},N}e_{2n,n+i})$$

as $n \to \infty$, where $e_{2n,j}$ denotes the $2n$-dimensional unit vector with its j-th element equal to 1 and all other elements equal to 0.

3.2 Variance estimation

For a practitioner who utilizes OD flow data, this subsection provides a method for spatial heteroskedasticity and autocorrelation consistent (spatial HAC) estimation of the covariance matrices in Theorems 3.1 and 3.2 of the PPMLE. In the main draft, we focus on estimating $\boldsymbol{\Omega}_\theta$. For details and additional results, refer to Section 2.3 in the supplement file. We basically follow the existing literature for cross-sectional data (e.g., Kelejian and Prucha (2007) and Kim and Sun (2011)). At first, we provide the regularity assumptions.

Assumption 3.2. (i) For the structure of the additive error $\mathbf{u} = (u_{11}, \cdots, u_{n1}, \cdots, u_{1n}, \cdots, u_{nn})'$, we assume

$$\mathbf{u} = \mathbf{B}\mathbf{H}\boldsymbol{\epsilon}, \tag{14}$$

where $\mathbf{B}$ denotes some $N \times N$ matrix, $\mathbf{H} = \text{diag}\left(\sigma_{11}^*, \cdots, \sigma_{n1}^*, \cdots, \sigma_{1n}^*, \cdots, \sigma_{nn}^*\right)$, and $\boldsymbol{\epsilon} = (\epsilon_{11}, \cdots, \epsilon_{n1}, \cdots, \epsilon_{1n}, \cdots, \epsilon_{nn})'$ is an $N \times 1$ vector of innovations.

(ii) $\epsilon_{ij} \overset{i.i.d.}{\sim} (0, 1)$ across ij with $\sup_{n,i,j} \mathbb{E}|\epsilon_{ij}|^4 < \infty$.

(iii) We assume $0 < \inf_{i,j,n} \sigma_{ij}^* \leq \sup_{i,j,n} \sigma_{ij}^* < \infty$.

(iv) We assume that $\mathbf{B}$ is nonsingular and $\sup_n \max\{\|\mathbf{B}\|_\infty, \|\mathbf{B}\|_1\} < \infty$.

Assumption 3.2 (i) describes the basic covariance structure. Hence, we have

$$\boldsymbol{\Omega}_{\theta,N} = \frac{1}{N}\mathbf{R}'\mathbf{R} = \frac{1}{N}\sum_{i,j,k,l=1}^{n} R_{ij}R_{kl}' = \frac{1}{N}\sum_{i,j,k,l=1}^{n} \mathbb{E}\left(\left(\mathbf{G}'\mathbf{S}^{-1\prime}\mathbf{M}_{\mathbf{D}}'\mathbf{u}\right)_{.,ij}\left(\mathbf{u}'\mathbf{M}_{\mathbf{D}}\mathbf{S}^{-1}\mathbf{G}\right)_{kl,.}\right),$$

where $\mathbf{R} = \mathbf{HB}'\mathbf{M_D}\mathbf{S}^{-1}\mathbf{G}$, R_{ij} denotes the $((j-1)n+i)$-th column of $\mathbf{R}$, $\left(\mathbf{G}'\mathbf{S}^{-1\prime}\mathbf{M}_\mathbf{D}'\mathbf{u}\right)_{.,ij}$ denotes the $(j-1)n+i$-th column of $\mathbf{G}'\mathbf{S}^{-1\prime}\mathbf{M}_\mathbf{D}'\mathbf{u}$ and $\left(\mathbf{u}'\mathbf{M_D}\mathbf{S}^{-1}\mathbf{G}\right)_{kl,.}$ is the $(l-1)n+k$-th row of $\mathbf{u}'\mathbf{M_D}\mathbf{S}^{-1}\mathbf{G}$. Section 3.3 provides an example of (14). Conditions (ii) and (iii) in Assumption 3.2 are conventional. Condition (iv) comes from Kelejian and Prucha (2007). Still, it can be relaxed by the ideas of Pesaran and Yang (2020, 2021).[15]

Assumption 3.3. (i) There exists a distance measure $d_{ij,kl}$ measuring the distance between ij and kl. There exists a constant $q_d > 0$ such that $\sup_n \frac{1}{N}\sum_{i,j,k,l=1}^{n} \|R_{ij}R_{kl}'\| d_{ij,kl}^{q_d} < \infty$.

(ii) Let $d_{ij,kl}^*$ be a feasible distance between ij and kl. We assume $d_{ij,kl}^* = d_{ij,kl} + \nu_{ij,kl}$, where $\nu_{ij,kl}$ is a measurement error. We assume that $\{\nu_{ij,kl}\}$ are independent of $\{\epsilon_{ij}\}$ and any component of $\mathbf{z}$, $\nu_{ij,kl} = o(d_N)$, where d_N is a bandwidth, and $\sup_n \frac{1}{N}\sum_{i,j,k,l}\|R_{ij}R_{kl}'\|\mathbb{E}|\nu_{ij,kl}|^{q_d} < \infty$.

Let kl be a pseudo-neighbor of ij when $d_{ij,kl}^* \le d_N$. Define $\deg_{ij}^* = \sum_{k,l=1}^{n}\mathbb{I}\{d_{ij,kl}^* \le d_N\}$ and $\deg^* = \frac{1}{N}\sum_{i,j=1}^{n}\deg_{ij}^*$. Based on these definitions, we define

$$\mathcal{E} = \{ij : \mathbb{E}|\deg_{ij}^* - \mathbb{E}(\deg^*)| = o\left(\deg^*\right)\},$$

(iii) For each $ij \in \mathcal{E}$, there is a constant $C > 0$ such that $\deg_{ij}^* \le C \cdot \mathbb{E}\left(\deg^*\right)$.

(iv) As $n \to \infty$, $\frac{N_2}{N} \to 0$, $\mathbb{E}\left(\deg^*\right) \to \infty$, $d_N \to \infty$, and $\frac{\mathbb{E}(\deg^*)}{N} \to 0$.

(v) For each $ij \in \mathcal{E}$,

$$\lim_{n\to\infty} \mathsf{Var}\left(\frac{1}{\sqrt{\mathbb{E}(\deg^*)}}\sum_{kl:d_{ij,kl}^*\le d_N}\left(\mathbf{G}'\mathbf{S}^{-1\prime}\mathbf{M}_\mathbf{D}'\mathbf{u}\right)_{.,kl}\right) = \mathbf{\Omega}_\theta.$$

Assumption 3.4. (i) The kernel $\mathsf{K} : \mathbb{R}_+ \cup \{0\} \to [-1,1]$, with $\mathsf{K}(0) = 1$, $\mathsf{K}(x) = \mathsf{K}(-x)$, $\mathsf{K}(x) = 0$ for $|x| > 1$, satisfies

$$|\mathsf{K}(x) - 1| \le c_\mathsf{K} \cdot |x|^{\rho_\mathsf{K}} \text{ for } |x| \le 1,$$

for some $0 < c_\mathsf{K} < \infty$ and $\rho_\mathsf{K} \ge 1$.

(ii) For every pair ij, $\frac{1}{\mathbb{E}(\deg^*)}\mathbb{E}\left(\sum_{k,l=1}^{n}\mathsf{K}^2\left(\frac{d_{ij,kl}^*}{d_N}\right)\right) \to \overline{\mathsf{K}} < \infty$.

Assumption 3.3 (i) characterizes an admissible type of dependence. It excludes the infill asymptotic. An example is $\|R_{ij}R_{kl}'\| \le \frac{C}{(1+d_{ij,kl})^{c+\Delta}}$ for some $C > 0$ and $\Delta > 2d$, which it means that the magnitude of the covariance factor $\|R_{ij}R_{kl}'\|$ diminishes when $d_{ij,kl} \to \infty$. Assumption 3.3 (ii) allows a feasible distance measure $d_{ij,kl}^*$ with a measurement error $\nu_{ij,kl}$. In practice, since a distance measure between two pairs is generally not available, practitioners need to construct a proxy distance from a feasible distance measure d_{ij}^*. In Section 3.3, we evaluate the simulation results for possible distance measures for pairs. Under Assumption 3.3 (iii), if $ij \in \mathcal{E}$ (i.e., ij is in the interior), the number of pseudo neighbors of ij is the same order as the average number of pseudo neighbors $\mathbb{E}(\deg^*)$. Assumption 3.3 (iv) states that (i) the proportion of boundary pairs is asymptotically negligible; (ii) the number of average neighboring pairs ($\mathbb{E}(\deg^*)$) and a bandwidth (d_N) are increasing functions of n; and (iii) $\mathbb{E}(\deg^*)$ increases but much slower than N. To understand Assumption 3.3 (v), note that $\frac{1}{\sqrt{\mathbb{E}(\deg^*)}}\sum_{kl:d_{ij,kl}^*\le d_N}\left(\mathbf{G}'\mathbf{S}^{-1\prime}\mathbf{M}_\mathbf{D}'\mathbf{u}\right)_{.,kl}$ is a local average around ij, while $\frac{1}{\sqrt{N}}\sum_{k,l=1}^{n}\left(\mathbf{G}'\mathbf{S}^{-1\prime}\mathbf{M}_\mathbf{D}'\mathbf{u}\right)_{.,kl}$ is the global average. If $ij \in \mathcal{E}$ (interior), the local average and the global average have the same asymptotic variance. Assumption 3.4 is conventional in spatial HAC literature (Kelejian and Prucha,

[15] If the error structure follows (16) in the simulation section, we can allow a finite number of moderate dominant units. We will leave this issue for future research.

2007; Kim and Sun, 2011).[16]

Hence, we define the spatial HAC estimator

$$\widehat{\mathbf{\Omega}}_{\theta,N} = \frac{1}{N}\sum_{i,j,k,l=1}^{n}\left(\widehat{\mathbf{G}}'\widehat{\mathbf{S}}^{-1\prime}\widehat{\mathbf{M}}_{\mathbf{D}}'\widehat{\mathbf{u}}\right)_{.,ij}\left(\widehat{\mathbf{u}}'\widehat{\mathbf{M}}_{\mathbf{D}}\widehat{\mathbf{S}}^{-1}\widehat{\mathbf{G}}\right)_{kl,.}\mathsf{K}\left(\frac{d^*_{ij,kl}}{d_N}\right),$$

and

$$\widetilde{\mathbf{\Omega}}_{\theta,N} = \frac{1}{N}\sum_{i,j,k,l=1}^{n}\left(\mathbf{G}'\mathbf{S}^{-1\prime}\mathbf{M}_{\mathbf{D}}'\mathbf{u}\right)_{.,ij}\left(\mathbf{u}'\mathbf{M}_{\mathbf{D}}\mathbf{S}^{-1}\mathbf{G}\right)_{kl,.}\mathsf{K}\left(\frac{d^*_{ij,kl}}{d_N}\right),$$

which is the infeasible spatial HAC estimator.

Theorem 3.3. Assume that Assumptions 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 3.1 and 3.2 hold for Theorems 3.1 and 3.2. Also, we suppose that Assumptions 3.2, 3.3, and 3.4 hold. Then, we have the following results:

(i) (Variance) $\lim_{n\to\infty}\frac{N}{\mathbb{E}(\deg^*)}\mathsf{Var}\left(\text{vec}\left(\widetilde{\mathbf{\Omega}}_{\theta,N}\right)\right) = \overline{\mathsf{K}}(1+C)(\mathbf{\Omega}_\theta\otimes\mathbf{\Omega}_\theta)$, where C denotes the $(3+K)^2\times(3+K)^2$ commutation matrix[17];

(ii) (Bias) $\lim_{n\to\infty} d_N^q\left(\mathbb{E}\left(\widetilde{\mathbf{\Omega}}_{\theta,N}\right)-\mathbf{\Omega}_{\theta,N}\right) = -K_q\mathbf{\Omega}_\theta^{(q)}$, where q denotes the Parzen characteristic exponent of $\mathsf{K}(\cdot)$ and $\mathbf{\Omega}_\theta^{(g)} = \lim_{n\to\infty}\frac{1}{N}\sum_{i,j,k,l=1}^n R_{ij}R_{kl}'\cdot\mathbb{E}\left((d^*_{ij,kl})^q\right)$ for each q. We assume $0\le q\le q_d$;[18] and

(iii) If $0<\lim_{n\to\infty}\frac{d_N^{2q}\mathbb{E}(\deg^*)}{N}<\infty$, $\sqrt{\frac{N}{\mathbb{E}(\deg^*)}}\left(\widehat{\mathbf{\Omega}}_{\theta,N}-\mathbf{\Omega}_{\theta,N}\right) = O_p(1)$ and $\sqrt{\frac{N}{\mathbb{E}(\deg^*)}}\left(\widehat{\mathbf{\Omega}}_{\theta,N}-\widetilde{\mathbf{\Omega}}_{\theta,N}\right) = O_p(1)$.

[16] In particular, Assumption 3.4 (ii) characterizes how pair units are distributed, how they are included in the support of a kernel function. By Lemma A.1 in Jenish and Prucha (2009), $\mathbb{E}(\deg^*) = C\cdot d_N^{2d}$ for some $C>0$ and the ij's number of neighboring pairs in the distance $[r, r+dr)$ is $\widetilde{C}\cdot r^{2d-1}dr$ for some $\widetilde{C}>0$. Hence,

$$\mathbb{E}\left(\sum_{k,l=1}^n\mathsf{K}^2\left(\frac{d^*_{ij,kl}}{d_N}\right)\right) = \int_0^{d_N}\widetilde{C}\cdot r^{2d-1}\mathsf{K}\left(\frac{r}{d_N}\right)dr = \widetilde{C}\cdot d_N^{2d}\cdot\int_0^1 u^{2d-1}\mathsf{K}^2(u)du.$$

Hence, $\frac{1}{\mathbb{E}(\deg^*)}\mathbb{E}\left(\sum_{k,l=1}^n\mathsf{K}^2\left(\frac{d^*_{ij,kl}}{d_N}\right)\right) = \frac{\widetilde{C}}{C}\int_0^1 u^{2d-1}\mathsf{K}^2(u)du$. Without loss of generality, we can consider $\overline{\mathsf{K}} = \int_0^1 u^{2d-1}\mathsf{K}^2(u)du$. If $\mathsf{K}(u) = 1-|u|$ for $|u|\le 1$ (Bartlett kernel), $\overline{\mathsf{K}} = \int_0^1 u^{2d-1}(1-u)^2du = \frac{1}{2d(2d+1)(d+1)}$. When $d=2$, $\overline{\mathsf{K}} = \frac{1}{60}$. Since our goal is to establish the HAC estimator $\widehat{\mathbf{\Omega}}_{\theta,N}$ and its infeasible version ($\widetilde{\mathbf{\Omega}}_{\theta,N}$) takes a form of $\frac{1}{N}\sum_{i,j,k,l=1}^n V_{ij}V_{kl}'\mathsf{K}\left(\frac{d^*_{ij,kl}}{d_N}\right)$ for some V_{ij}, its precision measure $\mathsf{Var}\left(\text{vec}(\widetilde{\mathbf{\Omega}}_{\theta,N})\right)$ is mainly characterized by $\frac{1}{N^2}\sum_{i,j,k,l=1}^n\mathsf{K}^2\left(\frac{d^*_{ij,kl}}{d_N}\right)\mathsf{Var}\left(\text{vec}\left(V_{ij}V_{kl}'\right)\right)$. In this case, the average weight is $\overline{\mathsf{K}} = \frac{1}{60}$.

[17] C satisfies $C\text{vec}(B) = \text{vec}(B')$ for a $K\times K$ matrix B. For example, if B is a 2×2 matrix,

$$C = \begin{bmatrix}1&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&1\end{bmatrix}.$$

[18] Here, q shows the smoothness of $\mathsf{K}(x)$ at $x=0$. When $\mathsf{K}(u) = 1-|u|$ for $|u|\le 1$ (Bartlett), $\frac{1-\mathsf{K}(u)}{|u|}\to 1$ as $|u|\to 0$. Hence, $q=1$ and $K_q = 1$. If $\mathsf{K}(u)$ is the Parzen kernel, $\frac{1-\mathsf{K}(u)}{u^2}\to 6$. Then, $q=2$ and $K_6 = 6$. If $\mathsf{K}(u)$ is the Tukey-Hanning kernel, $q=2$ and $K_q = \frac{\pi^2}{4}$. This quantity characterizes the bias of $\widetilde{\mathbf{\Omega}}_{\theta,N}$. Since $\mathsf{K}\left(\frac{d^*_{ij,kl}}{d_N}\right)-1\simeq -K_p\left(\frac{d^*_{ij,kl}}{d_N}\right)^q = -\frac{K_q}{d_N^q}\cdot(d^*_{ij,kl})^q$ around 0, we have

$$\mathbb{E}\left(\widetilde{\mathbf{\Omega}}_{\theta,N}\right)-\mathbf{\Omega}_{\theta,N} = \frac{1}{N}\sum_{i,j,k,l=1}^n R_{ij}R_{kl}'\left(\mathsf{K}\left(\frac{d^*_{ij,kl}}{d_N}-1\right)\right)\simeq\frac{1}{N}\sum_{i,j,k,l=1}^n R_{ij}R_{kl}'\cdot(d^*_{ij,kl})^q\simeq -\frac{K_q}{d_N^q}\cdot\mathbf{\Omega}_\theta^{(q)}.$$

3.3 Monte Carlo Simulations

Here, we introduce a computation issue and examine the finite sample properties of the PPMLE.

Efficient computation of the network multiplier matrix $\mathbf{S}^{-1}$. A challenge in our estimation method is the heavier computational costs, primarily attributed to the computation of $\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta})$. As n increases, the computation cost increases exponentially. By considering the structure of three networks $I_n \otimes W$, $W \otimes I_n$, and $W \otimes W$ generated by a single row-normalized connectivity matrix W, we can achieve efficient computation procedures.[footnote]

By Assumption 2.2 and the spectral decomposition theorem, $W = QDQ^{-1}$, where $D = \operatorname{diag}(\varphi_1, \cdots, \varphi_n)$ and Q denotes the eigenvector matrix. Since W is originated by a symmetric relationship, all φ_is are real. Since $I_n \otimes W$, $W \otimes I_n$, and $W \otimes W$ share the same eigenvector basis, an eigenvalue of $\mathbf{A}(\lambda)$ can be represented by

$$\mathbf{A}(\lambda)(q_i \otimes q_j) = (\lambda_d \varphi_j + \varphi_o \varphi_i + \lambda_w \varphi_i \varphi_j) \text{ for } i, j = 1, \cdots, n, \tag{15}$$

where q_i denotes the ith column vector of Q.

Define $Z^{\text{mat}}(\boldsymbol{\theta})$ satisfying $\mathbf{Z}(\boldsymbol{\theta}) = \operatorname{vec}(Z^{\text{mat}}(\boldsymbol{\theta}))$. Then, we want to obtain the matrix fixed-point $T(\boldsymbol{\theta})$ satisfying $T(\boldsymbol{\theta}) = \mathbf{S}^{-1}(\lambda)Z^{\text{mat}}(\boldsymbol{\theta})$.

Note that Q and D are invariant in the estimation procedure.

Observe

$$\widetilde{Z}^{\text{mat}}(\boldsymbol{\theta}) = \widetilde{T}(\boldsymbol{\theta}) - \lambda_d D\widetilde{T}(\boldsymbol{\theta}) - \lambda_o \widetilde{T}(\boldsymbol{\theta})D - \lambda_w D\widetilde{T}(\boldsymbol{\theta})D,$$

where $\widetilde{Z}^{\text{mat}}(\boldsymbol{\theta}) = Q^{-1}Z^{\text{mat}}(\boldsymbol{\theta})Q^{-1\prime}$ and $\widetilde{T}(\boldsymbol{\theta}) = Q^{-1}T(\boldsymbol{\theta})Q^{-1\prime}$. It implies

$$(\widetilde{T}(\boldsymbol{\theta}))_{ij} = \frac{(\widetilde{Z}^{\text{mat}}(\boldsymbol{\theta}))_{ij}}{1 - \lambda_d \varphi_i - \lambda_o \varphi_j - \lambda_w \varphi_i \varphi_j}$$

Then, we can easily recover $T(\boldsymbol{\theta}) = Q\widetilde{T}(\boldsymbol{\theta})Q'$.

Data generating process. We consider $n = 49$ regions. Following Pesaran and Yang (2020, 2021), the regional proximity matrix W is constructed to feature two "dominant" regions (units 1 and 2). For regions $i \in \{1, 2\}$, the code draws relatively large link weights $\sim \mathcal{U}[0.8, 1]$ to n^{δ_i} neighbors (with $\delta_1 = 0.25$, $\delta_2 = 0.1$), then spreads the remaining mass across two randomly chosen non-hub neighbors so that each row approximately sums to one. Self-links are zero. Let total world GDP be $G^{\text{World}} = 1400$. Country i's GDP G_i is proportional to its out-degree in W: $G_i = G^{\text{World}} \cdot s_i$ and $s_i = \frac{\sum_j (w_{ij} + 0.01)}{\sum_{i,k=1}^{n} (w_{ik} + 0.01)}$. We impose balanced trade targets so that exports-by-origin and imports-by-destination both equal G: ExTar = ImTar = G. Preferences follow an Armington structure with elasticity of substitution $\varrho = 5$ (Anderson and van Wincoop, 2003). First-stage parameters $\tilde{\lambda} = \left(\tilde{\lambda}_d, \tilde{\lambda}_o, \tilde{\lambda}_w\right)' = (0.05, 0.05, 0.025)'$ and $\tilde{\beta} = (-0.15, -0.05)'$ are mapped to reduced-form coefficients $\lambda^0 = (\varrho - 1)\tilde{\lambda} = (0.2, 0.2, 0.1)'$ and $\beta^0 = (1 - \varrho)\tilde{\beta} = (0.6, 0.2)'$. Let $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2] \in \mathbb{R}^{N \times 2}$ be bilateral covariates drawn from $\mathcal{U}[0, 0.75]$, and $X^U = (x_1^u, \cdots, x_n^u)' \in \mathbb{R}^{n \times 1}$ a standardized country characteristic. Second-stage loadings are $\tilde{\gamma}_o = \tilde{\gamma}_d = 0.01$ with the same scaling $\gamma = (1 - \varrho)\tilde{\gamma} = -0.04$.

Starting with this setting, the initial destination and origin fixed effects are specified by

$$\boldsymbol{\alpha}^{(0)} = X^U \gamma_o, \text{ and } \boldsymbol{\eta}^{(0)} = X^U \gamma_d.$$

Conditional on $(\boldsymbol{\alpha}^{(0)}, \boldsymbol{\eta}^{(0)})$, the initial systematic flows are determined, $\boldsymbol{\mu}^{(0)} = \exp\left(\mathbf{S}^{-1}\left(\mathbf{X}\beta + \boldsymbol{\alpha}^{(0)} \otimes \mathbf{1}_n + \mathbf{1}_n \otimes \boldsymbol{\eta}^{(0)}\right)\right)$. Given $\boldsymbol{\mu}^{(0)}$ and initial $\left(P^{(0)}, \Pi^{(0)}\right) = (\mathbf{1}_n, \mathbf{1}_n)$, we compute multilateral resistances via contraction mappings by (7):

$$\boldsymbol{\alpha}^{(\ell)} = c_0 \mathbf{1}_n + X^U \gamma_o + \log G + (\varrho - 1) \log \Pi^{(\ell-1)} \text{ and } \boldsymbol{\eta}^{(\ell)} = c_0 \mathbf{1}_n + X^U \gamma_d + \log G + (\varrho - 1) \log P^{(\ell-1)} \text{ for } \ell = 1, 2, \cdots,$$

where $c_0 = -\frac{1}{2} \log G^{\text{World}}$ and $G = (G_1, \cdots, G_n)'$. Then, we have $\boldsymbol{\mu}^{(\ell)} = \exp\left(\mathbf{S}^{-1}\left(\mathbf{X}\beta + \boldsymbol{\alpha}^{(\ell)} \otimes \mathbf{1}_n + \mathbf{1}_n \otimes \boldsymbol{\eta}^{(\ell)}\right)\right)$ for $\ell = 1, 2, \cdots$. Note that we apply the normalization $\sum_i \alpha_i^{(\ell)} - \sum_i \eta_i^{(\ell)} = 0$ for each iternation ℓ. We iterate until $\max\{\|\boldsymbol{\alpha}^{(\ell)} - \boldsymbol{\alpha}^{(\ell-1)}\|_\infty, \|\boldsymbol{\eta}^{(\ell)} - \boldsymbol{\eta}^{(\ell-1)}\|_\infty\} < 10^{-12}$. Condition in Assumption 2.4 is utilized for guaranteeing this convergence. For each simulation replication, we allow random variations in the fixed effects:

$$\boldsymbol{\alpha}^0 = \boldsymbol{\alpha}^{(\infty)} + \boldsymbol{\varepsilon}^\alpha, \text{ and } \boldsymbol{\eta}^0 = \boldsymbol{\eta}^{(\infty)} + \boldsymbol{\varepsilon}^\eta,$$

where $\boldsymbol{\varepsilon}^\alpha = (\varepsilon_1^\alpha, \cdots, \varepsilon_n^\alpha) \sim \mathcal{N}(\mathbf{0}, 0.08^2 I_n)$ and $\boldsymbol{\varepsilon}^\eta = (\varepsilon_1^\eta, \cdots, \varepsilon_n^\eta) \sim \mathcal{N}(\mathbf{0}, 0.08^2 I_n)$. Consequently, we have $\boldsymbol{\mu}^0 = \exp\left(\mathbf{S}^{-1}\left(\mathbf{X}\beta^0 + \boldsymbol{\alpha}^0 \otimes \mathbf{1}_n + \mathbf{1}_n \otimes \boldsymbol{\eta}^0\right)\right)$.

Next, we generate the error components for each simulation replication:

1. First, we generate $\xi_{ij}^* \overset{i.i.d.}{\sim} \text{Lognormal}\left(-\frac{1}{2}\sigma^2, \sigma^2\right)$ across ij with $\sigma^2 = 0.125^2$. Then, $\mathbb{E}\left(\xi_{ij}^* | \mathbf{z}\right) = \mathbb{E}\left(\xi_{ij}^*\right) = 1$.

2. Let $\epsilon_{ij}^* = \mu_{ij}^0$ for all ij. Then, $\mathbb{E}\left(\epsilon_{ij}^* | \mathbf{z}\right) = \mu_{ij}^0 \cdot \left(\mathbb{E}\left(\xi_{ij}^* | \mathbf{z}\right) - 1\right) = 0$ and $\mathsf{Var}\left(\epsilon_{ij}^* | \mathbf{z}\right) = \left(\mu_{ij}^0\right)^2 \cdot \mathsf{Var}\left(\xi_{ij}^*\right) = \left(\mu_{ij}^0\right)^2 \cdot \left(\exp(\sigma^2) - 1\right)$.

3. Last, we generate
$$u_{ij} = 0.008 \sum_{k=1}^n w_{ik}^* \epsilon_{kj}^* + 0.008 \sum_{l=1}^n w_{jl}^* \epsilon_{li}^* + 0.002 \sum_{k,l=1}^n w_{ik}^* w_{jl}^* \epsilon_{kl}^*, \tag{16}$$
where $W^* = (w_{ij}^*)$ is a row-normalized spatial weighting matrix defined by the adjacency based on W (i..e., $w_{ij}^* = \frac{\widetilde{w}_{ij}^*}{\sum_{k=1}^n \widetilde{w}_{ik}^*}$ where $\widetilde{w}_{ij}^* = \mathbb{I}\{w_{ij} + w_{ji} > 0\}$). This error structure follows (14) since $\sigma_{ij}^* = \mu_{ij}^0 \sqrt{\exp(\sigma^2) - 1}$ and $\mathbf{B} = 0.008(I_n \otimes W) + 0.008(W \otimes I_n) + 0.002(W \otimes W)$.

Basic information. Four criteria are used to evaluate the finite sample performance of the PPMLE: (i) empirical bias, (ii) empirical standard deviation (STD), (iii) standard error (s.e.), and (iv) the coverage probability of a nominal 95% confidence interval (CP). To evaluate the standard errors, we consider four kernel functions: (i) Bartlett, (ii) Parzen, (iii) Tukey-Hanning, and (iv) Quadratic Spectral (QS). For the distance measurem we first conduct the adjacency matrix $A = (a_{ij})$, $a_{ij} = \max\{\mathbb{I}(w_{ij} > 0), \mathbb{I}(w_{ji} > 0)\}$. Then, we evaluate the geodesic distance d_{ij}^*. To gauge the distance between two pairs, we consider the three types of measures:

1. Observed L^1 distance: $d_{ij,kl}^{*,1} = d_{ik}^* + d_{jl}^*$

2. Observed L^2 (Euclidean) distance: $d_{ij,kl}^{*,2} = \sqrt{(d_{ik}^*)^2 + (d_{jl}^*)^2}$

3. Observed L^∞ (max) distance: $d_{ij,kl}^{*,\infty} = \max\{d_{ik}^*, d_{jl}^*\}$

We set a bandwidth d_N to be the 25th-percentile of $\{d^*_{ij,kl}\}$. For the fixed-effect parameters, we report $\widehat{\alpha}_{49}$ and $\widehat{\eta}_{49}$ as representatives. We consider 1,000 sample repetitions. Table 1 summarizes the results.

Result interpretations. Across designs, the PPMLE performs well in finite samples. We see small upward biases in $\widehat{\lambda}_d$ and the fixed effects $(\widehat{\alpha}_{49}, \widehat{\eta}_{49})$, and mild downward biases in $\widehat{\lambda}_o$, $\widehat{\lambda}_w$, $\widehat{\beta}_1$, and $\widehat{\beta}_2$. For instance, the empirical biases are $+0.0123$ for $\widehat{\lambda}_d$ and -0.0163 for $\widehat{\lambda}_w$. These finite-sample biases attenuate as n increases; see Section 3.

For the main parameters $(\widehat{\lambda}, \widehat{\beta})$, our standard errors track the empirical standard deviations reasonably closely, albeit somewhat below them. As a result, coverage probabilities (CPs) are slightly below the 95% nominal rate. For example, under Parzen with an L^2 pair distance, the empirical STD and reported s.e. are $(0.0266, 0.0229)$ for $\widehat{\lambda}_d$ and $(0.0137, 0.0120)$ for $\widehat{\beta}_1$, with CPs of 0.933 and 0.897, respectively. Among kernel–distance combinations, we recommend the Parzen kernel with the L^2 pair distance for the main parameters, which delivers the most stable CPs across coefficients.

For the fixed-effect estimates $(\widehat{\alpha}_{49}, \widehat{\eta}_{49})$, the reported s.e. tend to be understated relative to the empirical STDs (e.g., $\widehat{\alpha}_{49}$: STD 0.0845 vs s.e. 0.0299 under Parzen–L^∞), yielding CPs around 0.86–0.90. The gap narrows with larger n (Section 3). Within our design, Parzen with an L^∞ pair distance performs best for fixed effects. Intuitively, the L^∞ metric better captures the effective dependence radius in the FE direction by guarding against long-range pairwise links, which helps reduce s.e. underestimation.

QS (with truncation) tends to under-cover at this sample size due to its long effective support, while Bartlett and Tukey–Hanning are slightly more variable than Parzen. As a rule of thumb, we recommend Parzen–L^2 for $(\widehat{\lambda}, \widehat{\beta})$ and Parzen–L^∞ for $(\widehat{\alpha}, \widehat{\eta})$. Monte Carlo uncertainty for a 95% CP with 1,000 replications is about 0.7 percentage points, so differences below $\approx$1–2 points are not statistically meaningful, whereas the 4–8 point gaps we observe are. In applications, a slightly larger bandwidth (e.g., a +5–10 percentile increase) can further mitigate undercoverage for the fixed effects; we illustrate this sensitivity in the supplement.

Table 1: Simulation results

	λ_d	λ_o	λ_w	β_1	β_2	α_{49}	η_{49}
Empirical bias	0.0123	-0.0004	-0.0163	-0.0011	-0.0001	0.0131	0.0044
Empirical STD	0.0266	0.0260	0.0355	0.0137	0.0130	0.0845	0.0850
s.e. (Bartlett, L^1)	0.0228	0.0226	0.0326	0.0117	0.0115	0.0283	0.0196
CP (Bartlett, L^1)	0.9180	0.9250	0.9050	0.9030	0.9110	0.8430	0.8290
s.e. (Bartlett, L^2)	0.0227	0.0227	0.0326	0.0117	0.0115	0.0241	0.0215
CP (Bartlett, L^2)	0.9190	0.9270	0.9120	0.9000	0.9160	0.8560	0.8630
s.e. (Bartlett, L^∞)	0.0227	0.0228	0.0326	0.0117	0.0116	0.0294	0.0224
CP (Bartlett, L^∞)	0.9180	0.9280	0.9120	0.8960	0.9070	0.8580	0.8830
s.e. (Parzen, L^1)	0.0228	0.0232	0.0330	0.0120	0.0118	0.0292	0.0216
CP (Parzen, L^1)	0.9310	0.9260	0.9200	0.9000	0.9180	0.8580	0.8710
s.e. (Parzen, L^2)	0.0229	0.0232	0.0331	0.0120	0.0118	0.0252	0.0228
CP (Parzen, L^2)	0.9330	0.9340	0.9280	0.8970	0.9180	0.8630	0.8910
s.e. (Parzen, L^∞)	0.0229	0.0233	0.0332	0.0120	0.0118	0.0299	0.0232
CP (Parzen, L^∞)	0.9320	0.9310	0.9290	0.8960	0.9160	0.8660	0.9030
s.e. (Tukey–Hanning, L^1)	0.0227	0.0228	0.0326	0.0118	0.0115	0.0284	0.0197
CP (Tukey–Hanning, L^1)	0.9170	0.9270	0.9080	0.9000	0.9140	0.8410	0.8220
s.e. (Tukey–Hanning, L^2)	0.0227	0.0229	0.0327	0.0118	0.0116	0.0242	0.0217
CP (Tukey–Hanning, L^2)	0.9140	0.9260	0.9080	0.9000	0.9170	0.8560	0.8680
s.e. (Tukey–Hanning, L^∞)	0.0227	0.0229	0.0327	0.0118	0.0116	0.0295	0.0224
CP (Tukey–Hanning, L^∞)	0.9220	0.9250	0.9130	0.8970	0.9150	0.8580	0.8840
s.e. (QS, L^1)	0.0225	0.0220	0.0320	0.0113	0.0111	0.0274	0.0173
CP (QS, L^1)	0.8880	0.9060	0.8890	0.8880	0.8890	0.8150	0.7600
s.e. (QS, L^2)	0.0224	0.0221	0.0320	0.0114	0.0112	0.0228	0.0199
CP (QS, L^2)	0.9020	0.9150	0.8940	0.8950	0.8940	0.8360	0.8300
s.e. (QS, L^∞)	0.0223	0.0219	0.0318	0.0113	0.0112	0.0287	0.0211
CP (QS, L^∞)	0.8880	0.9090	0.8900	0.8850	0.8910	0.8400	0.8470

4 Empirical Application

Motivation and basic setting. Understanding the evolution of global trade requires recognizing that international linkages are not static but shaped by major institutional, political, and economic shifts. Over time, trade agreements, policy reforms, and structural changes in the world economy have altered how countries connect and interact within the global network. These shifts provide natural breakpoints that help us examine whether and how the underlying patterns of interdependence—and the resulting network effects—change across different historical contexts.

We suspect that the structure and sources of network effects vary across four key phases of global trade: Phase 1 (1986, trade liberalization), Phase 2 (1997, active NAFTA implementation), Phase 3 (2007, emergence of the China trade shock), and Phase 4 (2016, expansion of global supply chains). The number of countries included in each phase is 136, 142, 146, and 147, respectively.[19]

[19] The full list of country names is provided in Table A.1 and A.2.

Descriptive statistics. Table 2 presents the descriptive statistics for the main variables used in the analysis across the four phases of global trade. The sample includes 136 countries in Phase 1 (1986, trade liberalization), 142 countries in Phase 2 (1997, active NAFTA implementation), 146 countries in Phase 3 (2007, emergence of the China trade shock), and 147 countries in Phase 4 (2016, expansion of global supply chains). The statistics reveal substantial variation in trade volumes, reflecting the evolution of global trade integration and the expansion of international linkages over time.

The mean logged bilateral trade flow (y_{ij}), measured in 2015 constant U.S. dollars, increases markedly from Phase 1 to Phase 3 before stabilizing in Phase 4. This pattern reflects the steady expansion of global trade and the progressive deepening of international production networks. Meanwhile, the share of zero trade flows declines over time—from roughly 53% in Phase 1 to about 28% in Phase 4—indicating that more country pairs established trading relationships as global markets became increasingly interconnected.

The time-invariant bilateral variables offer additional insights into trade frictions and integration patterns. The mean logged *Distance* between trading partners remains stable across periods, as expected, while the *Border* variable indicates that only a small fraction of pairs share a common border, underscoring the predominance of long-distance trade relationships. Institutional and cultural similarities vary moderately over time: the proportion of country pairs sharing the same legal system (*Legal*) or language (*Language*) remains around 30–37%, suggesting persistent institutional diversity. Colonial ties (*Colony*) and common currency arrangements (*Currency*) are rare and relatively unchanged across phases, while the share of country pairs classified as islands or landlocked (*Islands*, *Landlock*) remains stable, reflecting enduring geographic constraints on trade.

Finally, the incidence of regional trade agreements (*FTA*) rises gradually across phases, from 0.04% in 1986 to 1.0% in 2016, capturing the growing prevalence of formal trade cooperation. Together, these patterns highlight the dynamic transformation of the global trading system—from one dominated by fragmented and geographically limited exchanges to one characterized by broad participation, institutional integration, and extensive cross-border linkages.

To operationalize these phases, we constructed the spatial weighting matrix (W) as the average trade flows between exporters and importers in the years immediately preceding each referential year. This approach captures the stability of trade connections among countries while anchoring each phase to a major historical turning point in global trade. Specifically, W is based on trade flows from 1984–1985 for Phase 1 (as data collection begins in 1984), 1993–1996 for Phase 2, 2000–2006 for Phase 3, and 2010–2015 for Phase 4.

Table 3: Network statistics for the countries' connectivity matrix

	linear-in-mean	Bipartite	Phase 1	Phase 2	Phase 3	Phase 4
$\varphi_{(2)}$	-0,0067	$\simeq 0$	0.5413	0.5653	0.5752	0.5553
$\varphi_{\min}$	-0.0067	-1	-0.5296	-0.5151	-0.5088	-0.4419
Density	1	0.5034	0.4948	0.7560	0.8910	0.9105

Note: Phase 1 (1986, trade liberalization), Phase 2 (1997, active NAFTA implementation), Phase 3 (2007, emergence of the China trade shock), and Phase 4 (2016, expansion of global supply chains). $\varphi_{(2)}$ denotes the second-largest eigenvalue, $\varphi_{\min}$ denotes the smallest eigenvalue, and *Density* represents the proportion of nonzero elements (edges) in the network. The *linear-in-means* network follows Manski (1993), while the *bipartite* network represents a perfectly polarized structure in which nodes are divided into two mutually connected groups with no intra-group links.

Table 2: Descriptive statistics

Phase	1			2			3			4		
	Mean	Median	STD	Mean	Median	STD	Mean	Median	STD	Mean	Median	STD
y_{ij}	201,347	0	2,604,211	380,573	382	4,151,519	753,486	1,127	7,055,185	718,068	865	7,174,976
(Zero freq.)	(0.5307)			(0.3007)			(0.2488)			(0.2835)		
y_{ij}^{+}	429,005	8,105	3,788,568	544,197	4,726	4,955,465	1,003,100	7,249	8,125,016	1,002,158	8,909	8,459,553
Distance	0.2898	0.2635	0.1821	0.2864	0.2598	0.1801	0.2853	0.2589	0.1796	0.2829	0.2543	0.1787
Border	0.0184	0.0000	0.1344	0.0188	0.0000	0.1357	0.0190	0.0000	0.1365	0.0186	0.0000	0.1352
Legal	0.3760	0.0000	0.4844	0.3629	0.0000	0.4808	0.3623	0.0000	0.4807	0.3673	0.0000	0.4821
Language	0.3254	0.0000	0.4685	0.3100	0.0000	0.4625	0.3027	0.0000	0.4594	0.3035	0.0000	0.4598
Colony	0.0117	0.0000	0.1073	0.0110	0.0000	0.1042	0.0105	0.0000	0.1019	0.0104	0.0000	0.1016
Currency	0.0109	0.0000	0.1038	0.0100	0.0000	0.0994	0.0094	0.0000	0.0967	0.0093	0.0000	0.0961
Islands	0.3824	0.0000	0.5541	0.3662	0.0000	0.5450	0.3699	0.0000	0.5472	0.3673	0.0000	0.5457
Landlock	0.3088	0.0000	0.5091	0.3099	0.0000	0.5099	0.3151	0.0000	0.5134	0.2993	0.0000	0.5028
FTA	0.0004	0.0000	0.0209	0.0010	0.0000	0.0316	0.0060	0.0000	0.0775	0.0107	0.0000	0.1030

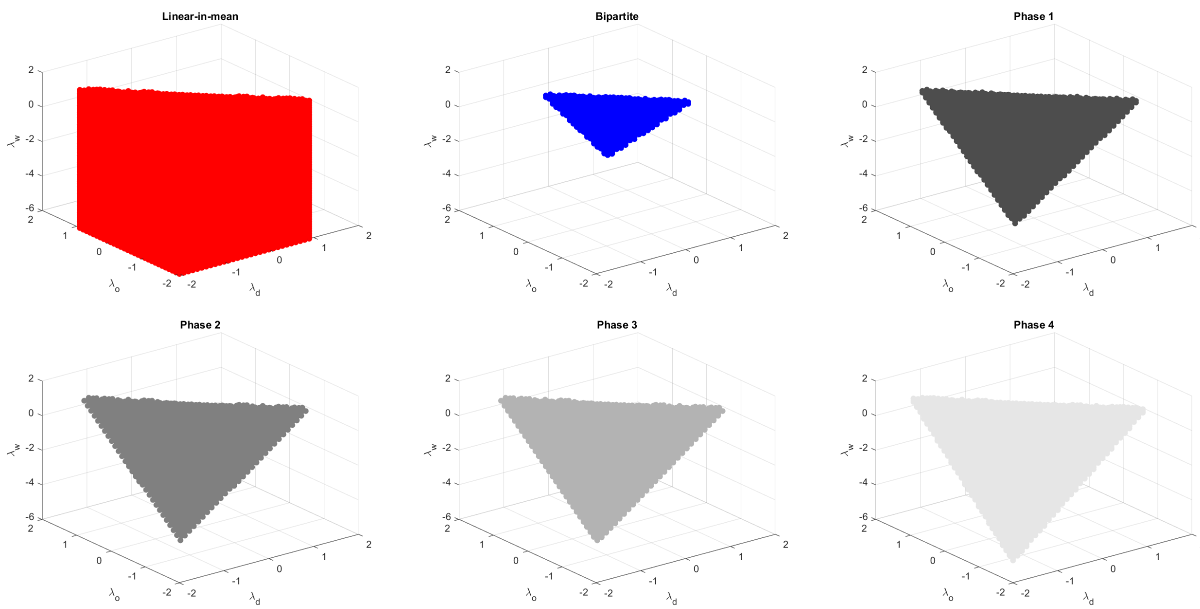
Note: STD stands for the standard deviation. Phase 1 (1986, trade liberalization), Phase 2 (1997, active NAFTA implementation), Phase 3 (2007, emergence of the China trade shock), and Phase 4 (2016, expansion of global supply chains). The definitions of explanatory variables are adopted from Helpman et al. (2008) as follows: *Distance:* the distance between importer i's and exporter j's capitals; *Border:* a binary variable that equals one if country j and country i are neighbors that meet a common physical boundary, and zero otherwise; *Legal:* a binary variable that equals one if country j and country i share the same legal system, and zero otherwise; *Language:* a binary variable that equals one if country i and j share the same language system, and zero otherwise; *Colony:* a binary variable that equals one if country j ever colonized country i or vice versa, and zero otherwise; *Currency*: a binary variable that equals one if the country j and country i use the same currency or if within the country pair money was interchangeable at a 1:1 exchange rate for an extended period of time; *Islands:* a binary variable that equals one if both importer i and exporter j are islands, and zero otherwise; *Landlock:* a binary variable that equals one if both exporting country j and importing country i have no coastline or direct access to sea, and zero otherwise; *FTA:* a binary variable that equals one if country j and country i belong to a common regional trade agreement, and zero otherwise.

Figure 3 visualizes the admissible parameter spaces for λ across different network structures and phases. The admissible space is widest under the linear-in-means network (i.e., a uniform connection structure), indicating that the corresponding λ values are less restrictive and may capture more diffuse or noisy network effects. In contrast, the bipartite network—which represents a highly polarized structure—exhibits the narrowest admissible parameter space, reflecting its rigidity and limited capacity for capturing intermediate forms of interdependence. Our spatial weighting matrices (W) across the four phases lie between these two extremes, suggesting that our estimated trade networks are neither uniformly dense nor trivial or polarized. Instead, they occupy an intermediate region of the parameter space, consistent with networks that evolve over time to reflect varying degrees of connectivity, clustering, and heterogeneity in global trade relationships.

To better understand the role of W, we describe its main properties and their implications. Since W is row-normalized, its maximum eigenvalue is naturally equal to 1. Because the matrix also has zero diagonal elements, its minimum eigenvalue must be negative, though still greater than -1. Also, when the number of observations is large, having a minimum eigenvalue far from zero indicates that the network contains substantial heterogeneity. In this case, the weighted influence of neighboring units does not collapse into something close to a constant (such as a uniform average across units), but instead reflects meaningful variation driven by the structure of the network. In our case, W remain substantially distant from zero even with a large number of countries, implying that the network encodes strong and diverse interdependencies across countries rather than converging to a homogeneous or trivial structure.

The sparsity of W declines markedly over time, falling from 0.4021 in Phase 1 to 0.0583 in Phase 4. This trend indicates that the global trade network has become substantially more connected, with countries increasingly linked through trade relationships. A denser matrix not only reflects the expansion and diversification of trading partners but also suggests greater potential for leveraging network effects, whereby gains or shocks in one country can more readily spread across the system, amplifying both opportunities and vulnerabilities in global trade.

Figure 3: Admissible parameter spaces for λ



Estimation results. Table 4 presents the estimation results, describing the evolution of network effects across phases. The estimates of the network parameters, λ_d, λ_o, and λ_w, reveal how the structure of trade interdependence evolved across four distinct phases of globalization and regional integration. In Phase 1 (1986, early trade liberalization), both λ_d and λ_o are negative and statistically significant, indicating that trade flows sharing the same exporter or the same importer behaved as *substitutes*. This suggests that, in the early stage of market liberalization, exporters faced limited capacity or market-access constraints, such that expanding exports to one destination crowded out exports to others. Similarly, importers reallocated their demand across competing sources of supply. At this stage, network competition dominated network complementarity.

In Phase 2 (1997, active NAFTA implementation), both λ_d and λ_o turn positive, reflecting a structural shift toward *complementarity* among trade flows. Regional integration under NAFTA reduced trade frictions and deepened production linkages across member countries, promoting mutual reinforcement among export destinations and import sources. The positive signs of λ_d and λ_o thus capture the emergence of regional value chains and increasing returns to network connectivity.

In Phase 3 (2007, the emergence of the China trade shock), λ_d and λ_o again become negative. The sharp rise of China as a global exporter introduced strong competitive pressures in world markets, leading to substitution effects across both exporters and destinations. Exporters increasingly competed for global market share, while importers rebalanced sourcing patterns in response to China's dominance. This phase marks a reemergence of network competition and market crowding-out effects within the trade system.

By Phase 4 (2016, expansion of global supply chains), λ_d and λ_o move closer to zero, implying a stabilization of trade interdependencies under mature global value chains. Competition and complementarity coexist, reflecting more stable network structures shaped by deep integration of intermediate inputs and production networks.

Across all phases, λ_w remains positive and highly significant, indicating persistent and robust *third-party spillover effects*. Trade flows between two countries are positively associated with those among third-country pairs, capturing the complementarity and propagation of trade through multilateral linkages. The magnitude of λ_w rises notably during Phase 3, consistent with the expansion of cross-country production sharing and the consolidation of China's central position in global value chains. This pattern suggests that, while bilateral trade relationships may exhibit competitive substitution, the broader trade network increasingly amplifies interdependence through global production and supply-chain connectivity.

Beyond the network parameters, the coefficient on *Distance* remains negative and highly significant across all phases, confirming that geographic separation continues to impose substantial trade costs even in an increasingly interconnected global economy. The magnitude of the distance elasticity becomes larger over time, particularly during Phase 4, suggesting that despite technological advances and reduced communication costs, spatial frictions persist in shaping international trade patterns.

The *Border* variable is positive and significant throughout all phases, indicating that countries sharing a common border trade more intensively than others, likely due to reduced transportation costs and institutional proximity. Similarly, sharing the same *Legal* system increases bilateral trade volumes, reflecting the role of institutional similarity in lowering transaction costs and enhancing contract enforcement. The impact of *Language* becomes positive and statistically significant in later phases, particularly during Phase 3, implying that cultural and informational frictions gain importance as global competition intensifies and trade networks expand into more diverse markets.

Other structural and historical factors show heterogeneous patterns. The effect of *Colony* is small and mostly insignificant, suggesting that historical colonial ties have weakened over time as new trade alliances emerged. In contrast, the coefficient on *FTA* is positive and significant across all phases, confirming the trade-creating effect of regional trade agreements such as NAFTA. The positive and persistent influence of *FTA* underscores the continued relevance of policy-driven integration alongside endogenous network formation.

Geographic constraints, captured by the *Islands* and *Landlock* variables, exhibit more nuanced patterns. Island countries tend to trade less during the early phases but show substantial trade increases in later phases, particularly during Phase 3, when global value chains expanded maritime trade routes. Landlocked countries, by contrast, consistently face trade disadvantages, though their negative effects diminish in the later phases, possibly reflecting improved regional infrastructure and logistics integration. Finally, the *Currency* variable shows mixed signs and lacks statistical significance in most phases, consistent with the limited prevalence of currency unions in the sample.

Counterfactual analysis. Based on the substantial variations in spillover patterns, we investigate amplified responsiveness across the four phases of global trade. The counterfactual analysis highlights three key measures—the *network multiplier*, *average improvement*, and *heterogeneity*—which together describe how trade shocks and policies propagate through the global network.

The *network multiplier* measures the total strength of amplification within the trade network, quantifying the overall degree of amplification embedded in the trade network. If the multiplier is close to one, network effects are weak—each bilateral change mainly affects its own trade pair. If the multiplier exceeds one, a shock to one link spreads through the network, magnifying its total impact on global trade flows. A higher multiplier thus indicates a more interdependent and responsive global trade system. Specifically, the *network multiplier* corresponds to an element of $\mathbf{S}^{-1}\mathbf{1}_N$. Since W is row-normalized, all elements of $\mathbf{S}^{-1}\mathbf{1}_N$ are identical (homogeneous).

Table 4: Estimation Results by Phase

Phase	1	2	3	4
λ_d	-0.1724**	0.1883**	-0.3370**	-0.0215
	(0.0891)	(0.0844)	(0.1645)	(0.1383)
λ_o	-0.2761**	0.2504***	-0.3343***	0.0895
	(0.1268)	(0.0766)	(0.1247)	(0.1467)
λ_w	0.3743*	0.2728***	0.6249**	0.4566***
	(0.2102)	(0.0582)	(0.3135)	(0.0982)
Distance	-1.4799**	-2.0462***	-1.8372***	-2.2233***
	(0.7555)	(0.4078)	(0.4482)	(0.4437)
Border	1.1377***	1.0922***	0.3338	1.0085***
	(0.3802)	(0.2529)	(0.3826)	(0.1796)
Legal	0.2644***	0.217***	0.3752***	0.1260
	(0.0696)	(0.0721)	(0.1007)	(0.0849)
Language	0.0901	-0.1003	0.3500***	0.0001
	(0.1140)	(0.0794)	(0.0681)	(0.0825)
Colony	0.0701	0.0549	-0.284	0.1904*
	(0.1403)	(0.1151)	(0.2158)	(0.1156)
Currency	0.5330	-0.1882	0.0936	-0.1512
	(0.3559)	(0.5164)	(0.3299)	(0.3499)
Islands	9.3548*	-4.4006***	10.4907**	2.2854***
	(5.3965)	(0.0970)	(4.3980)	(0.6494)
Landlock	3.6492	-0.4001***	3.5766	0.1563
	(10.5348)	(0.1092)	(10.4987)	(0.0987)
FTA	0.7785*	0.4941***	0.2094*	0.5098***
	(0.2586)	(0.1866)	(0.1270)	(0.1044)
# of observations	18,360	20,022	21,170	21,462
Log-likelihood	54,499,334,220	115,633,972,489	246,566,921,010	240,363,503,195

Note: Significance levels: *** (1%), ** (5%), * (10%). Standard errors are evaluated by the Parzen kernel with the L^2-based distance, and are reported in parentheses. Phase 1 (1986, trade liberalization), Phase 2 (1997, active NAFTA implementation), Phase 3 (2007, emergence of the China trade shock), and Phase 4 (2016, expansion of global supply chains). The definitions of explanatory variables are adopted from Helpman et al. (2008) as follows: *Distance:* the distance between importer i's and exporter j's capitals; *Border:* a binary variable that equals one if country j and country i are neighbors that meet a common physical boundary, and zero otherwise; *Legal:* a binary variable that equals one if country j and country i share the same legal system, and zero otherwise; *Language:* a binary variable that equals one if country i and j share the same language system, and zero otherwise; *Colony:* a binary variable that equals one if country j ever colonized country i or vice versa, and zero otherwise; *Currency*: a binary variable that equals one if the country j and country i use the same currency or if within the country pair money was interchangeable at a 1:1 exchange rate for an extended period of time; *Islands:* a binary variable that equals one if both importer i and exporter j are islands, and zero otherwise; *Landlock:* a binary variable that equals one if both exporting country j and importing country i have no coastline or direct access to sea, and zero otherwise; *FTA:* a binary variable that equals one if country j and country i belong to a common regional trade agreement, and zero otherwise.

The *average improvement* quantifies the average percentage change in predicted trade flows when network interdependence is fully activated, indicating how much trade increases relative to a benchmark where network effects are absent. Specifically, the *average improvement* is defined as

$$\frac{1}{n(n-1)} \sum_{i,j=1}^{n} \frac{\widehat{\mu}_{ij}}{\widetilde{\mu}_{ij}},$$

where $\widehat{\mu}_{ij}$ represents predicted trade flow from j to i using the PPML estimates and $\widetilde{\mu}_{ij}$ represents predicted trade flow from j to i using the PPML estimates with $\lambda_d = \lambda_o = \lambda_w = 0$ imposed.

The *heterogeneity* measure compares the dispersion of trade improvements across country pairs, showing how unevenly the benefits of network effects are distributed—larger values suggest that gains are concentrated among specific countries or trade routes, whereas smaller values imply more uniform effects. Specifically, the *heterogeneity* measure is given by

$$\frac{\mathrm{STD}(\widehat{\mu}_{ij})}{\mathrm{STD}(\widetilde{\mu}_{ij})} = \frac{\sqrt{\sum_{i,j=1}^{n} \left(\widehat{\mu}_{ij} - \overline{\widehat{\mu}}_{ij}\right)^2}}{\sqrt{\sum_{i,j=1}^{n} \left(\widetilde{\mu}_{ij} - \overline{\widetilde{\mu}}_{ij}\right)^2}}.$$

Table 5 shows that differences in spillover patterns shape how trade policies transmit through the global network. During the early liberalization period (Phase 1), all three measures are modest, consistent with limited network propagation and relatively uniform gains from liberalization. Under active NAFTA implementation (Phase 2), all indicators rise sharply, suggesting that regional integration substantially magnified spillover effects—changes in one country generated large and uneven responses across the network. During the China trade shock (Phase 3), the measures are moderate, indicating that intensified global competition dampened propagation effects and reduced disparities in trade responses. By the expansion of global supply chains (Phase 4), the multiplier and average improvement rise again, reflecting renewed amplification of network effects, while heterogeneity remains elevated. This pattern suggests that gains from interdependence became increasingly uneven across trade pairs, though less pronounced than in Phase 2, consistent with a more mature and stabilized global value chain structure.

Table 5: Statistics from the counterfactual analysis

Phase	1	2	3	4
Network multiplier	0.9310	3.4656	0.9556	2.1035
Average improvement	1.9325	17270.6262	5.4500	2965.1577
Heterogeneity	0.2193	122706.2996	0.1839	5086.2405

Note: Phase 1 (1986, trade liberalization), Phase 2 (1997, active NAFTA implementation), Phase 3 (2007, emergence of the China trade shock), and Phase 4 (2016, expansion of global supply chains). The three measures—the network multiplier, average improvement, and heterogeneity—capture structural shifts in global trade interdependence. The *network multiplier* measures the total strength of amplification within the trade network, quantifying the overall degree of amplification embedded in the trade network. The *average improvement* represents the overall percentage increase in predicted trade flows when network interdependence is fully activated, capturing the magnitude of network-induced trade gains relative to a benchmark scenario without such effects. The *heterogeneity* measure compares the dispersion of these trade improvements across country pairs, reflecting how unevenly the benefits of interdependence are distributed.

5 Conclusion

The gravity equation has long served as a foundational framework for analyzing origin–destination flows, tracing back to Isard (1954) and Tinbergen (1962). Although these early formulations were groundbreaking in highlighting the role of geographic distance in trade, they remained within an exogenous framework: trade costs were treated as predetermined, and trade flows were assumed to be independent, overlooking the multilateral interactions and network linkages that characterize modern global trade. This bilateral focus isolated pairwise effects—such as shared borders, common languages, or trade agreements—but failed to capture how indirect and higher-order connections among countries jointly shaped trade outcomes. Recognizing the growing importance of trade networks, we developed a microfoundation-based specification and econometric framework that endogenized trade costs as a function of network structure, thereby making them both endogenous and interdependent, which in turn amplified heterogeneity across trade pairs.

From a theoretical standpoint, we derived the specification from microfoundations and demonstrated that it could be represented as a spatial autoregressive model that naturally captured multilateral interdependence. Our framework departed from the traditional iceberg-cost assumption by allowing trade costs to evolve through network linkages, in which countries leveraged their trade connections as a resource. This approach contrasted with conventional gravity models, which assumed predetermined and symmetric trade costs. In practice, our specification captured how connectivity, competition, and spillovers jointly determined trade frictions and outcomes.

Methodologically, we extended the Poisson Pseudo Maximum Likelihood Estimator (Gourieroux et al., 1984) to account for network dependence and proposed heteroskedasticity- and autocorrelation-robust standard errors to ensure valid inference in the presence of dominant units and arbitrary correlation in the error structure. We addressed the limitations of existing spatial gravity models, including their lack of microfoundations, reliance on log-linearized specifications, and inability to handle zero trade flows. By formulating the model in its original level, we overcame the log-transformation issue, accommodated zero flows, and formally established the conditions for a unique equilibrium. We also characterized the properties of the spatial weight matrix that revealed the network connectivity structure, linked these properties to the eigenvalue spectrum, and derived the uniqueness condition for the network-based trade flow equilibrium.

Empirically, we identified distinct patterns of network effects across four key phases of global trade. During early liberalization, network propagation was limited and trade gains were relatively uniform. Under NAFTA, spillover effects intensified, producing large but uneven trade improvements. The rise of China reintroduced competition and substitution among trade links, while the subsequent expansion of global supply chains led to a more stable yet persistently heterogeneous network structure. The counterfactual analysis further demonstrated that network effects amplified trade and welfare gains, with their magnitude and distribution varying across different phases of globalization. These results contributed to the growing recognition of correlated and interdependent trade flows (Lind and Moreno, 2023) and moved beyond the traditional iceberg-cost paradigm by explicitly modeling how trade networks functioned as a resource in shaping global trade patterns.

As a final remark, although our primary focus was the gravity equation, our proposed framework is broadly applicable to a wider class of constant elasticity models up to their semi-reduced forms (i.e., up to constant elasticity models with individual fixed effects). While the detailed structure of individual fixed effects may differ across specific constant elasticity systems, the core principle of endogenizing trade costs through network interactions remained effective.

Appendix

Table A.1: Countries List (1)

	Phase					Phase			
Countries	1	2	3	4	Countries	1	2	3	4
Afghanistan	NA	NA	*	*	China	*	*	*	*
Albania	*	*	*	*	Colombia	*	*	*	*
Algeria	*	*	*	*	Comoros	*	*	*	*
Angola	*	*	*	*	Costa Rica	*	*	*	*
Argentina	*	*	*	*	Cuba	*	*	*	*
Australia	*	*	*	*	Cyprus	*	*	*	*
Austria	*	*	*	*	Czechia	NA	*	*	*
Bahamas	*	*	*	*	Democratic Republic of the Congo	*	*	*	*
Bahrain	*	*	*	*	Denmark	*	*	*	*
Bangladesh	*	*	*	*	Djibouti	NA	NA	NA	*
Barbados	*	*	*	*	Dominican Republic	*	*	*	*
Belgium	NA	NA	*	*	Ecuador	*	*	*	*
Belize	*	*	*	*	Egypt	*	*	*	*
Benin	*	*	*	*	El Salvador	*	*	*	*
Bermuda	*	*	*	*	Equatorial Guinea	*	*	*	*
Bhutan	*	*	*	*	Ethiopia	*	*	*	*
Bolivia	*	*	*	*	Fiji	*	*	*	*
Brazil	*	*	*	*	Finland	*	*	*	*
Brunei	*	*	*	*	France	*	*	*	*
Bulgaria	*	*	*	*	French Guiana	NA	NA	NA	NA
Burkina Faso	*	*	*	*	Gabon	*	*	*	*
Burundi	*	*	*	*	Gambia	*	*	*	*
Cote D'Ivoire	*	*	*	*	Germany	*	*	*	*
Cambodia	*	*	*	*	Ghana	*	*	*	*
Cameroon	*	*	*	*	Greece	*	*	*	*
Canada	*	*	*	*	Greenland	*	*	*	*
Cayman Islands	NA	NA	*	*	Guadeloupe	NA	NA	NA	NA
Central African Republic	*	*	*	*	Guatemala	*	*	*	*
Chad	*	*	*	*	Guinea	*	*	*	*
Chile	*	*	*	*	Guinea-Bissau	*	*	*	*

Table A.2: Countries List (2)

Countries	Phase 1	2	3	4	Countries	Phase 1	2	3	4	Countries	Phase 1	2	3	4
Guyana	*	*	*	*	Mauritius	*	*	*	*	Saint Kitts and Nevis	*	*	*	*
Haiti	*	*	*	*	Mexico	*	*	*	*	Saudi Arabia	*	*	*	*
Honduras	*	*	*	*	Mongolia	*	*	*	NA	Senegal	*	*	*	*
Hong Kong	*	*	*	*	Morocco	*	*	*	*	Serbia	NA	NA	*	*
Hungary	*	*	*	*	Mozambique	*	*	*	*	Seychelles	*	*	*	*
Iceland	*	*	*	*	Myanmar	*	*	*	*	Sierra Leone	*	*	*	*
India	*	*	*	*	Nepal	*	*	*	*	Singapore	*	*	*	*
Indonesia	*	*	*	*	Netherlands	*	*	*	*	Solomon Islands	*	*	*	*
Iran	*	*	*	*	Netherlands Antilles	NA	NA	NA	NA	Somalia	*	*	*	*
Iraq	*	*	*	*	New Caledonia	NA	NA	NA	NA	South Africa	*	*	*	*
Ireland	*	*	*	*	New Zealand	*	*	*	*	South Korea	*	*	*	*
Israel	*	*	*	*	Nicaragua	*	*	*	*	Spain	*	*	*	*
Italy	*	*	*	*	Niger	*	*	*	*	Sri Lanka	*	*	*	*
Jamaica	*	*	*	*	Nigeria	*	*	*	*	Sudan	*	*	*	*
Japan	*	*	*	*	North Korea	NA	NA	NA	NA	Suriname	*	*	*	*
Jordan	*	*	*	*	Norway	*	*	*	*	Sweden	*	*	*	*
Kenya	*	*	*	*	Oman	*	*	*	*	Switzerland	*	*	*	*
Kiribati	*	*	*	*	Pakistan	*	*	*	*	Syria	*	*	*	*
Kuwait	*	*	*	*	Panama	*	*	*	*	Taiwan	NA	NA	NA	NA
Laos	*	*	*	*	Papua New Guinea	*	*	*	*	Tanzania	*	*	*	*
Lebanon	NA	*	*	*	Paraguay	*	*	*	*	Thailand	*	*	*	*
Liberia	*	*	*	*	Peru	*	*	*	*	Togo	*	*	*	*
Libya	*	*	*	*	Philippines	*	*	*	*	Trinidad and Tobago	*	*	*	*
Madagascar	*	*	*	*	Poland	NA	*	*	*	Tunisia	*	*	*	*
Malawi	*	*	*	*	Portugal	*	*	*	*	Turkey	*	*	*	*
Malaysia	*	*	*	*	Qatar	*	*	*	*	Turks and Caicos Islands	NA	NA	NA	*
Maldives	*	*	*	*	Reunion	NA	NA	NA	NA	Uganda	*	*	*	*
Mali	*	*	*	*	Romania	NA	*	*	*	United Arab Emirates	*	*	*	*
Malta	*	*	*	*	Russia	NA	*	*	*	United Kingdom	*	*	*	*
Mauritania	*	*	*	*	Rwanda	*	*	*	*	United States	*	*	*	*
										Uruguay	*	*	*	*
										Venezuela	NA	NA	NA	NA
										Vietnam	*	*	*	*
										Western Sahara	NA	NA	NA	NA
										Yemen	NA	*	*	*
										Zambia	*	*	*	*

References

Anderson, J. (1979). A theoretical foundation for the gravity equation. *The American economic review*, 69(1):106–116.

Anderson, J. and van Wincoop, E. (2003). Gravity with gravitas: a solution to the border puzzle. *American Economic Review*, 93:170–192.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858.

Behrens, K., Ertur, C., and W., K. (2012). 'dual' gravity: using spatial econometrics to control for multilateral resistance. *Journal of Applied Econometrics*, 27:773–794.

Bramoullé, Y., Kranton, R., and D'Amours, M. (2014). Strategic interaction and networks. *American Economic Review*, 104(3):898–930.

Chaney, T. (2008). Distorted gravity: the intensive and extensive margins of international trade. *American Economic Review*, 98(4):1707–21.

Chen, J. and Roth, J. (2024). Logs with zeros? some problems and solutions. *The Quarterly Journal of Economics*, 139(2):891–936.

Chen, M., Fernandez-Val, I., and Weidner, M. (2021). Nonlinear factor models for network and panel data. *Journal of Econometrics*, 220:296–324.

Chung, F. R. (1997). *Spectral graph theory*, volume 92. American Mathematical Soc.

Cliff, A. D. and Ord, J. (1995). *Spatial Autocorrelation*. Pion Ltd, London.

Conley, T., Gonçalves, S., Kim, M., and Peron, B. (2023). Bootstrap inference under cross-sectional dependence. *Quantitative Economics*, 14:511–569.

Eaton, J. and Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5):1741–1779.

Fernandez-Val, I. and Weidner, M. (2016). Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics*, 192:291–312.

Fernandez-Val, I. and Weidner, M. (2018). Fixed effects estimation of large-t panel data models. *Annual Review of Economics*, 10:109–138.

Gourieroux, C., Monfort, A., and Trognon, A. (1984). Pseudo maximum likelihood methods: applications to poisson models. *Econometrica*, 52:701–720.

Head, K. and Mayer, T. (2014). Chapter 3 - gravity equations: Workhorse,toolkit, and cookbook. In Gopinath, G., Helpman, E., and Rogoff, K., editors, *Handbook of International Economics*, volume 4 of *Handbook of International Economics*, pages 131–195. Elsevier.

Helpman, E. and Krugman, P. (1985). *Market Structure and Foreign Trade*. Cambridge, MA: The MIT Press.

Helpman, E., Melitz, M., and Rubinstein, Y. (2008). Estimating trade flows: Trading partners and trading volumes. *The Quarterly Journal of Economics*, 123(2):441–487.

Isard, W. (1954). Location theory and trade theory: Short-run analysis. *Quarterly Journal of Economics*, 68(2):305–320.

Jenish, N. and Prucha, I. (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics*, 150:86–98.

Jenish, N. and Prucha, I. (2012). On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics*, 170:178–190.

Jeong, H. and Lee, L. (2024). Maximum likelihood estimation of a spatial autoregressive model for origin-destination flow variables. *Journal of Econometrics*, 242:105790.

Jeong, H., Lin, Y., and Lee, L. (2023). Estimation of spatial autoregressive models for origin-destination flows: A partial likelihood approach. *Economics Letters*, 229:111202.

Jin, F., Lee, L., and Yu, J. (2023). Estimating flow data models of international trade: Dual gravity and spatial interactions. *Econometric Reviews*, 42:157–194.

Kapetanios, G., Serlenga, L., and Shin, Y. (2021). Estimation and inference for multi-dimensional heterogeneous panel datasets with hierarchical multi-factor error structure. *Journal of Econometrics*, 220(2):504–531. Annals Issue: Celebrating 40 Years of Panel Data Analysis: Past, Present and Future.

Kelejian, H. and Prucha, I. (2007). HAC estimation in a spatial framework. *Journal of Econometrics*, 140:131–154.

Kim, M. and Sun, Y. (2011). Spatial heteroskedasticity and autocorrelation consistent estimation of covariance matrix. *Journal of Econometrics*, 160:349–371.

Krugman, P. (1995). *Handbook of International Economics - Increasing returns, imperfect competition and the positive theory of international trade*, volume 3. Elsevier.

Kwon, O., Yoon, J., and Yotov, Y. V. (2025). A generalized poisson-pseudo maximum likelihood estimator*. *Journal of Business & Economic Statistics*, 0(ja):1–27.

Lee, L. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models. *Econometrica*, 72:1899–1925.

Lee, L. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics*, 137:489–514.

Lee, L. and Yu, J. (2010). A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory*, 26:564–597.

LeSage, J. and Fischer, M. (2010). *Spatial econometric methods for modeling origin-destination flows*. In: Fischer M., Getis A. (eds) Handbook of Applied Spatial Analysis. Springer, Berlin, Heidelberg.

LeSage, J. and Pace, R. (2008). Spatial econometric modeling of origin-destination flows. *Journal of Regional Science*, 48:941–967.

Lind, N. and Ramondo, N. (2023). Trade with correlation. *American Economic Review*, 113(2):317–53.

McCallum, J. (1995). National borders matter: Canada-u.s. regional trade patterns. *American Economic Review*, 85:615–623.

Melitz, M. and Ottaviano, G. (2008). Market size, trade, and productivity. *Review of Economic Studies*, 75(1):295–316.

Melitz, M. J. (2003). *The impact of trade on intra-industry reallocations and aggregate industry productivity.*, volume 71(6). econometrica.

Mullahy, J. and Norton, E. (2024). Why transform Y? A the pitfalls of transformed regressions with a mass at zero. *Oxford Bulletin of Economics and Statistics*, 86:417 – 447.

Nagengast, A. J. and Yotov, Y. V. (2025). Staggered difference-in-differences in gravity settings: Revisiting the effects of trade agreements. *American Economic Journal: Applied Economics*, 17(1):271–96.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708.

Ord, J. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70:120–126.

Pesaran, M. H. and Yang, C. F. (2020). Econometric analysis of production networks with dominant units. *Journal of Econometrics*, 219(2):507–541. Annals Issue: Econometric Estimation and Testing: Essays in Honour of Maxwell King.

Pesaran, M. H. and Yang, C. F. (2021). Estimation and inference in spatial models with dominant units. *Journal of Econometrics*, 221(2):591–615.

Qu, X. and Lee, L. (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics*, 184:209–232.

Samuelson, P. (1952). The transfer problem and transport costs: The terms of trade when impediments are absent. *Economic Journal*, pages 265–289.

Samuelson, P. (1954). The transfer problem and trasponrt costs ii: Analysis of effects of trade impediments. *Economic Journal*, pages 265–289.

Santos Silva, J. and Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics*, 88:641–658.

Santos Silva, J. and Tenreyro, S. (2022). The log of gravity at 15. *Portuguese Economic Journal*, 21:423–437.

Tinbergen, J. (1962). *Shaping the World Economy.* Net York: The Twentieth Century Fund.

Weidner, M. and Zylkin, T. (2021). Bias and consistency in three-way gravity models. *Journal of International Economics*, 132:103513.

Xu, X. and Lee, L. (2015a). Maximum likelihood estimation of a spatial autoregressive Tobit model. *Journal of Econometrics*, 188:264–280.

Xu, X. and Lee, L. (2015b). A spatial autoregressive model with a nonlinear transformation of the dependent variable. *Journal of Econometrics*, 186:1–18.

Xu, X. and Lee, L. (2018). Sieve maximum likelihood estimation of the spatial autoregressive Tobit model. *Journal of Econometrics*, 203:96–112.

Supplement to "Specification and Estimation of Spatial Interaction Models for Nonnegative Origin-destination Flows with Dominant Units"

Hanbat Jeong* and Jieun Lee†

Abstract

This document contains some technical proofs, additional MC, and empirical results for Jeong and Lee (2025). Section 1 reviews previous findings and provides further interpretations of the model specification. Sections 1.1 and 1.2 examine issues with the log-transformed specification in the existing literature. Section 1.3 then reviews and extends the conventional gravity equations into a spatial-gravity framework, followed by a detailed interpretation of our model. Section 2 provides the theoretical framework of our model, with Section 2.1 outlining the first- and second-order conditions, Section 2.2 detailing the NED properties, and Section 2.3 discussing the asymptotic distribution, bias, and variance estimation.

1 Discussion on Model Specification and Its Implications

1.1 Log-transformation

In this subsection, we summarize and extend the previous findings. For simplicity, consider the stochastic version of a simple constant elasticity model and assume $\dim(x_{ij}) = 1$:

$$y_{ij} = \underbrace{\exp\left(\beta_0^0 + \beta_1^0 x_{ij}\right)}_{=\mu_{ij}=\mathbb{E}(y_{ij}|x_{ij})} \cdot \xi_{ij} \Leftrightarrow y_{ij} = \mu_{ij} + u_{ij}, \text{ where } u_{ij} = \mu_{ij}(\xi_{ij} - 1), \tag{1}$$

and β_0^0 and β_1^0 are the main parameters of interests.

To estimate β_0^0 and β_1^0, we utilize the conditional distribution information, $y_{ij}|x_{ij}$. The PPML estimation method uses only the first conditional moment, $\mathbb{E}\left(\xi_{ij}|x_{ij}\right) = 1$, for estimation. Note that $\mathbb{E}(\xi_{ij}|x_{ij}) = 1$ is equivalent to $\mathbb{E}(u_{ij}|x_{ij}) = 0$. It implies $\mathbb{E}\left(y_{ij}|x_{ij}\right) = \mu_{ij} = \exp\left(\beta_0^0 + \beta_1^0 x_{ij}\right)$. Then, the following moment conditions are:

$$[\beta_0]: \quad \mathbb{E}(u_{ij}) = \mathbb{E}\left(y_{ij} - \exp\left(\beta_0^0 + \beta_1^0 x_{ij}\right)\right) = 0, \text{ and} \tag{2}$$

$$[\beta_1]: \quad \mathbb{E}(x_{ij}u_{ij}) = \mathbb{E}\left(x_{ij}\left(y_{ij} - \exp\left(\beta_0^0 + \beta_1^0 x_{ij}\right)\right)\right) = 0. \tag{3}$$

We now consider the log transformation of (1) to estimate β_0^0 and β_1^0:

$$\ln(y_{ij}) = \beta_0^0 + \beta_1^0 x_{ij} + v_{ij}, \tag{4}$$

*Department of Economics, Macquarie Business School, Macquarie University. E-mail: *hanbat.jeong@mq.edu.au*
†Hubert Department of Global Health, Emory Rollins School of Public Health. E-mail: *jieun.lee@emory.edu*.

where $v_{ij} = \ln(\xi_{ij})$. By Jensen's inequality, $\mathbb{E}(\xi_{ij}|x_{ij}) = 1$ does not imply $\mathbb{E}(v_{ij}|x_{ij}) = 0$ (hence, $\ln(\mathbb{E}(y_{ij}|x_{ij})) \neq \mathbb{E}(\ln(y_{ij})|x_{ij})$). Santos Silva and Tenreyro (2006) point out that the gap $\mathbb{E}(\ln(y_{ij})) - \ln(\mathbb{E}(y_{ij})) < 0$ characterizes the bias. This gap becomes larger when (i) there are many zero values or (ii) some y_{ij}'s take significantly large positive values, leading to a large variance. To see this, consider the following examples:

1. Suppose $y_{ij} \overset{i.i.d.}{\sim}$ Bernoulli(0.5). Observe that $\ln(\mathbb{E}(y_{ij})) = \ln(1 \cdot 0.5 + 0 \cdot 0.5) \simeq -0.6931$. Now consider $\mathbb{E}\ln(y_{ij})$. As zero is not defined in the log function, we need to add some arbitrary constant, say 1, so that $\mathbb{E}(\ln(y_{ij}+1)) = 0.5 \cdot \ln(1+1) + 0.5 \cdot \ln(0+1) \simeq 0.3466$. The gap is about 1.0397.

2. Suppose $y_{ij} = \exp(\tilde{y}_{ij})$, where $\tilde{y}_{ij} \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$. Then, $\mathbb{E}(\ln(y_{ij})) - \ln(\mathbb{E}(y_{ij})) = \mu - \left(\mu + \frac{1}{2}\sigma^2\right) = -\frac{1}{2}\sigma^2$. We observe that this gap increases as σ^2 increases.

Note that the examples above imply the bias from the logarithmic transformation model is highly sensitive to the *unit* of the outcome, against the original purpose of the model (1) to estimate the constant elasticities. To see this, suppose $y^*_{ij} := 100 \cdot y_{ij}$, i.e., $y^*_{ij} \overset{i.i.d.}{\sim} 100 \cdot$ Bernoulli(0.5). Observe that $\ln\left(\mathbb{E}(y^*_{ij})\right) = \ln(100 \cdot 1 \cdot 0.5 + 100 \cdot 0 \cdot 0.5) \simeq 3.9120$. Now consider $\mathbb{E}\ln(y^*_{ij})$. For zero outcomes, we need to add some arbitrary constant (e.g., 1), where $\mathbb{E}\left(\ln(y^*_{ij}+1)\right) = \ln(100 \cdot 1 + 1) \cdot 0.5 + \ln(100 \cdot 0 + 1) \cdot 0.5 \simeq 2.3076$. The gap between $\mathbb{E}(\ln(y^*_{ij}))$ and $\ln(\mathbb{E}(y^*_{ij}))$ is about 1.6045, which is larger than that between $\mathbb{E}(\ln(y_{ij}))$ and $\ln(\mathbb{E}(y_{ij}))$ (1.0397). Conversely, suppose $y_{*,ij} := 0.01 \cdot y_{ij}$, i.e., $y_{*,ij} \overset{i.i.d.}{\sim} 0.01 \cdot$ Bernoulli(0.5). Observe that $\ln(\mathbb{E}(y_{*,ij})) = \ln(0.01 \cdot 1 \cdot 0.5 + 0.01 \cdot 0 \cdot 0.5) \simeq -5.2983$. Now consider $\mathbb{E}\ln(y_{*,ij})$, where some arbitrary constant (e.g., 1) is added for zero outcomes to be defined so that $\mathbb{E}(\ln(y_{*,ij}+1)) = \ln(0.01 \cdot 1 + 1) \cdot 0.5 + \ln(0.01 \cdot 0 + 1) \cdot 0.5 \simeq 0.0050$. The gap between $\mathbb{E}(\ln(y_{*,ij}))$ and $\ln(\mathbb{E}(y_{*,ij}))$ is then about 5.3033, which is much larger than that between $\mathbb{E}(\ln(y_{ij}))$ and $\ln(\mathbb{E}(y_{ij}))$ (1.0397).

Now we analytically investigate if the log-transformed error, $v_{ij} = \ln(\xi_{ij})$, preserve the moment conditions. Suppose that we consider two moments, $\mathbb{E}(v_{ij})$ and $\mathbb{E}(x_{ij}v_{ij})$, for estimation even though the true DGP is (1). When the two moment conditions are valid, we should have $\mathbb{E}(v_{ij}) = 0$ and $\mathbb{E}(x_{ij}v_{ij}) = 0$ under the true parameter values $\beta^0 = \left(\beta^0_0, \beta^0_1\right)'$.

Regarding (2), by the Maclaurin series expansion for $\mathbb{E}(\ln(\xi_{ij}))$, observe that

$$\mathbb{E}(v_{ij}) = \mathbb{E}(\ln(\xi_{ij})) = \sum_{p=2}^{\infty} \frac{(-1)^{p-1}}{p} \mathbb{E}\left(\left(\xi^-_{ij}\right)^p\right)$$

where $\xi^-_{ij} = \xi_{ij} - 1$ with $\mathbb{E}\left(\xi^-_{ij}|x_{ij}\right) = 0$, followed by $\mathbb{E}\left(\xi^-_{ij}\right) = 0$ by the law of iterated expectation. Hence, $\mathbb{E}(\ln(\xi_{ij}))$ could deviate from zero when the higher-order moments of ξ^-_{ij} are non-zero (i.e., $\mathbb{E}\left((\xi^-_{ij})^p\right) \neq 0$ for $p = 2, 3, \cdots$). Since ξ^-_{ij} is the error term of the level, it might exhibit large variance, heavy tails, or high skewness. As a consequence, this discrepancy may lead to large biases in the OLS estimator based on (4).

Regarding (3), observe that

$$\mathbb{E}(x_{ij}v_{ij}) = \mathbb{E}(x_{ij}\ln(\xi_{ij})) = \sum_{p=2}^{\infty} \frac{(-1)^{p-1}}{p} \mathbb{E}\left(x_{ij}\left(\xi^-_{ij}\right)^p\right),$$

where $\mathbb{E}\left(x_{ij}\xi^-_{ij}\right) = \mathbb{E}\left(x_{ij}\mathbb{E}\left(\xi^-_{ij}|x_{ij}\right)\right) = 0$ by the law of iterated expectation. $\mathbb{E}(x_{ij}v_{ij}) = 0$ holds if (i) x_{ij} and ξ^-_{ij} are independent and $\mathbb{E}\left((\xi^-_{ij})^p\right) = 0$ for $p = 2, 3, \cdots$ or (ii) all conditional moments are constant (i.e., $\mathbb{E}\left((\xi^-_{ij})^p|x_{ij}\right) = c_p$ for $p = 2, 3, \cdots$) and $\mathbb{E}(x_{ij}) = 0$ for all $i, j = 1, \cdots, n$.

Note that (i) and (ii) hold only in very restricted cases. There are numerous cases where $\mathbb{E}\left(\xi_{ij}^{-}|x_{ij}\right)=0$ holds but x_{ij} and ξ_{ij}^{-} are not independent. To see this, recall that we only assume $\mathbb{E}\left(\xi_{ij}^{-}|x_{ij}\right)=0$ without imposing assumptions on the higher moments. Thus, higher conditional moments can be supposed to take the form $\mathbb{E}\left((\xi_{ij}^{-})^p|x_{ij}\right)=h_p(x_{ij})$ for $p=2,3,\cdots$. Notably, for $p=2$, (i) and (ii) fail under heteroskedasticity. Moreover, when the conditional moment $\mathbb{E}\left((\xi_{ij}^{-})^p|x_{ij}\right)$ is not a constant function, the interaction term can be a highly nonlinear moment of x_{ij}, i.e., $\mathbb{E}\left(x_{ij}(\xi_{ij}^{-})^p\right)=\mathbb{E}\left(x_{ij}\mathbb{E}\left((\xi_{ij}^{-})^p|x_{ij}\right)\right)=\mathbb{E}\left(x_{ij}h_p(x_{ij})\right)$. Hence, we expect $\mathbb{E}\left(x_{ij}v_{ij}\right)$ to be far from zero in general.

In consequence, we can characterize the magnitudes of the asymptotic bias of the OLS estimator $\hat{\beta}^{+}=\left(\hat{\beta}_0^{+},\hat{\beta}_1^{+}\right)'$ from the log-transformed model. The asymptotic bias of $\hat{\beta}^{+}$ is characterized by the following difference:

$$\hat{\beta}^{+}-\beta^{0}=\begin{bmatrix}1 & \frac{1}{N}\sum_{i,j=1}^{n}x_{ij}\\ \frac{1}{N}\sum_{i,j=1}^{n}x_{ij} & \frac{1}{N}\sum_{i,j=1}^{n}x_{ij}^2\end{bmatrix}^{-1}\cdot\begin{pmatrix}\frac{1}{N}\sum_{i,j=1}^{n}v_{ij}\\ \frac{1}{N}\sum_{i,j=1}^{n}x_{ij}v_{ij}\end{pmatrix}.$$

Under some regularity conditions, by the law of large numbers,

1. $\begin{bmatrix}1 & \frac{1}{N}\sum_{i,j=1}^{n}x_{ij}\\ \frac{1}{N}\sum_{i,j=1}^{n}x_{ij} & \frac{1}{N}\sum_{i,j=1}^{n}x_{ij}^2\end{bmatrix}^{-1}\xrightarrow{p}\frac{1}{\mu_{x,2}-\mu_{x,1}^2}\begin{bmatrix}\mu_{x,2} & -\mu_{x,1}\\ -\mu_{x,1} & 1\end{bmatrix}$,
2. $\frac{1}{N}\sum_{i,j=1}^{n}v_{ij}\xrightarrow{p}\lim_{n\to\infty}\frac{1}{N}\sum_{i,j=1}^{n}\mathbb{E}(v_{ij})=\lim_{n\to\infty}\frac{1}{N}\sum_{i,j=1}^{n}\sum_{p=2}^{\infty}\frac{(-1)^{p-1}}{p}\mathbb{E}\left(h_p(x_{ij})\right)$,
3. $\frac{1}{N}\sum_{i,j=1}^{n}x_{ij}v_{ij}\xrightarrow{p}\lim_{n\to\infty}\frac{1}{N}\sum_{i,j=1}^{n}\mathbb{E}(x_{ij}v_{ij})=\lim_{n\to\infty}\frac{1}{N}\sum_{i,j=1}^{n}\sum_{p=2}^{\infty}\frac{(-1)^{p-1}}{p}\mathbb{E}\left(x_{ij}h_p(x_{ij})\right)$,

as $n\to\infty$, where $\mu_{x,1}=\lim_{n\to\infty}\frac{1}{N}\sum_{i,j=1}^{n}\mathbb{E}(x_{ij})$ and $\mu_{x,2}=\lim_{n\to\infty}\frac{1}{N}\sum_{i,j=1}^{n}\mathbb{E}(x_{ij}^2)$. Consequently, we have

$$\begin{pmatrix}\hat{\beta}_0^{+}-\beta_0^{0}\\ \hat{\beta}_1^{+}-\beta_1^{0}\end{pmatrix}\xrightarrow{p}\frac{1}{\mu_{x,2}-\mu_{x,1}^2}\begin{pmatrix}\mu_{x,2}\lim_{n\to\infty}\frac{1}{N}\sum_{i,j=1}^{n}\sum_{p=2}^{\infty}\frac{(-1)^{p-1}}{p}\mathbb{E}\left(h_p(x_{ij})\right)-\mu_{x,1}\lim_{n\to\infty}\frac{1}{N}\sum_{i,j=1}^{n}\sum_{p=2}^{\infty}\frac{(-1)^{p-1}}{p}\mathbb{E}\left(x_{ij}h_p(x_{ij})\right)\\ -\mu_{x,1}\lim_{n\to\infty}\frac{1}{N}\sum_{i,j=1}^{n}\sum_{p=2}^{\infty}\frac{(-1)^{p-1}}{p}\mathbb{E}\left(h_p(x_{ij})\right)+\lim_{n\to\infty}\frac{1}{N}\sum_{i,j=1}^{n}\sum_{p=2}^{\infty}\frac{(-1)^{p-1}}{p}\mathbb{E}\left(x_{ij}h_p(x_{ij})\right)\end{pmatrix}$$

as $n\to\infty$.

Our model specification accounts for network spillovers in OD flows. Based on the distribution of $y_{ij}|\mathbf{x}$, we consider a stronger-type moment condition:

$$\mathbb{E}\left(\xi_{ij}^{-}|\mathbf{x}\right)=0,$$

where $\mathbf{x}=(x_{11},\cdots,x_{n1},\cdots,x_{1n},\cdots,x_{nn})'$. That is, the conditional expectation of ξ_{ij}^{-} is zero when all connected characteristics are known. Since our method is based on the distribution of $y_{ij}|\mathbf{x}$, we allow a more general structure on the higher-order conditional moments: $\mathbb{E}\left(\left(\xi_{ij}^{-}\right)^p|\mathbf{x}\right)=h_p(\mathbf{x})$ for $p=2,3,\cdots$. For example, suppose $\mathbb{E}\left(\left(\xi_{ij}^{-}\right)^2|\mathbf{x}\right)=c_0+c_1x_{ij}^2+c_2x_{kj}^2+c_3x_{il}^2$, where $c_0,c_1,c_2,c_3>0$, k is an i's neighbor, and l is a j's neighbor. In this case,

$$\mathbb{E}\left(x_{ij}(\xi_{ij}^{-})^2\right)=\mathbb{E}\left(x_{ij}\mathbb{E}\left((\xi_{ij}^{-})^2|x_{ij},x_{kj},x_{il}\right)\right)=c_0\mathbb{E}\left(x_{ij}\right)+c_1\mathbb{E}\left(x_{ij}^3\right)+c_2\mathbb{E}\left(x_{ij}x_{kj}^2\right)+c_3\mathbb{E}\left(x_{ij}x_{il}^2\right).$$

Comparing this expression with the special case with $c_2=c_3=0$ (no spillovers) highlights how $\mathbb{E}(x_{ij}v_{ij})$ can deviate further from zero. This deviation arises from the inclusion of the nonzero terms $\mathbb{E}(x_{ij}x_{kj}^2)$ and $\mathbb{E}\left(x_{ij}x_{il}^2\right)$, which are absent in the non-spillover scenario.

1.2 Adding some constant $c > 0$ to y_{ij} in the log-transformation

We consider the effect of adding some constant $c > 0$ in the logarithmic transformation. First of all, we review the results studied by Mullahy and Norton (2024). Consider the quantity $\frac{d\ln(y_{ij}+c)}{dy_{ij}} = \frac{1}{y_{ij}+c}$ for $c > 0$ around $y_{ij} = 0$, that is, $\frac{d\ln(y_{ij}+c)}{dy_{ij}}\Big|_{y_{ij}=0} = \frac{1}{c}$. This quantity means the marginal change of the log-transformed outcome $\ln(y_{ij} + c)$ when $y_{ij} = 0$. Then,

$$\frac{d\ln(y_{ij}+c)}{dy_{ij}}\Bigg|_{y_{ij}=0} = \frac{1}{c}\begin{cases} \to 0 \text{ as } c \to \infty \\ \to \infty \text{ as } c \to 0 \end{cases}.$$

A small change around $y_{ij} = 0$ produces significantly different $\ln(y_{ij} + c)$ values depending on c. When c is close to zero, the changed quantity from $\ln(0+c)$ to $\ln(y_{ij}+c)$ becomes extremely large for any $y_{ij} > 0$. On the other hand, if c is sufficiently large, the difference between $\ln(0+c)$ and $\ln(y_{ij}+c)$ is close to zero. Hence, considering $c \to 0$ highlights the distinct structures of $y_{ij} = 0$ and $y_{ij} > 0$, while considering $c \to \infty$ is similar to the non-transformed model. Note that, however, adding $c \to \infty$ involves an asymptotic bias that grows to infinity for y_{ij} close to zero, as shown in (5).

We go beyond the existing works to study the impact of adding "$c > 0$" on the OLS estimator's bias. Let $\hat{\beta}^+(c)$ be the OLS estimator when we employ $\ln(y_{ij} + c)$ as the dependent variable in (4). The asymptotic bias of $\hat{\beta}^+(c)$ can be characterized by the following difference:

$$\begin{aligned}\hat{\beta}^+(c) - \beta^0 = &\begin{bmatrix} 1 & \frac{1}{N}\sum_{i,j=1}^n x_{ij} \\ \frac{1}{N}\sum_{i,j=1}^n x_{ij} & \frac{1}{N}\sum_{i,j=1}^n x_{ij}^2 \end{bmatrix}^{-1} \cdot \begin{pmatrix} \frac{1}{N}\sum_{i,j=1}^n v_{ij} \\ \frac{1}{N}\sum_{i,j=1}^n x_{ij}v_{ij} \end{pmatrix} \\ &+ \begin{bmatrix} 1 & \frac{1}{N}\sum_{i,j=1}^n x_{ij} \\ \frac{1}{N}\sum_{i,j=1}^n x_{ij} & \frac{1}{N}\sum_{i,j=1}^n x_{ij}^2 \end{bmatrix}^{-1} \cdot \begin{pmatrix} \frac{1}{N}\sum_{i,j=1}^n \Delta_{y,ij}(c) \\ \frac{1}{N}\sum_{i,j=1}^n x_{ij}\Delta_{y,ij}(c) \end{pmatrix},\end{aligned} \tag{5}$$

where $\Delta_{y,ij}(c) := \begin{cases} \ln\left(1 + \frac{c}{y_{ij}}\right) = \ln(y_{ij} + c) - \ln(y_{ij}) & \text{if } y_{ij} > 0 \\ \ln\left(1 + \frac{c}{\varepsilon_y}\right) = \ln(\varepsilon_y + c) - \ln(\varepsilon_y) & \text{if } y_{ij} = 0, \text{ where } \varepsilon_y > 0 \text{ denotes an infinitesimal number.} \end{cases}$

Observe that the first part of $\hat{\beta}^+(c) - \beta^0$ is the same as $\hat{\beta}^+ - \beta^0$. Hence, the second part of $\hat{\beta}^+(c) - \beta^0$ describes the source of the asymptotic bias arising from $c > 0$. Then, the second bias part is characterized by $\lim_{n\to\infty} \frac{1}{N}\sum_{i,j=1}^n \mathbb{E}(\Delta_{y,ij}(c))$ and $\lim_{n\to\infty} \frac{1}{N}\sum_{i,j=1}^n \mathbb{E}(x_{ij}\Delta_{y,ij}(c))$. Consider the quantity $\lim_{n\to\infty} \frac{1}{N}\sum_{i,j=1}^n \mathbb{E}(\Delta_{y,ij}(c))$ for a simple explanation. Then,

$$\frac{1}{N}\sum_{i,j=1}^n \mathbb{E}(\Delta_{y,ij}(c)) = \frac{1}{N}\sum_{i,j=1}^n \mathbf{1}\{0 \le y_{ij} < \varepsilon_y\} \cdot \mathbb{E}(\Delta_{y,ij}(c)) + \frac{1}{N}\sum_{i,j=1}^n \mathbf{1}\{y_{ij} \ge \varepsilon_y\} \cdot \mathbb{E}(\Delta_{y,ij}(c)),$$

Note that for $y_{ij} \in [0, \varepsilon_y]$, $\mathbb{E}(\Delta_{y,ij}(c))$ can be extremely large as $c \to 0$, although $\mathbb{E}(\Delta_{y,ij}(c))$ may take a moderately bounded value under some regularity conditions. Since

$$\frac{1}{N}\sum_{i,j=1}^n \mathbf{1}\{0 \le y_{ij} < \varepsilon_y\} \cdot \mathbb{E}(\Delta_{y,ij}(c)) \ge \underbrace{\frac{\sum_{i,j=1}^n \mathbf{1}\{0 \le y_{ij} < \varepsilon_y\}}{N}}_{\substack{\text{proportion of } y_{ij}\text{'s} \\ \text{zero or close to zero}}} \cdot \inf_{\substack{n,i,j, \\ 0 \le y_{ij} < \varepsilon_y}} \mathbb{E}(\Delta_{y,ij}(c)),$$

we expect a large bias of $\hat{\beta}^+(c)$ when a sample includes many zero values or positive infinitesimal values.

1.3 Interpretations of our model

This subsection rigorously examines the key properties of our model. In our application, we focus on the international trade flow and extend the previous discussion summarized by Head and Mayer (2014).

Let

- y_{ij} = trade flow from j to i,
- $\mu_{ij} = \mathbb{E}(y_{ij}|\mathbf{z})$, where $\mathbf{z}$ denotes a vector of exogenous characteristics,
- G_i^I = importer i's total expenditure,
- G_j^E = exporter j's total production,
- G_i = country i's GDP,
- π_{ij} = a measure of bilateral frictions (costs),
- D_{ij} = geographic distance between i and j.

A simple multiplicative gravity model (Tinbergen (1962)) is specified by

$$\mu_{ij} = \mu \cdot G_i^I \cdot G_j^E \cdot \pi_{ij}, \tag{6}$$

where μ is a constant. When the triple identity (of GDP) holds (e.g., $G_i^I = G_i$ and $G_j^E = G_j$), equation (6) is simplified by $\mu_{ij} = \mu \cdot G_i \cdot G_j \cdot \pi_{ij}$. If π_{ij} is a function of the inverse distance, this conventional equation reflects two stylized facts about gravity well: (i) trade is proportional to capacity, and (ii) trade is inversely proportional to distance (see Figure 3.1 in Head and Mayer (2014)).

Conventional specifications (e.g., equation (6)) only consider the bilateral trade cost between two countries. For example, McCallum (1995) considers the following specification on π_{ij}:

$$\ln \pi_{ij} = \beta_w \ln D_{ij} + \beta_b B_{ij},$$

where $B_{ij} = \mathbf{1}\{\text{Regions } i \text{ and } j \text{ are in Canada}\}$. By estimating positively significant β_b, McCallum (1995) finds that trade between two provinces in Canada is over 22 times larger than trade between a Canadian province and a U.S. state. This result implies that the Canada-U.S. border is a significant barrier to trade (McCallum border puzzle).

1.3.1 Demand-side-based Gravity Equation (Anderson and van Wincoop 2003)

Anderson and van Wincoop (2003) establish the structural gravity equation by including the concept of multilateral resistance, based on the demand side. Our model extends their framework using the spatial autoregressive model's structure. To address the McCallum border puzzle, the structural gravity equation specification is:

$$\mu_{ij} = \frac{G_i \cdot G_j}{G^W} \cdot \left(\frac{\pi_{ij}}{\Pi_j \cdot P_i}\right)^{1-\varrho}, \tag{7}$$

where $G^W \equiv \sum_{k=1}^n G_k$ represents the world GDP, Π_j denotes the outward resistance, P_i is the inward resistance, and $\varrho > 1$ stands for the elasticity of substitution between all goods.

First, the outward resistance Π_j shows how exporter j faces trade barriers across all potential export destinations: the overall difficulty of sending goods from j to other countries around the world. This Π_j can be interpreted as a price index. In the (partial) equilibrium, given $(P_1, \cdots, P_n)$,

$$\Pi_j = \left(\sum_{k=1}^{n} \frac{G_k}{G^W} \left(\frac{\pi_{kj}}{P_k} \right)^{1-\varrho} \right)^{\frac{1}{1-\varrho}}. \tag{8}$$

Hence, the outward resistance Π_j represents the overall trade cost from j since each $\frac{\pi_{kj}}{P_k}$ illustrates the normalized trade cost from j to k and Π_j consists of aggregated $\frac{\pi_{kj}}{P_k}$ for $k = 1, \cdots, n$ weighted by the GDP shares $\frac{G_k}{G^W}$ for $k = 1, \cdots, n$. For example, suppose that π_{kj} (= trade cost from j to k) for some k decreases. A drop in π_{kj} means that country j has a more attractive (less costly) export route to k. From the country j's perspective, this lowers the overall export barrier it faces in the world since one key route becomes cheaper.

Second, the inward resistance P_i captures how importer i experiences trade barriers across all possible foreign suppliers (= a measure of the overall difficulty of importing from the rest of the world into i). In the (partial) equilibrium, given $(\Pi_1, \cdots, \Pi_n)$,

$$P_i = \left(\sum_{k=1}^{n} \frac{G_k}{G^W} \left(\frac{\pi_{ik}}{\Pi_k} \right)^{1-\varrho} \right)^{\frac{1}{1-\varrho}}. \tag{9}$$

Like the outward resistance, the inward resistance P_i captures the overall trade cost to i. Like the outward resistance, if π_{ik} decreases for some k, it leads to cheaper access to one key supplier k. This then lowers the overall "import barrier" faced by importer country i. In consequence, decreasing π_{ik} causes lower P_i.

As the third component, the elasticity of substitution among goods $\varrho > 1$ generates the main motivation of trade (Dixit and Stiglitz (1993)). That is, goods (from monopolistic competition) are imperfect substitutes, and consumers prefer to have variety. If ϱ is close to 1, consumers have strong preferences for specific varieties (less substitutability). On the other hand, $\varrho = \infty$ indicates perfect substitutability. When π_{ij} increases under large ϱ, μ_{ij} in equation (7) significantly decreases. When $\varrho \to \infty$, $P_i \to \min_{k=1,\cdots,n}\{\frac{\pi_{ik}}{\Pi_k}\}$ and $\Pi_j \to \min_{k=1,\cdots,n}\{\frac{\pi_{kj}}{P_k}\}$. In the case of perfect substitutability, trade flows are dominated by the route with the lowest resistance (i.e., the smallest $\frac{\pi_{ik}}{\Pi_k}$ or $\frac{\pi_{kj}}{P_k}$). On the other hand, π_{ij} does not play a role in μ_{ij} if $\varrho \to 1$. As $\varrho \to 1$, $P_i = \sum_{k=1}^{n} \frac{\pi_{ik}}{\Pi_k} \cdot \frac{G_k}{G^W}$ and $\Pi_j = \sum_{k=1}^{n} \frac{\pi_{kj}}{P_k} \cdot \frac{G_k}{G^W}$. In the case of perfect complementarity, all trade links are treated in an additive way (i.e., the full average of all links).

From an econometric perspective, the McCallum border puzzle arises due to omitted variable bias. When equation (7) is the true model, conventional gravity specification (e.g., equation (6)) omits the multilateral resistance terms. Since the multilateral resistance terms (8) and (9) contain $\{G_k, \pi_{ik}, \pi_{kj}\}_{k=1}^{n}$, the omitted terms in the traditional gravity equation are dependent on the original components G_i, G_j, and π_{ij}.

1.3.2 Detailed solutions to our model

Note that our model's theoretical foundation is a modification of Anderson and van Wincoop (2003). Here we introduce the details of the model's solution.

Step 1 (solving Stage 2). We will apply the backsolving procedure. Suppose that the trade cost factors $\{\pi_{ij}\}$ were determined in **Stage 1**.

Step 1.1: Demand function. First, we will derive a demand function of country i (importer). Let c_{ij} be consumption by country i consumers of goods from country j. A representative consumer in country i chooses $\{c_{i1}, \cdots, c_{in}\}$ by maximizing the following problem:

$$\max_{\{c_{ij}\}_{j=1}^n} U_i = \left(\sum_{j=1}^n \chi_j^{\frac{1}{\varrho}} \cdot c_{ij}^{\frac{\varrho-1}{\varrho}} \right)^{\frac{\varrho}{\varrho-1}} \text{ subject to } \sum_{j=1}^n p_{ij} c_{ij} = G_i, \tag{10}$$

where χ_j denotes a preference parameter for country j's good and p_{ij} is the price of country i of consuming one unit from country j. Importantly, note that $G_1, \cdots, G_n$ are exogenously given. We will discuss p_{ij} in **Step 1.2**.

To solve (10), we set up the Lagrangian:

$$\mathcal{L} = \left(\sum_{j=1}^n \chi_j^{\frac{1}{\varrho}} \cdot c_{ij}^{\frac{\varrho-1}{\varrho}} \right)^{\frac{\varrho}{\varrho-1}} - \lambda_i \left(\sum_{j=1}^n p_{ij} c_{ij} - G_i \right),$$

where λ_i denotes the Lagrange multiplier. For notational convenience, define $C_i = \sum_{j=1}^n \chi_j^{\frac{1}{\varrho}} \cdot c_{ij}^{\frac{\varrho-1}{\varrho}}$ for $i = 1, \cdots, n$. Then, the first-order condition generates:

$$\begin{aligned} 0 &= \frac{\partial \mathcal{L}}{\partial c_{ij}} = \frac{\partial U_i}{\partial c_{ij}} - \lambda_i p_{ij} \\ &\Leftrightarrow C_i^{\frac{1}{\varrho-1}} \chi_j^{\frac{1}{\varrho}} c_{ij}^{-\frac{1}{\varrho}} = \lambda_i p_{ij} \Leftrightarrow c_{ij}^{\frac{1}{\varrho}} = \frac{C_i^{\frac{1}{\varrho-1}} \chi_j^{\frac{1}{\varrho}}}{\lambda_i p_{ij}} \end{aligned}$$

since $\frac{\partial U_i}{\partial c_{ij}} = \frac{\varrho}{\varrho-1} C_{ij}^{\frac{1}{\varrho-1}} \frac{\partial C_{ij}}{\partial c_{ij}} = C_{ij}^{\frac{1}{\varrho-1}} \chi_j^{\frac{1}{\varrho}} c_{ij}^{-\frac{1}{\varrho}}$. This implies

$$c_{ij}^* = \frac{C_i^{\frac{\varrho}{\varrho-1}} \chi_j}{(\lambda_i p_{ij})^{\varrho}}. \tag{11}$$

Next, we will derive the CES price index P_i by the cost minimization problem:

$$\min_{\{c_{ij}\}_{j=1}^n} \sum_{j=1}^n p_{ij} c_{ij} \text{ subject to } \left(\sum_{j=1}^n \chi_j^{\frac{1}{\varrho}} \cdot c_{ij}^{\frac{\varrho-1}{\varrho}} \right)^{\frac{\varrho}{\varrho-1}} = \overline{U}_i \tag{12}$$

for some $\overline{U}_i$. We set up the Lagrangian to solve (12):

$$\mathcal{L}^{**} = \sum_{j=1}^n p_{ij} c_{ij} + \lambda_i^{**} \left(\overline{U}_i - C_i^{\frac{\varrho}{\varrho-1}} \right).$$

The first-order condition is

$$\begin{aligned}
0 = \frac{\partial \mathcal{L}^{**}}{\partial c_{ij}} &= p_{ij} - \lambda_i^{**} \frac{\varrho}{\varrho - 1} C_i^{\frac{1}{\varrho-1}} \cdot \frac{\partial C_i}{\partial c_{ij}} \\
\Leftrightarrow p_{ij} &= \lambda_i^{**} C_i^{\frac{1}{\varrho-1}} \chi_j^{\frac{1}{\varrho}} c_{ij}^{-\frac{1}{\varrho}} \\
\Leftrightarrow c_{ij}^{**} &= \left(\lambda_i^{**} C_i^{\frac{1}{\varrho-1}} \chi_j^{\frac{1}{\varrho}} p_{ij}^{-1} \right)^{\varrho}.
\end{aligned}$$

The utility constraint in (12) is equivalent that $\overline{U}_i = C_i^{\frac{\varrho}{\varrho-1}}$. Hence,

$$\begin{aligned}
C_i &= \sum_{j=1}^{n} \chi_j^{\frac{1}{\varrho}} (c_{ij}^{**})^{\frac{\varrho-1}{\varrho}} \\
&= \sum_{j=1}^{n} \chi_j^{\frac{1}{\varrho}} \left(\lambda_i^{**} C_i^{\frac{1}{\varrho-1}} \chi_j^{\frac{1}{\varrho}} p_{ij}^{-1} \right)^{\varrho-1} \\
&= (\lambda_i^{**})^{\varrho-1} C_i \sum_{j=1}^{n} \chi_j p_{ij}^{1-\varrho}.
\end{aligned}$$

Hence,

$$\lambda_i^{**} = \left(\sum_{j=1}^{n} \chi_j p_{ij}^{1-\varrho} \right)^{\frac{1}{1-\varrho}}. \tag{13}$$

Then, the minimum expenditure of country i's consumer is

$$\begin{aligned}
E_i^{**} &= \sum_{j=1}^{n} p_{ij} \left(\lambda_i^{**} C_i^{\frac{1}{\varrho-1}} \chi_j^{\frac{1}{\varrho}} p_{ij}^{-1} \right)^{\varrho} \\
&= (\lambda_i^{**})^{\varrho} C_i^{\frac{\varrho}{\varrho-1}} \sum_{j=1}^{n} \chi_j p_{ij}^{1-\varrho} \\
&= \lambda_i^{**} \cdot \overline{U}_i
\end{aligned} \tag{14}$$

by the constraint $\overline{U}_i = C_i^{\frac{\varrho}{\varrho-1}}$ and (13). This gives $\lambda_i^{**} = \frac{\partial E_i^{**}}{\partial \overline{U}_i}$.

In the consumer's minimization problem, λ_i^{**} means the marginal cost of utility (shadow price for one unit of utility): the marginal expenditure to gain one unit of utility. Thus, $E_i^{**} = P_i \cdot \overline{U}_i = \lambda_i^{**} \cdot \overline{U}_i$, so that $P_i = \lambda_i^{**}$, where

$$P_i = \left(\sum_{j=1}^{n} \chi_j p_{ij}^{1-\varrho} \right)^{\frac{1}{1-\varrho}}$$

is the summary of prices for country i.

Now we return to the consumer's maximization problem. When we apply (11) to the budget constraint,

$$G_i = \sum_{j=1}^{n} p_{ij} \left(\underbrace{C_i^{\frac{\varrho}{\varrho-1}} \chi_j \lambda_i^{-\varrho} p_{ij}^{-\varrho}}_{c_{ij}^*} \right) = C_i^{\frac{\varrho}{\varrho-1}} \lambda_i^{-\varrho} P_i^{1-\varrho}$$

by the definition of P_i. Hence,

$$\lambda_i = C_i^{\frac{1}{\varrho-1}} P_i^{\frac{1-\varrho}{\varrho}} G_i^{-\frac{1}{\varrho}}. \tag{15}$$

Then, $\lambda_i = \frac{\partial U_i}{\partial G_i} = C_i^{\frac{1}{\varrho-1}} P_i^{\frac{1-\varrho}{\varrho}} G_i^{-\frac{1}{\varrho}}$ presents the increased utility when G_i increases by one unit. In consequence, (11) generates the demand function:

$$c_{ij}^* = C_i^{\frac{\rho}{\varrho-1}} \chi_j p_{ij}^{-\varrho} C_i^{-\frac{\rho}{\varrho-1}} P_i^{\varrho-1} G_i = \chi_j \left(\frac{p_{ij}}{P_i} \right)^{-\varrho} \frac{G_i}{P_i}. \tag{16}$$

Step 1.2: Market clearing. The existence of trade costs leads to heterogeneous prices. We assume

$$p_{ij} = p_j \cdot \pi_{ij},$$

where p_j is the exporter's supply price.

Firstly, we assume that each p_j $(j = 1, \cdots, n)$ is exogenously given. When each country's market is assumed to be perfectly competitive, the exporter's supply price p_j is determined by the marginal cost in country j, i.e., $p_j = \frac{w_j}{A_j}$ where w_j denotes a wage and A_j represents the productivity of a worker.[1] Alternatively, if we consider monopolistic competition, each exporter j produces its differentiated variety at the marginal cost $\frac{w_j}{A_j}$. In this case, $p_j = \frac{\varrho}{\varrho-1} \cdot \frac{w_j}{A_j}$ implying a constant markup $\frac{\varrho}{\varrho-1}$ above the marginal cost.

The nominal value of exports from country j to country i (= country i's payment to j) is

$$\mu_{ij} = p_{ij} c_{ij} = \underbrace{p_j c_{ij}}_{\text{Value of production at the origin } j} + \underbrace{(\pi_{ij} - 1) p_j c_{ij}}_{\text{Trade cost that exporter passes on to the importer}}.$$

When $\pi_{ij} = 1$, $p_{ij} = p_j$ which implies that no additional cost occurs. If $\pi_{ij} > 1$, the extra cost $\pi_{ij} - 1$ for a unit good in exports from j to i arises.

Hence, we have

$$\mu_{ij}^* = p_{ij} c_{ij}^* = \chi_j p_{ij}^{1-\varrho} P_i^{-(1-\varrho)} G_i = \chi_j (p_j \pi_{ij})^{1-\varrho} P_i^{-(1-\varrho)} G_i. \tag{17}$$

The market-clearing condition imposes

$$G_j = \sum_{i=1}^{n} \mu_{ij}^* = \chi_j p_j^{1-\varrho} \sum_{i=1}^{n} \pi_{ij}^{1-\varrho} P_i^{-(1-\varrho)} G_i. \tag{18}$$

[1] Note that A_j is the amount of a good each worker can produce.

By imposing $p_1 = p_2 = \cdots = p_n = 1$ (price normalization)[2], we then obtain

$$\chi_j = \frac{G_j}{\sum_{i=1}^n \left(\frac{\pi_{ij}}{P_i}\right)^{1-\varrho} G_i} = \frac{G_j}{G^W} \frac{1}{\sum_{i=1}^n \left(\frac{\pi_{ij}}{P_i}\right)^{1-\varrho} \frac{G_i}{G^W}} = \frac{G_j}{G^W} \Pi_j^{-(1-\varrho)}$$

by the definition in (8). Hence, equation (17) becomes

$$\mu_{ij} = \frac{G_i G_j}{G^W} \left(\frac{\pi_{ij}}{\Pi_j P_i}\right)^{1-\varrho}. \tag{19}$$

Further, we can verify that

$$P_i^{1-\varrho} = \sum_{j=1}^n \chi_j p_{ij}^{1-\varrho} = \sum_{j=1}^n \left(\frac{\pi_{ij}}{\Pi_j}\right)^{1-\varrho} \cdot \frac{G_j}{G^W},$$

which is the same as (9).

Step 2 (solving Stage 1). The next step is to characterize the equilibrium negotiated trade cost factor π_{ij}. Our specification on π_{ij} is following:

$$\pi_{ij} = \left(\mu_{ij}^{\text{proxy}}\right)^{-1} \cdot \underbrace{D_{ij,1}^{\tilde{\beta}_1} \cdots D_{ij,K}^{\tilde{\beta}_K} \cdot E_{i,1}^{\tilde{\gamma}_{1,d}} \cdots E_{i,L}^{\tilde{\gamma}_{L,d}} \cdot E_{j,1}^{\tilde{\gamma}_{1,o}} \cdots E_{j,L}^{\tilde{\gamma}_{L,o}}}_{\equiv \pi_{ij}^{+}}. \tag{20}$$

π_{ij} consists of two parts: (i) endogenous factor from routing and negotiation $\left(\mu_{ij}^{\text{proxy}}\right)^{-1}$ and (ii) usual cost specification part (π_{ij}^{+}) specifying information costs, design costs, legal and regulatory costs, and transport costs. In detail,

- $D_{ij,k}$ $(k = 1, \cdots, K)$ presents a bilateral characteristic with structural parameters $\tilde{\beta}_1, \cdots, \tilde{\beta}_K$; and
- $E_{i,l}$ $(E_{j,l})$ is a destination-specific (origin-specific) characteristic for $l = 1, \cdots, L$. The relevant structural parameters are $\tilde{\gamma}_{1,d}, \cdots, \tilde{\gamma}_{L,d}$ $(\tilde{\gamma}_{1,o}, \cdots, \tilde{\gamma}_{L,o})$.

μ_{ij}^{proxy}, a new term in our model, captures a discounting factor for the trade barrier for μ_{ij}. Specifically, we assume

$$\mu_{ij}^{\text{proxy}} = \left(\prod_{k=1}^n \mu_{kj}^{w_{ik}}\right)^{\tilde{\lambda}_d} \left(\prod_{l=1}^n \mu_{il}^{w_{jl}}\right)^{\tilde{\lambda}_o} \left(\prod_{k,l=1}^n \mu_{kl}^{w_{ik} w_{jl}}\right)^{\tilde{\lambda}_w}, \tag{21}$$

where w_{ij} is a network link between i and j satisfying $\sum_{j=1}^n w_{ij} = 1$ and $w_{ii} = 0$ for all $i = 1, \cdots, n$, and $\tilde{\lambda}_d$, $\tilde{\lambda}_o$ and $\tilde{\lambda}_w$ are the main structural parameters. Hence, μ_{ij}^{proxy} is the three-type geometric averages of other flows:

(i) $\overline{\mu}_{\cdot j}^{i} = \prod_{k=1}^n \mu_{kj}^{w_{ik}}$ is the average of outflows from country j,

(ii) $\overline{\mu}_{i\cdot}^{j} = \prod_{l=1}^n \mu_{il}^{w_{jl}}$ denotes the average of inflows to country i, and

(iii) $\overline{\mu}_{\cdot\cdot}^{ij} = \prod_{k,l=1}^n \mu_{kl}^{w_{ik} w_{jl}}$ represents the average of flows among third-party units. Note that $\overline{\mu}_{\cdot\cdot}^{ij}$ contains μ_{ji} as a component (i.e., $\mu_{ji}^{w_{ij} w_{ji}}$).

[2]This normalization does not affect the gravity equation form.

This specification originates from LeSage and Pace (2008): from an $n \times n$ network matrix W with $w_{ii} = 0$ for $i = 1, \cdots, n$, we clearly separate the three-type flows. Moreover, these classifications are mutually exclusive and collectively exhaustive. When $\tilde{\lambda}_d > 0$, $\tilde{\lambda}_o > 0$, and $\tilde{\lambda}_w > 0$, we have $\mu_{ij}^{\text{proxy}} > 1$. In this case, the trade cost π_{ij} is reduced ($\pi_{ij} \leq \pi_{ij}^{+}$) by utilizing information about the trade cost. On the other hand, if $\tilde{\lambda}_d \simeq \tilde{\lambda}_o \simeq \tilde{\lambda}_w \simeq 0$, $\pi_{ij} \simeq \pi_{ij}^{+}$ since $\mu_{ij}^{\text{proxy}} \simeq 1$. We will provide the detailed interpretations of those geometric averages later.

Define $\lambda_d = (\varrho - 1)\tilde{\lambda}_d$, $\lambda_o = (\varrho - 1)\tilde{\lambda}_o$, $\lambda_w = (\varrho - 1)\tilde{\lambda}_w$, $\beta_k = (1 - \varrho)\tilde{\beta}_k$ for $k = 1, \cdots, K$, $\gamma_{l,d} = (1 - \varrho)\tilde{\gamma}_{l,d}$ and $\gamma_{l,o} = (1 - \varrho)\tilde{\gamma}_{l,o}$ for $l = 1, \cdots, L$. Let

$$\mu_{ij}^{+} = D_{ij,1}^{\beta_1} \cdots D_{ij,K}^{\beta_K} \cdot E_{i,1}^{\gamma_{1,d}} \cdots E_{i,L}^{\gamma_{L,d}} \cdot E_{j,1}^{\gamma_{1,o}} \cdots E_{j,L}^{\gamma_{L,o}}$$

denote the pure exogenous part of μ_{ij}. Note that equation (19) can be alternatively represented by

$$\begin{aligned}
\mu_{ij} &= \frac{G_i G_j}{G^W} \left(\frac{\pi_{ij}}{P_i \Pi_j} \right)^{1-\varrho} \\
&= \frac{G_i G_j}{G^W} \cdot P_i^{\varrho-1} \Pi_j^{\varrho-1} \cdot (\mu_{ij}^{\text{proxy}})^{\varrho-1} \cdot (\pi_{ij}^{+})^{1-\varrho} \\
&= \underbrace{(\overline{\mu}_{\cdot j}^{i})^{\lambda_d} (\overline{\mu}_{i \cdot}^{j})^{\lambda_o} (\overline{\mu}_{\cdot\cdot}^{ij})^{\lambda_w}}_{\text{Part A}} \cdot \underbrace{P_i^{\varrho-1} \Pi_j^{\varrho-1}}_{\text{Part B}} \cdot \underbrace{G_i G_j \cdot (G^W)^{-1} \cdot \mu_{ij}^{+}}_{\text{Part C}}.
\end{aligned} \tag{22}$$

Step 2.1: Unique form of the optimal trade flow μ_{ij}^{*}. Our next goal is to obtain the uniqueness of the optimal trade flow μ_{ij}^{*} satisfying equation (22), i.e., the unique representation of μ_{ij}^{*} as a function of the components in μ_{kl}^{+} for $k, l = 1, \cdots, n$. In this step, we will derive a sufficient condition guaranteeing the uniqueness of μ_{ij}^{*}.

From equation (22), μ_{ij}^{*} consists of three parts: (i) explicitly endogenous term (Part A), (ii) implicitly endogenous term (Part B), and (iii) purely exogenous term (Part C). To be consistent with the econometric framework, let

$$\mu_{ij}^{+} = \mu_{ij}^{B} \cdot \mu_i^{D} \cdot \mu_j^{O},$$

where $\mu_{ij}^{B} = D_{ij,1}^{\beta_1} \cdots D_{ij,K}^{\beta_K}$ denotes the bilaternal exogenous characteristic components, $\mu_i^{D} = E_{i,1}^{\gamma_{1,d}} \cdots E_{i,L}^{\gamma_{L,d}}$ is the combination of destination-specific exogenous characteristics, and $\mu_j^{O} = E_{j,1}^{\gamma_{1,o}} \cdots E_{j,L}^{\gamma_{L,o}}$ stands for the part of origin-specific exogenous characteristics. Further, we denote

$$\begin{aligned}
\Pi_j(\boldsymbol{\mu}) &= \left(\sum_{i=1}^{n} \left(\frac{\pi_{ij}(\boldsymbol{\mu})}{P_i(\boldsymbol{\mu})} \right)^{1-\varrho} \frac{G_i}{G^W} \right)^{\frac{1}{1-\varrho}}, \text{ for } j = 1, \cdots, n \text{ and} \\
P_i(\boldsymbol{\mu}) &= \left(\sum_{j=1}^{n} \left(\frac{\pi_{ij}(\boldsymbol{\mu})}{\Pi_j(\boldsymbol{\mu})} \right)^{1-\varrho} \frac{G_j}{G^W} \right)^{\frac{1}{1-\varrho}}, \text{ for } i = 1, \cdots, n,
\end{aligned}$$

for each $\boldsymbol{\mu}$, where $\boldsymbol{\mu} = (\mu_{11}, \cdots, \mu_{n1}, \cdots, \mu_{1n}, \cdots, \mu_{nn})'$. Note that these notations highlight that the components above rely on $\boldsymbol{\mu}$. In our econometric framework, note that the fixed-effect components have their own structures:

$$\begin{aligned}
\tilde{\alpha}_j(\boldsymbol{\mu}) &= (G^W)^{-\frac{1}{2}} \cdot G_j \cdot \Pi_j^{\varrho-1}(\boldsymbol{\mu}) \cdot \mu_j^{O} \text{ for } j = 1, \cdots, n, \text{ and} \\
\tilde{\eta}_i(\boldsymbol{\mu}) &= (G^W)^{-\frac{1}{2}} \cdot G_i \cdot P_i^{\varrho-1}(\boldsymbol{\mu}) \cdot \mu_i^{D} \text{ for } i = 1, \cdots, n
\end{aligned}$$

to have $\alpha_j(\boldsymbol{\mu}) = \ln(\tilde{\alpha}_j(\boldsymbol{\mu}))$ for $j = 1, \cdots, n$ and $\eta_i(\boldsymbol{\mu}) = \ln(\tilde{\eta}_i(\boldsymbol{\mu}))$ for $i = 1, \cdots, n$. Then, equation (22) can be rewritten as an implicit funtion form:

$$\mu_{ij}^* = (\overline{\mu}_{.j}^{i*})^{\lambda_d} (\overline{\mu}_{i.}^{j*})^{\lambda_o} (\overline{\mu}_{..}^{ij*})^{\lambda_w} \cdot \tilde{\alpha}_j(\boldsymbol{\mu}^*) \cdot \tilde{\eta}_i(\boldsymbol{\mu}^*) \cdot \mu_{ij}^B. \tag{23}$$

The superscript "$*$" in the equation above denotes the optimal flow. Since all the components in equation (23) are positive, we can have the following log-transformed vector notation:

$$\ln \boldsymbol{\mu}^* = \mathbf{A} \ln \boldsymbol{\mu}^* + \tilde{\mathbf{z}}(\boldsymbol{\mu}^*), \tag{24}$$

where $\mathbf{A} = \lambda_d(I_n \otimes W) + \lambda_o(W \otimes I_n) + \lambda_w(W \otimes W)$, and $\tilde{\mathbf{z}}(\boldsymbol{\mu}^*)$ is an $N \times 1$ vector having $\tilde{z}_{ij}(\boldsymbol{\mu}^*) = \ln(\tilde{\alpha}_j(\boldsymbol{\mu}^*) \cdot \tilde{\eta}_i(\boldsymbol{\mu}^*) \cdot \mu_{ij}^B)$ as its $(j-1)n + i$th element.

As a first intermediate step, we will find a unique representation of $\boldsymbol{\mu}^*$ as a function of $\tilde{\mathbf{z}}(\boldsymbol{\mu}^*)$. If $\rho_{\text{spec}}(\mathbf{A}) < 1$, we have a unique solution to equation (24): $\ln \boldsymbol{\mu}^* = \mathbf{S}^{-1}\tilde{\mathbf{z}}(\boldsymbol{\mu}^*)$ where $\mathbf{S} = I_N - \mathbf{A}$. Then,

$$\begin{aligned} \mu_{ij}^* &= \prod_{k=1}^n \prod_{l=1}^n \exp\left(s_{ij,kl}\tilde{z}_{kl}(\boldsymbol{\mu}^*)\right) \\ &= \exp\left(\sum_{k=1}^n \sum_{l=1}^n s_{ij,kl}\tilde{z}_{kl}(\boldsymbol{\mu}^*)\right) \\ &= \exp\left(\sum_{k=1}^n \sum_{l=1}^n s_{ij,kl}\left(\sum_{m=1}^K \beta_m \ln(D_{kl,m}) + \alpha_l(\boldsymbol{\mu}^*) + \eta_k(\boldsymbol{\mu}^*)\right)\right) \end{aligned} \tag{25}$$

since $\alpha_l(\boldsymbol{\mu}) = \ln(\tilde{\alpha}_l(\boldsymbol{\mu}))$ and $\eta_k(\boldsymbol{\mu}) = \ln(\tilde{\eta}_k(\boldsymbol{\mu}))$. Since $x_{kl}'\beta = \sum_{m=1}^K \beta_m \ln(D_{kl,m})$ (i.e., $x_{kl} = (\ln(D_{kl,1}), \cdots, \ln(D_{kl,K}))'$), our econometric model constitutes the semi-reduced form (25) as the conditional expectation of y_{ij}.

If representation (25) is (fully) unique as a function of the exogenous factors, we can identify $\lambda_d^0, \lambda_o^0, \lambda_w^0, \beta_1^0, \cdots, \beta_K^0$, $\alpha_1^0, \cdots, \alpha_n^0, \eta_1^0, \cdots, \eta_n^0$ from our econometric model. Suppose that we identify those parameters. It implies that μ_{ij}^* is identified. The remaining task is to identify/estimate $\varrho^0, \gamma_{1,d}^0, \cdots, \gamma_{L,d}^0, \gamma_{1,o}^0, \cdots, \gamma_{L,o}^0$. Given the exogenous factors, we need to guarantee the uniqueness of $\boldsymbol{\mu}^*$. Under $\rho_{\text{spec}}(\mathbf{A}) < 1$, the weights $s_{ij,kl}$ and the exogenous part $\mu_{ij}^{++} \equiv \exp\left(\sum_{k=1}^n \sum_{l=1}^n s_{ij,kl} \sum_{m=1}^K \beta_m \ln(D_{kl,m})\right)$ are well-defined.

Then, equation (25) can be rewritten as

$$\mu_{ij}^* = \mu_{ij}^{++} \cdot \left(\prod_{k=1}^n \prod_{l=1}^n \tilde{\alpha}_l^{s_{ij,kl}}(\boldsymbol{\mu}^*)\right) \cdot \left(\prod_{k=1}^n \prod_{l=1}^n \tilde{\eta}_k^{s_{ij,kl}}(\boldsymbol{\mu}^*)\right), \text{ for } i, j = 1, \cdots, n, \tag{26}$$

where

$$\ln(\tilde{\alpha}_l(\boldsymbol{\mu})) = -\frac{1}{2}\ln(G^W) + \ln(G_l) + \ln(\Pi_l^{\varrho-1}(\boldsymbol{\mu})) + \ln(\mu_l^O)$$

and

$$\ln(\tilde{\eta}_k(\boldsymbol{\mu})) = -\frac{1}{2}\ln(G^W) + \ln(G_k) + \ln(P_k^{\varrho-1}(\boldsymbol{\mu})) + \ln(\mu_k^D).$$

Consequently, equation (26) can be simplified as the following additive form:

$$\begin{aligned} \ln(\mu^*_{ij}) =& \ln(\mu^{++}_{ij}) + \sum_{k=1}^n \sum_{l=1}^n s_{ij,kl} \left(\ln(\tilde{\alpha}_l(\boldsymbol{\mu}^*)) + \ln(\tilde{\eta}_k(\boldsymbol{\mu}^*))\right) \\ \Leftrightarrow \ln\left(\boldsymbol{\mu}^*\right) =& \Psi\left(\boldsymbol{\mu}^*, \boldsymbol{\mu}^{++}, \boldsymbol{\mu}^+\right), \end{aligned} \tag{27}$$

where $\boldsymbol{\mu}^{++} = \left(\mu^{++}_{11}, \cdots, \mu^{++}_{n1}, \cdots, \mu^{++}_{1n}, \cdots, \mu^{++}_{nn}\right)'$ and $\boldsymbol{\mu}^{+} = \left(\mu^{+}_{11}, \cdots, \mu^{+}_{n1}, \cdots, \mu^{+}_{1n}, \cdots, \mu^{+}_{nn}\right)'$. Given $\boldsymbol{\mu}^{++}$ and $\boldsymbol{\mu}^+$, hence, we want to find conditions to make $\Psi(\cdot, \boldsymbol{\mu}^{++}, \boldsymbol{\mu}^+)$ be a contraction mapping.

A sufficient condition for the uniqueness of $\boldsymbol{\mu}^*$ is that the maximum absolute row sum of the Jacobian matrix is less than one:

$$\left\| \frac{\partial \Psi\left(\boldsymbol{\mu}, \boldsymbol{\mu}^{++}, \boldsymbol{\mu}^+\right)}{\partial \ln(\boldsymbol{\mu})'} \right\|_\infty < 1.$$

For this, consider $\frac{\partial \Psi_{ij}\left(\boldsymbol{\mu}, \boldsymbol{\mu}^{++}, \boldsymbol{\mu}^+\right)}{\partial \ln(\mu_{kl})}$, which is the $((j-1)n+i, (l-1)n+k)$-element of $\frac{\partial \Psi\left(\boldsymbol{\mu}, \boldsymbol{\mu}^{++}, \boldsymbol{\mu}^+\right)}{\partial \ln(\boldsymbol{\mu})'}$:

$$\begin{aligned} \frac{\partial \Psi_{ij}\left(\boldsymbol{\mu}, \boldsymbol{\mu}^{++}, \boldsymbol{\mu}^+\right)}{\partial \ln(\mu_{kl})} =& \sum_{p=1}^n \sum_{q=1}^n s_{ij,pq} \left(\frac{\partial \ln(\tilde{\alpha}_q(\boldsymbol{\mu}))}{\partial \ln(\mu_{kl})} + \frac{\partial \ln(\tilde{\eta}_p(\boldsymbol{\mu}))}{\partial \ln(\mu_{kl})} \right) \\ =& - \mu_{kl} \sum_{p=1}^n \sum_{q=1}^n s_{ij,pq} \left(\frac{1}{\Pi_q^{\varrho-1}(\boldsymbol{\mu})} \frac{\partial \Pi_q^{\varrho-1}(\boldsymbol{\mu})}{\partial \mu_{kl}} + \frac{1}{P_p^{\varrho-1}(\boldsymbol{\mu})} \frac{\partial P_p^{\varrho-1}(\boldsymbol{\mu})}{\partial \mu_{kl}} \right). \end{aligned}$$

Consequently, a sufficient condition can be provided by

$$\sup_{i,j} \sum_{k=1}^n \sum_{l=1}^n \left| \sum_{p=1}^n \sum_{q=1}^n s_{ij,pq} \left(\frac{\partial \Pi_q^{\varrho-1}(\boldsymbol{\mu})}{\partial \mu_{kl}} \frac{\mu_{kl}}{\Pi_q^{\varrho-1}(\boldsymbol{\mu})} + \frac{\partial P_p^{\varrho-1}(\boldsymbol{\mu})}{\partial \mu_{kl}} \frac{\mu_{kl}}{P_p^{\varrho-1}(\boldsymbol{\mu})} \right) \right| < 1.$$

This condition restricts the cumulative influence on the fixed-effect components from a marginal change of μ_{kl}. Note that the multilateral resistance terms are affected by a marginal change of μ_{kl}, and these terms are only varying factors in $\alpha_1(\boldsymbol{\mu})$, $\cdots$, $\alpha_n(\boldsymbol{\mu})$, $\eta_1(\boldsymbol{\mu})$, $\cdots$, and $\eta_n(\boldsymbol{\mu})$ (Note that $G_1, \cdots, G_n$ themselve are exogenously given. In contrast, each distribution in G_l is affected by a change of μ_{kl}). Hence, this condition is satisfied when a small change of μ_{kl} does not yield dramatic changes in the multilateral resistance terms.

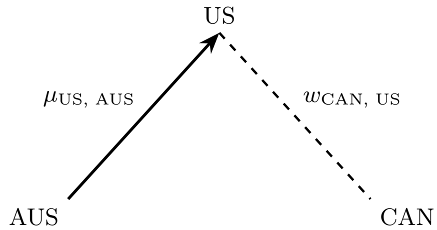
Interpretations. Each w_{ij} captures the strength of proximity (connectivity) between i and j. For intuition, consider a nearest-neighbor specification where $w_{ij} = 1$ if j is the nearest neighbor of i and $w_{ij} = 0$ otherwise. Under this extreme case,

- $\overline{\mu}^i_{\cdot j} = \mu_{kj}$ where k is the country most similar to i (cross-destination weighting on j's outflows);
- $\overline{\mu}^j_{i\cdot} = \mu_{il}$ where l is the country most similar to j (cross-origin weighting on i's inflows);
- $\overline{\mu}^{ij}_{\cdot\cdot} = \mu_{kl}$ where k (resp. l) is the country most similar to i (resp. j).

For concreteness, suppose Canada is the nearest neighbor to the US, and Australia is the nearest neighbor to New Zealand.

1. Common origin+cross-destination linkage: The diagram below illustrates how $\mu_{\text{CAN, AUS}}$ is affected by $\mu_{\text{US, AUS}}$ when CAN and US are close.

Figure 1: Common origin + Cross-destination linkage

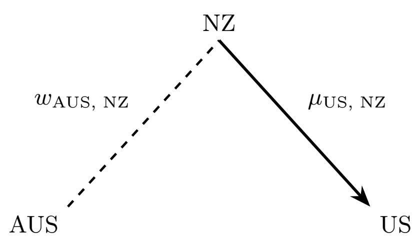


   – If $\lambda_d > 0$, $\mu_{\text{US,AUS}}$ and $\mu_{\text{CAN,AUS}}$ move in the same direction. As AUS $\rightarrow$ US increases, AUS $\rightarrow$ CAN also expands through shared scheduling, fixed logistics, and backhaul synergies via the hub US.[3]

   – When $\lambda_d < 0$, $\mu_{\text{US,AUS}}$ and $\mu_{\text{CAN,AUS}}$ move in opposite directions. With a binding transport capacity from AUS to North America, AUS $\rightarrow$ CAN must shrink when AUS $\rightarrow$ US increases.

2. Common destination+cross-origin linkage: The diagram below shows how $\mu_{\text{US, AUS}}$ is affected by $\mu_{\text{US, NZ}}$ when Australia and New Zealand are close.

Figure 2: Common destination + cross-origin linkage



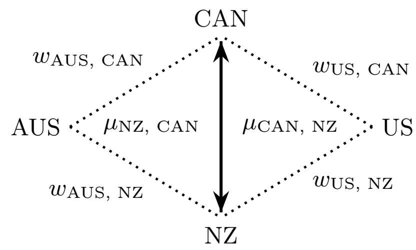
   – If $\lambda_o > 0$, $\mu_{\text{US,AUS}}$ and $\mu_{\text{US,NZ}}$ are positively associated. AUS and NZ coordinate their exports to the US through joint scheduling, consolidation, or hub sharing, lowering costs.

   – If $\lambda_o < 0$, $\mu_{\text{US,AUS}}$ and $\mu_{\text{US,NZ}}$ move in opposite directions. Capacity constraints or competition for slots into the US imply that one country's larger shipments reduce the other's.

3. Cross-origin/-destination linkages: The diagram below describes how $\mu_{\text{US, AUS}}$ is influenced by $\mu_{\text{CAN, NZ}}$ (or $\mu_{\text{NZ, CAN}}$).

[3] Backhaul is the use of return-leg capacity to carry paying cargo, allowing fixed and scheduling costs to be shared across both directions. In our framework, higher expected flows in the reverse or neighboring lanes endogenously reduce bilateral trade costs via consolidation, hub sharing, and multi-leg routing.

Figure 3: Cross-origin/-destination linkages



- If $\lambda_w > 0$, $\mu_{\text{US,AUS}}$ and $\mu_{\text{CAN,NZ}}$ move in the same direction. Strong third-party flows (CAN$\leftrightarrow$NZ) support AUS $\to$ US through hub-and-spoke coordination and multi-leg routing.
- If $\lambda_w < 0$, $\mu_{\text{US,AUS}}$ and $\mu_{\text{CAN,NZ}}$ are negatively associated. Third-party routes absorb transport resources (slots, hub capacity), crowding out AUS $\to$ US.

The nearest-neighbor example clarifies intuition, but in practice, W is row-normalized based on historical trade flows. Then $\overline{\mu}^i_{.j} = \prod_k \mu_{kj}^{w_{ik}}$, $\overline{\mu}^j_{i.} = \prod_l \mu_{il}^{w_{jl}}$, and $\overline{\mu}^{ij}_{..} = \prod_{k,l} \mu_{kl}^{w_{ik}w_{jl}}$ become geometric averages over multiple neighbors. This smooths discrete neighbor switches and allows gradual spillovers.

Positive coefficients $(\lambda_d, \lambda_o, \lambda_w > 0)$ capture coordination, consolidation, and network density that reduce effective costs as neighboring flows expand. On the other hand, negative coefficients accommodate capacity constraints, slot competition, and congestion that raise effective costs when related flows expand.

Econometric point of view. Since connectivities of country i to other countries are heterogeneous across countries, $\overline{\mu}^i_{.j}$, $\overline{\mu}^j_{i.}$ and $\overline{\mu}^{ij}_{..}$ are pair-specific characteristics instead unit-specific ones.

Now taking the natural logarithm on (23), we obtain

$$\begin{aligned}
\ln(\mu_{ij}) = &- \ln(G^W) + \lambda_d \ln(\overline{\mu}^i_{.j}) + \lambda_o \ln(\overline{\mu}^j_{i.}) + \lambda_w \ln(\overline{\mu}^{ij}_{..}) \\
&+ \ln(G_i) + \ln(G_j) + (\varrho - 1)\ln(\Pi_i) + (\varrho - 1)\ln(P_j) \\
&+ \sum_{k=1}^{K} \beta_k \ln(D_{ij,k}) + \sum_{l=1}^{L} \gamma_{l,o} \ln(E_{j,l}) + \sum_{l=1}^{L} \gamma_{l,d} \ln(E_{i,l}).
\end{aligned} \tag{28}$$

The two-way fixed effects, α_j and η_i, absorb the unit-specific terms $\ln(G_j)$, $(\varrho - 1)\ln(P_j)$, $\ln(G_i)$, and $(\varrho - 1)\ln(\Pi_i)$. Since $\overline{\mu}^i_{.j}$, $\overline{\mu}^j_{i.}$ and $\overline{\mu}^{ij}_{..}$ in equation (28) are pair-specific characteristics, the conventional fixed-effect approach omits these terms.

1.3.3 Production-side-based Gravity Equation (Eaton and Kortum 2002)

This subsection reviews the production-side gravity developed by (Eaton and Kortum, 2002). The primary purpose of this review is to highlight the role of the iceberg cost specification in the conventional gravity equation framework. We show how the traditional setting changes once we move beyond this specification.

Suppose there is a continuum of goods, indexed by $\omega \in [0, 1]$, where any country $j = 1, \ldots, n$ can produce any good

ω. Let $\vartheta_j(\omega)$ denote the efficiency or productivity at producing good ω of country j, where $\vartheta_j(\omega)$ is randomly drawn from a Fréchet distribution with parameters $A_j > 0$ (technology/scale parameter, higher means better on average), and $b > 1$ (shape parameter, higher means lower dispersion) such that $F_j(v) := \Pr[\vartheta_j(\omega) \le v] = \exp[-A_j v^{-b}]$ for $v > 0$.

Let $w_j > 0$ be country j's wage. Let $\pi_{ij} \ge 1$ be the trade cost from country j to i. If country j draws productivity $\vartheta_j(\omega)$ for good ω, the unit cost p_{ij} to produce and *deliver* to i is

$$p_{ij}(\omega) := \pi_{ij} \times w_j \times \frac{1}{\vartheta_j(\omega)}.$$

Here, $\frac{w_j}{\vartheta_j(\omega)}$ represents the cost of producing a unit of good ω in country j by constant returns to scale. As an essential assumption, this work supposes that π_{ij} follows the conventional iceberg assumption. Krugman (1995) points out an advantage of this iceberg specification since this assumption implies:

$$\frac{p_{ij}(\omega)}{p_{ij}(\omega')} = \frac{\frac{w_j}{\vartheta_j(\omega)}\pi_{ij}}{\frac{w_j}{\vartheta_j(\omega')}\pi_{ij}} = \frac{\vartheta_j(\omega')}{\vartheta_j(\omega)} \text{ for } \omega \neq \omega'. \tag{29}$$

That is, country j's relative cost of producing any two goods does not rely on the destination.

Our model specification endogenously specifies the cost function, which is beyond the conventional iceberg cost specification. Under Eaton and Kortum's (2002) specification, $\pi_{ii} = 1$ for all i, while $\pi_{ij} > 1$ for $i \neq j$ illustrating positive geographic barrier. Eaton and Kortum (2002) additionally assume that the cross-border arbitrage condition holds based on the iceberg cost specification: it implies effective geographic barriers implied by the triangle inequality. For example, $\pi_{ij} \le \pi_{ik} \cdot \pi_{kj}$ for arbitrary three countries i, j, and k.

In our framework, however, it is not necessary to hold this hypothesis. As an example from Figure 1, if there is a routing advantage, it is possible to have[4]

$$\underbrace{\pi_{\text{US, AUS}}(\boldsymbol{\mu}) + \pi_{\text{CAN, US}}(\boldsymbol{\mu})}_{\text{cost for AUS to US and CAN by routing}} \quad \le \quad \underbrace{\pi^{+}_{\text{US, AUS}} + \pi^{+}_{\text{CAN, AUS}}}_{\text{separated costs for AUS to US and CAN}}.$$

The left-hand side above describes the total trade costs when AUS tries to send its products to CAN through the US, while the right-hand side shows the total cost of AUS when AUS sends its products to the US and to CAN separately (If we consider possible backhaul synergies, the difference between the two scenarios might be larger). Intuitively, the trade cost of AUS to the US and CAN can be reduced by leveraging network information compared to the scenario where AUS separately sends its products to the US and CAN (when $\tilde{\lambda}_d > 0$). The second example from Figure 2 describes the following scenario:

$$\underbrace{\pi_{\text{US, AUS}}(\boldsymbol{\mu}) + \pi_{\text{US, NZ}}(\boldsymbol{\mu})}_{\text{cost for AUS and NZ to US by consolidating shipments}} \quad \le \quad \underbrace{\pi^{+}_{\text{US, AUS}} + \pi^{+}_{\text{US, NZ}}}_{\text{cost for AUS to US + that for NZ to US}}.$$

This means that the costs of two countries, AUS and NZ, can be lower than the costs when AUS and NZ send their products to the US without negotiation.

Krugman's (1995) point from the iceberg cost specification is that the relative price between two goods (produced

[4] In levels, the triangle inequality under iceberg costs is multiplicative. For intuition, we use its additive (log) form here.

in country j) does not depend on the destination. Since our framework does not specify a product-specific trade cost, our framework also satisfies (29). As an extension, if we specify a trade cost as a function of product-specific factors (i.e., $\pi_{ij}(\boldsymbol{\mu}, \omega) = \pi^e_{ij}(\boldsymbol{\mu}, \omega) \cdot \pi^+_{ij}$), (29) would be violated.

Now let's return to solving the model. Consider country i's side. Country i would buy from whichever j is cheapest, i.e., country i selects

$$J_i(\omega) := \arg\min_{j=1,\cdots,n} p_{ij}(\omega) = \arg\min_{j=1,\cdots,n} \left\{ \frac{w_j \pi_{ij}}{\vartheta_j(\omega)} \right\}.$$

Also, we define

$$p_i(\omega) = \min_{j=1,\cdots,n} p_{ij}(\omega).$$

Then, the CDF of p_{ij} is

$$\begin{aligned}
G_{ij}(p) &= \Pr[p_{ij}(\omega) \le p] \\
&= \Pr\left[\frac{w_j \pi_{ij}}{\vartheta_j(\omega)} \le p\right] \\
&= \Pr\left[\vartheta_j(\omega) \ge \frac{w_j \pi_{ij}}{p}\right] \\
&= 1 - F_j\left(\frac{w_j \pi_{ij}}{p}\right) \\
&= 1 - \exp\left(-A_j \left(\frac{w_j \pi_{ij}}{p}\right)^{-b}\right)
\end{aligned} \tag{30}$$

by the assumption on $\vartheta_j(\omega)$. Based on this, we can also derive the CDF of $p_i(\omega)$:

$$\begin{aligned}
G_i(p) &= \Pr[p_i(\omega) \le p] \\
&= \Pr\left[\min_{j=1,\cdots,n} p_{ij}(\omega) \le p\right] \\
&= 1 - \Pr\left[\min_{j=1,\cdots,n} p_{ij}(\omega) > p\right] \\
&= 1 - \Pr\left[\{p_{i1}(\omega) > p\} \cap \{p_{i2}(\omega) > p\} \cap \cdots \cap \{p_{in}(\omega) > p\}\right] \\
&= 1 - \prod_{j=1}^{n} \Pr[p_{ij}(\omega) > p] \\
&= 1 - \prod_{j=1}^{n} \left(1 - \Pr[p_{ij}(\omega) \le p]\right) \\
&= 1 - \prod_{j=1}^{n} \exp\left(-A_j \left(\frac{w_j \pi_{ij}}{p}\right)^{-b}\right) \quad \text{by (30)} \\
&= 1 - \exp\left(-\sum_{j=1}^{n} A_j \left(\frac{w_j \pi_{ij}}{p}\right)^{-b}\right).
\end{aligned} \tag{31}$$

Equation (31) is the answer to the one key question of Eaton and Kortum (2002): what is the distribution of product prices in destination i Notably, the fifth equality in (31) holds when $p_{i1}(\omega), \cdots, p_{in}(\omega)$ are mutually independent (it follows from i.i.d. Fréchet draws across origins.). On the other hand, our framework does not allow us to hold the fifth equality in (31) since π_{ij} in $p_{ij}(\omega)$ is endogenized. Further, $p_i(\omega) = \min_{j=1,\cdots,n} p_{ij}(\omega)$ might not hold in our framework since $p_i(\omega)$ is determined by the entire trade network with countries' proximities (i.e., $\pi_{ij}(\boldsymbol{\mu})$ creates cross-origin dependence among $p_{ij}(\omega)$s). Instead, we expect that $p_i(\omega)$ is characterized by the joint distribution of $p_{ij}(\omega)$ for $i, j = 1, \cdots, n$ under our specification. By the motivation of extending the independent assumption on productivity draws, Lind and Ramondo (2023) consider the joint distribution specification of productivity across countries.[5]

By (30) and (31), we are ready to provide an answer to the second question of Eaton and Kortum (2002): what is the fraction s_{ij} of products in country i that originate from j? Observe

$$
\begin{aligned}
s_{ij} = & \underbrace{\Pr[p_{ij}(\omega) < \min_{k\neq j} p_{ik}(\omega)]}_{\text{probability that country j's price to i is the lowest one}} \\
= & \int_0^\infty \underbrace{\int_0^\infty \mathbb{I}\{p_{ij}(\omega) < \min_{k\neq j} p_{ik}(\omega)\} dG^*_{ij}(p')}_{=\Pr[\min_{k\neq j} p_{ik}(\omega)>p] \text{ when } p_{ij}(\omega)=p} dG_{ij}(p) \text{ by the definition} \\
= & \int_0^\infty \Pr[\min_{k\neq j} p_{ik}(\omega)] dG_{ij}(p) \\
= & \int_0^\infty \Pr\left[\cap_{k\in\{1,\cdots,n\}\setminus\{j\}} \{p_{ik}(\omega) > p\}\right] dG_{ij}(p) \\
= & \int_0^\infty \left(\prod_{k\in\{1,\cdots,n\}\setminus\{j\}} (1 - G_{ik}(p)) \right) dG_{ij}(p)
\end{aligned} \tag{33}
$$

[5] In detail, Lind and Ramondo's (2023) assumption specifies:

$$
\Pr[\vartheta_{i1}(\omega) \le v_1, \cdots, \vartheta_{in}(\omega) \le v_n] = \exp\left[- \left(\sum_{j=1}^n \left(A_{ij} v_j^{-b} \right)^{\frac{1}{1-\varpi}} \right)^{1-\varpi} \right]. \tag{32}
$$

Here,

- A_{ij} is the scale parameter showing absolute advantage of countries;
- $b > 0$ is the shape parameter (leading to $\Pr[\vartheta_{ij}(\omega) \le v] = \exp(-A_{ij} v^{-b})$); and
- $\varpi \in [0, 1)$ characterizes correlation in origins' productivities. If $\varpi = 0$, this specification implies the independent productivity draws (Eaton and Kortum, 2002). On the other hand, if $\varpi \to 1$, the relative productivity between any two products is identical across countries (no comparative advantage in any product, implying no gains from trade).

Indeed, (32) is extended from a univariate Fréchet distribution, i.e.,

$$
\Pr[\vartheta_{i1}(\omega) \le v_1, \cdots, \vartheta_{in}(\omega) \le v_n] = \exp\left[-G^i \left(A_{i1} v_1^{-b}, \cdots, A_{in} v_n^{-b} \right) \right],
$$

where $G^i(\cdot)$ is a correlation function. In this case, the CES correlation function is employed: $G^i(x_1, \cdots, x_n) = \left(\sum_{j=1}^n x_j^{\frac{1}{1-\varpi}} \right)^{1-\varpi}$. Note that a function $G : \mathbb{R}^n_+ \to \mathbb{R}_+$ is a **correlation function** if $\exp[-\ln(u_1, \cdots, u_n)]$ is a **max-stable** copula. Recall that $C : [0,1]^n \to [0,1]$ is a copula if there exists a random vector $(U_1, \cdots, U_n)$ on $[0,1]^n$ such that

$$
C(u_1, \cdots, u_n) = \Pr[U_1 \le u_1, \cdots, U_n \le u_n],
$$

for each $(u_1, \cdots, u_n) \in [0,1]^n$. Given a random vector $(X_1, \cdots, X_n)$, hence, its copula is

$$
C(u_1, \cdots, u_n) = \Pr[F_1(X_1) \le u_1, \cdots, F_n(X_n) \le u_n],
$$

where $F_i(x) = \Pr[X_i \le x]$ for $i = 1, \cdots, n$. This is because $F(X) \sim \mathcal{U}[0,1]$ for any random variable X. Then, if $C(u_1, \cdots, u_n) = C\left(u_1^{1/m}, \cdots, u_n^{1/m} \right)^m$ for any $m > 0$ and for all $(u_1, \cdots, u_n) \in [0,1]^n$, C is **max-stable**.

where $p^*_{ij}(\omega) = \min_{k\neq j} p_{ik}(\omega)$ and $G^*_{ij}(\cdot)$ denotes the CDF of $p^*_{ij}(\omega)$. Note that

$$\prod_{k\in\{1,\cdots,n\}\setminus\{j\}} (1 - G_{ik}(p)) = \prod_{k\in\{1,\cdots,n\}\setminus\{j\}} \left(-A_k \left(\frac{w_k\pi_{ik}}{p}\right)^{-b}\right) = \exp\left(-\sum_{k\neq j} A_k \left(\frac{w_k\pi_{ik}}{p}\right)^{-b}\right),$$

and

$$dG_{ij}(p) = \frac{d}{dp}\left(1 - \exp\left(-A_j\left(\frac{w_j\pi_{ij}}{p}\right)^{-b}\right)\right) dp = bp^{b-1}\cdot A_j(w_j\pi_{ij})^{-b}\cdot \exp\left(-A_j\left(\frac{w_j\pi_{ij}}{p}\right)^{-b}\right)$$

since $\frac{d}{dp}\left(1 - \exp\left(-A_j\left(\frac{w_j\pi_{ij}}{p}\right)^{-b}\right)\right) = -\exp\left(-A_j\left(\frac{w_j\pi_{ij}}{p}\right)^{-b}\right)\cdot -bp^{b-1}A_j(w_j\pi_{ij})^{-b}$. From (33), we have

$$\begin{aligned} s_{ij} &= \int_0^\infty \left(\prod_{k\in\{1,\cdots,n\}\setminus\{j\}} (1 - G_{ik}(p))\right) dG_{ij}(p) \\ &= \int_0^\infty \exp\left(-\sum_{j=1}^n A_j\left(\frac{w_j\pi_{ij}}{p}\right)^{-b}\right)\cdot bp^{b-1}A_j(w_j\pi_{ij})^{-b}dp \\ &= A_j(w_j\pi_{ij})^{-b}\cdot\int_0^\infty bp^{b-1}\cdot\exp\left(-p^b\Upsilon_i\right)dp \\ &= \frac{A_j(w_j\pi_{ij})^{-b}}{\Upsilon_i} \end{aligned} \tag{34}$$

where $\Upsilon_i := \sum_{j=1}^n A_j(w_j\pi_{ij})^{-b}$. The last relation holds since

$$\int_0^\infty bp^{b-1}\cdot\exp\left(-p^b\Upsilon_i\right)dp = \int_0^\infty \exp\left(-\Upsilon_i x\right)dx = -\frac{1}{\Upsilon_i}\exp\left(-\Upsilon_i x\right)|_0^\infty = \frac{1}{\Upsilon_i}.$$

Thus,

$$s_{ij} = \frac{A_j(w_j\pi_{ij})^{-b}}{\sum_{k=1}^n A_k(w_k\pi_{ik})^{-b}}.$$

Given a fraction s_{ij} of goods originated from country j, the total value of imports from j to i is

$$\mu_{ij} = G_i \times s_{ij} = G_i \times \frac{A_j(w_j\pi_{ij})^{-b}}{\sum_{k=1}^n A_k(w_k\pi_{ik})^{-b}} = \underbrace{\frac{G_i}{\sum_{k=1}^n A_k(w_k\pi_{ik})^{-b}}}_{\text{country } i\text{-specific factor}} \times \underbrace{A_j w_j^{-b}}_{\text{country } j\text{-specific factor}} \times \pi_{ij}^{-b}, \tag{35}$$

which is the Eaton and Kortum (2002) gravity equation. In contrast to Anderson and van Wincoop (2003), it is not possible to endogenize π_{ij} in the same way since equation (35) is derived from the price distributions. To relate the price determination mechanism and leverage network information, we may need to specify the joint distribution of $p_{ij}(\omega)$s.

1.3.4 LeSage and Pace's (2008) model

This subsection reviews the spatial OD-flow specification of LeSage and Pace (2008), a reduced-form (non-microfounded) yet well-defined network model that underlies subsequent OD-flow frameworks (e.g., our model; Jeong and Lee, 2024).

We emphasize how an $N \times N$ ($N = n^2$) *network multiplier* matrix arises from an $n \times n$ connectivity matrix.

Model. LeSage and Pace (2008) posit the log-additive OD SAR model:

$$\ln y_{ij} = \lambda_d \sum_{k=1}^{n} w_{ik} \ln y_{kj} + \lambda_o \sum_{l=1}^{n} w_{jl} \ln y_{il} + \lambda_w \sum_{k=1}^{n}\sum_{l=1}^{n} w_{ik}w_{jl} \ln y_{kl} + x_{ij}'\beta + v_{ij}, \tag{36}$$

with $v_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_v^2)$. In vector form,

$$\ln(\mathbf{y}) = (\lambda_d(I_n \otimes W) + \lambda_o(W \otimes I_n) + \lambda_w(W \otimes W)) \ln(\mathbf{y}) + \mathbf{X}\beta + \mathbf{v}.$$

$I_n \otimes W$, $W \otimes I_n$, and $W \otimes W$ encode destination-, origin-, and cross-origin/destination spillovers, respectively.

Link-level interpretation. Note that the $((j-1)n+i, (l-1)n+k)$ element of each matrix component characterizes the network influence from pair kl to ij. The details are below:

- $I_n \otimes W$: $\mathbb{I}\{j = l\}\, w_{ik}$ is active if (i) $\lambda_d \neq 0$, (ii) common origin $j = l$, and (iii) destination i is connected to k ($w_{ik} > 0$).
- $W \otimes I_n$: $\mathbb{I}\{i = k\}\, w_{jl}$ is active if (i) $\lambda_o \neq 0$, (ii) common destination $i = k$, and (iii) origin j is connected to l ($w_{jl} > 0$).
- $W \otimes W$: $w_{ik}w_{jl}$ is active if (i) $\lambda_w \neq 0$, (ii) i is connected to k and j is connected to l.

Equilibrium uniqueness and network multiplier matrix. Recall $\mathbf{S} = I_N - \mathbf{A}$ where $\mathbf{A} = \lambda_d(I_n \otimes W) + \lambda_o(W \otimes I_n) + \lambda_w(W \otimes W)$. If $\mathbf{S}$ is invertible,

$$\ln(\mathbf{y}) = \mathbf{S}^{-1}(\mathbf{X}\beta + \mathbf{v}).$$

Then, $\mathbf{S}^{-1}$ serves as the $N \times N$ network multiplier that aggregates higher-order OD-path spillovers induced by W. Hence, two important issues exist here: (i) the invertibility condition of $\mathbf{S}$ and (ii) the detailed structure of $\mathbf{S}^{-1}$.

Issue 1: Invertibility of S. Assumption 2.4 (i) in the main draft (i.e., $\rho_{\text{spec}}(\mathbf{A}) < 1$) is introduced for well-definedness of the Neumann series expansion.

Here, we elaborate on this condition by assuming that W is a row-normalized matrix from a symmetric matrix $\widetilde{W} = (\widetilde{w}_{ij})$. That is, $W = \text{Diag}^{\text{sum}}(\widetilde{W})^{-1}\widetilde{W}$ with $\widetilde{W}' = \widetilde{W}$. Here, note that $\text{Diag}^{\text{sum}}(\widetilde{W}) = \text{diag}\left(\sum_{j=1}^{n} \widetilde{w}_{1j}, \cdots, \sum_{j=1}^{n} \widetilde{w}_{nj}\right)$. By the spectral theorem,

$$\widetilde{W} = \tilde{\tilde{Q}} D \tilde{\tilde{Q}}',$$

where $D = \text{diag}(\varphi_1, \cdots, \varphi_n)$ the (real-valued) eigenvalue matrix and $\tilde{\tilde{Q}}$ is the eigenvector matrix. Then,

$$\begin{aligned}
W =& \text{Diag}^{\text{sum}}(\widetilde{W})^{-1}\widetilde{W} \\
=& \text{Diag}^{\text{sum}}(\widetilde{W})^{-\frac{1}{2}} \text{Diag}^{\text{sum}}(\widetilde{W})^{-\frac{1}{2}} \widetilde{W} \text{Diag}^{\text{sum}}(\widetilde{W})^{-\frac{1}{2}} \text{Diag}^{\text{sum}}(\widetilde{W})^{\frac{1}{2}} \\
=& \text{Diag}^{\text{sum}}(\widetilde{W})^{-\frac{1}{2}} \widetilde{Q} D \widetilde{Q}' \text{Diag}^{\text{sum}}(\widetilde{W})^{\frac{1}{2}} \\
=& QDQ^{-1},
\end{aligned}$$

where $\widetilde{Q} = \text{Diag}^{\text{sum}}(\widetilde{W})^{-\frac{1}{2}}\tilde{\widetilde{Q}}$ and $Q = \text{Diag}^{\text{sum}}(\widetilde{W})^{-\frac{1}{2}}\widetilde{Q}$. For the last relationship, observe that $Q^{-1} = \left(\text{Diag}^{\text{sum}}(\widetilde{W})^{-\frac{1}{2}}\widetilde{Q}\right)^{-1} = \widetilde{Q}^{-1}\text{Diag}^{\text{sum}}(\widetilde{W})^{\frac{1}{2}} = \widetilde{Q}'\text{Diag}^{\text{sum}}(\widetilde{W})^{\frac{1}{2}}$ since $\widetilde{Q}^{-1} = \widetilde{Q}'$.

Observe that the three matrices, $I_n \otimes W$, $W \otimes I_n$, and $W \otimes W$, share the same eigenvector basis. In detail,

$$
\begin{aligned}
(I_n \otimes W)(q_i \otimes q_j) =& q_i \otimes W q_j = q_i \otimes \varphi_j q_j = \varphi_j (q_i \otimes q_j), \\
(W \otimes I_n)(q_i \otimes q_j) =& W q_i \otimes q_j = \varphi_i q_i \otimes q_j = \varphi_i (q_i \otimes q_j), \text{ and} \\
(W \otimes W)(q_i \otimes q_j) =& W q_i \otimes W q_j = \varphi_i q_i \otimes \varphi_j q_j = \varphi_i \varphi_j (q_i \otimes q_j),
\end{aligned}
$$

where q_i is the ith column vector of Q. Consequently, we have

$$\mathbf{A}(q_i \otimes q_j) = \left(\lambda_d \varphi_j + \lambda_o \varphi_i + \lambda_w \varphi_i \varphi_j\right)(q_i \otimes q_j) \text{ for } i, j = 1, \cdots, n.$$

There are two notable features in characterization of $\rho_{\text{spec}}(\mathbf{A}) < 1$. First, the minimum eigenvalue of W plays a key role here. To see this, consider the traditional SAR model (equation (1) in the main draft) and note that W is row-normalized and its diagonal elements are zero. It implies that $\varphi_{\max} := \max\{\varphi_1, \cdots, \varphi_n\} = 1$ and $\varphi_{\min} := \min\{\varphi_1, \cdots, \varphi_n\} < 0$ since $\text{tr}(W) = \sum_{i=1}^n \varphi_i = 0$. The lemma below describes the properties of $\varphi_{\min}$.

Lemma 1.1. $-1 \leq \varphi_{\min} < 0$. If W is bipartite, $\varphi_{\min} = -1$ (vice versa). Otherwise, $-1 < \varphi_{\min} < 0$.

Proof of Lemma 1.1. By the eigenvalue and eigenvector relationship,

$$Wq = \varphi q.$$

First, find k such that $q_k = \max_{i=1,\cdots,n} |q_i| > 0$. Since $\varphi q_k = (Wq)_k = \sum_{j=1}^n w_{kj} q_j$,

$$|\varphi q_k| = |\sum_{j=1}^n w_{kj} q_j| \leq \sum_{j=1}^n w_{kj} |q_j| \leq |q_k| \sum_{j=1}^n w_{kj} = |q_k|.$$

This implies $|\varphi| \leq 1$.[6]

Suppose that W is constructed by a bipartite network. Thenm all the vertices (agents) can be divided into two disjoint and independent sets $\mathcal{U}$ and $\mathcal{V}$, i.e., $\{1, \cdots, n\} = \mathcal{U} \cup \mathcal{V}$, $\mathcal{U} \cap \mathcal{V} = \varnothing$, and $w_{ij} > 0$ if $i \in \mathcal{U}$ and $j \in \mathcal{V}$; $w_{ij} = 0$,

[6] By the Gershgorin circle theorem, we also have

$$|\varphi - w_{ii}| = |\varphi| \leq \sum_{j=1, j \neq i}^n |w_{ij}| = \sum_{j=1}^n w_{ij} = 1.$$

since $w_{ii} = 0$ for all $i = 1, \cdots, n$ and $\sum_{j=1}^n w_{ij} = 1$.

otherwise. Define $z = (z_1, \cdots, z_n)'$, where $z_i = 1$ if $i \in \mathcal{U}$ and $z_i = -1$ if $i \in \mathcal{V}$. Then, for arbitrary i, observe

$$\begin{aligned}
(Wz)_i &= \sum_{j=1}^{n} w_{ij} z_j \\
&= \frac{1}{\sum_{k=1}^{n} \widetilde{w}_{ik}} \sum_{j=1}^{n} \widetilde{w}_{ij} z_j \\
&= \frac{1}{\sum_{k=1}^{n} \widetilde{w}_{ik}} \sum_{j=1}^{n} \mathbb{I}\{j \text{ is an opponent of } i\} \widetilde{w}_{ij}(-z_i) \\
&= -z_i \frac{1}{\sum_{k=1}^{n} \widetilde{w}_{ik}} \sum_{j=1}^{n} \mathbb{I}\{j \text{ is an opponent of } i\} \widetilde{w}_{ij} = -z_i.
\end{aligned}$$

Hence, $\varphi_{\min} = -1$.

Conversely, suppose that there exists $z \neq 0$ such that $Wz = z$. For arbitrary i,

$$-z_i = \sum_{j=1}^{n} w_{ij} z_j.$$

Since all w_{ij}s are nonnegative, $z_j = -z_i$ if $w_{ij} > 0$ to hold the equality above. It implies that W comes from a bipartite network.■

Note that $\varphi_{\min}$ measures the periodicity of a network, describing how much the network exhibits oscillatory or polarized patterns. In the economic literature, Bramoulle et al. (2014) conduct a detailed analysis of this issue. When $\varphi_{\min}$ approaches -1, W becomes strong bipartiteness (i.e., odd–even oscillations). On the other hand, if $\varphi_{\min} \to 0$, W tends to have a high averaging rate (i.e., W averages out heterogeneity, so that each node's value becomes a smooth local average of its neighbors, and differences vanish quickly). Indeed, the averaging rate is governed by $\max\{|\varphi_2|, |\varphi_{\min}|\}$ in a row-normalized undirected network. In detail, if $\varphi_{\min} \to 0$, the number of odd cycles becomes richer (on the other hand, there is no odd cycle if $\varphi_{\min} = -1$). On the other hand, φ_2 captures expansion/contractive properties. When $\varphi_2 \to 1$, it implies a small spectral gap $1 - \varphi_2$ entailing slow averaing. For details, refer to Chung (1997). As an example, consider $W = \frac{1}{n-1}(l_n l_n' - I_n)$ illustrating the linear-in-mean model's implication. In this case, $\varphi_{\max} = \varphi_1 = 1$ and $\varphi_{\min} = \varphi_2 = \cdots = \varphi_n = \frac{1}{n-1}$ since $\mathrm{tr}(W) = \sum_{i=1}^{n} \varphi_i = 0$. Under a large n, $\max\{|\varphi_2|, |\varphi_{\min}|\} \simeq 0$.

Let $A = \lambda W$ be the counterpart of $\mathbf{A}$ in equation (1). Since an eigenvalue of A is $\lambda \varphi_i$, $\rho_{\text{spec}}(A) = |\lambda|$ if we allow $\lambda > 0$. Hence, the stability condition simply becomes $|\lambda| < 1$. If we restrict the case of $\lambda < 0$, $\rho_{\text{spec}}(A) = \lambda \varphi_{\min} \geq |\lambda|$ since $-1 \leq \varphi_{\min} < 0$. Hence, if $W = \frac{1}{n-1}(l_n l_n' - I_n)$ and $\lambda < 0$, the possible parameter space for λ becomes quite wider. On the other hand, if $W = \begin{bmatrix} \mathbf{0} & \frac{1}{n_1} l_{n_1} l_{n_2}' \\ \frac{1}{n_2} l_{n_2} l_{n_1}' & \mathbf{0} \end{bmatrix}$, the admissible parameter space is always $|\lambda| < 1$.[7]

---

[7]To intuitively explain the two extreme cases, consider the structure of Wy in equation (1). When $W = \begin{bmatrix} \mathbf{0} & \frac{1}{n_1} l_{n_1} l_{n_2}' \\ \frac{1}{n_2} l_{n_2} l_{n_1}' & \mathbf{0} \end{bmatrix}$ (bipartite network) where n_1 denotes the number of the first group and n_2 is the number of the second group,

$$Wy \simeq \begin{pmatrix} \overline{y}_2 \\ \overline{y}_1 \end{pmatrix}$$

under a large n. In this case, if n is large, Wy consists of two distinct values ($\overline{y}_1$ and $\overline{y}_2$). Hence, the source of variation for identifying λ is $\overline{y}_1 \neq \overline{y}_2$.

On the other hand, if $W = \frac{1}{n-1}(l_n l_n' - I_n)$, $Wy \simeq \overline{y} l_n$ when n is large. Then, Wy and the intercept term cannot be distinguished when

Second, we observe that an eigenvalue of $\mathbf{A}$ is $\lambda_d\varphi_j + \lambda_o\varphi_i + \lambda_w\varphi_i\varphi_j$, which is a bilinear map. That is, $b(\varphi_i, \varphi_j) = \lambda_d\varphi_j + \lambda_o\varphi_i + \lambda_w\varphi_i\varphi_j$ for $(\varphi_i, \varphi_j) \in [\varphi_{\min}, 1]^2$ (note that $\varphi_{\min} < 0$). Then, we have the following observations:

- When φ_i is fixed, $b(\varphi_i, \varphi_j) = \lambda_o\varphi_i + (\lambda_d + \lambda_w\varphi_i)\varphi_j$ is a linear function of φ_j. For each $\varphi_i \in [\varphi_{\min}, 1]$, hence,

$$\max_{\varphi_j \in [\varphi_{\min}, 1]} b(\varphi_i, \varphi_j) = \max\{b(\varphi_i, \varphi_{\min}), b(\varphi_i, 1)\}.$$

- Now we observe that the two functions from above,

$$\begin{aligned} b(\varphi_i, \varphi_{\min}) =& \lambda_d\varphi_{\min} + \lambda_o\varphi_i + \lambda_w\varphi_i\varphi_{\min} \text{ and} \\ b(\varphi_i, 1) =& \lambda_d + \lambda_o\varphi_i + \lambda_w\varphi_i, \end{aligned}$$

  are linear in φ_i.

- Hence,

$$\begin{aligned} \max_{\varphi_i \in [\varphi_{\min}, 1]} b(\varphi_i, \varphi_{\min}) =& \max\{\underbrace{\lambda_d\varphi_{\min} + \lambda_o\varphi_{\min} + \lambda_w\varphi_{\min}^2}_{=b(\varphi_{\min}, \varphi_{\min})}, \underbrace{\lambda_d\varphi_{\min} + \lambda_o + \lambda_w\varphi_{\min}}_{=b(1, \varphi_{\min})}\}, \text{ and} \\ \max_{\varphi_i \in [\varphi_{\min}, 1]} b(\varphi_i, 1) =& \max\{\underbrace{\lambda_d + \lambda_o\varphi_{\min} + \lambda_w\varphi_{\min}}_{=b(\varphi_{\min}, 1)}, \underbrace{\lambda_d + \lambda_o + \lambda_w}_{=b(1,1)}\}. \end{aligned}$$

- Hence, we have

$$\rho_{\text{spec}}(\mathbf{A}) = \max\{b(1,1), b(1, \varphi_{\min}), b(\varphi_{\min}, 1), b(\varphi_{\min}, \varphi_{\min})\} < 1, \tag{37}$$

  as a stability condition, where

$$\begin{aligned} b(1,1) =& \lambda_d + \lambda_o + \lambda_w, \\ b(1, \varphi_{\min}) =& \lambda_d\varphi_{\min} + \lambda_o + \lambda_w\varphi_{\min}, \\ b(\varphi_{\min}, 1) =& \lambda_d + \lambda_o\varphi_{\min} + \lambda_w\varphi_{\min}, \text{ and} \\ b(\varphi_{\min}, \varphi_{\min}) =& \lambda_d\varphi_{\min} + \lambda_o\varphi_{\min} + \lambda_w\varphi_{\min}^2. \end{aligned}$$

Here, the arguments for the maximum above are $(1,1)$, $(1, \varphi_{\min})$, $(\varphi_{\min}, 1)$, and $(\varphi_{\min}, \varphi_{\min})$.

Issue 2: Structure of $\mathbf{S}^{-1}$. Our spatial OD flow model captures the intricate spatial relationships among flow outcomes, with each relationship characterized by $s_{ij,kl}$, an element of $\mathbf{S}^{-1}$. In detail,

$$\begin{aligned} \frac{\partial\mu_{ij}}{\partial x_{kl}} &= \beta \cdot \mu_{ij}s_{ij,kl}, \\ \frac{\partial\mu_{ij}}{\partial\alpha_l} &= \mu_{ij}\sum_{k=1}^{n} s_{ij,kl}, \text{ and} \\ \frac{\partial\mu_{ij}}{\partial\eta_k} &= \mu_{ij}\sum_{l=1}^{n} s_{ij,kl}. \end{aligned}$$

n is large.

The signal $s_{ij,kl}$ from one destination-origin pair kl to another ij is determined by a complex network structure that includes two sets of origins and destinations. Hence, understanding the structure of $\mathbf{S}^{-1}$ is critical for explaining the spatial influences that shape flow outcomes.

The trinomial expansion formula gives

$$
\begin{aligned}
s_{ij,kl} &= \left(e'_{n,j} \otimes e'_{n,i}\right) \left(I_N + \sum_{p=1}^{\infty} (\lambda_d(I_n \otimes W) + \lambda_o(W \otimes I_n) + \lambda_w(W \otimes W))^p \right) (e_{n,l} \otimes e_{n,k}) \\
&= \mathbb{I}(j = l, i = k) + \sum_{p=1}^{\infty} \left(e'_{n,j} \otimes e'_{n,i}\right) \mathbf{A}^p \left(e_{n,l} \otimes e_{n,k}\right) \\
&= \mathbb{I}(j = l, i = k) + \sum_{p=1}^{\infty} \sum_{p_1+p_2+p_3=p} \frac{p!}{p_1!p_2!p_3!} \lambda_d^{p_1} \lambda_o^{p_2} \lambda_w^{p_3} \left(W^{p_1+p_3}\right)_{ik} \left(W^{p_2+p_3}\right)_{jl}.
\end{aligned}
$$

Then, the p-th order effect contains (i) the $(p_1 + p_3)$-th order connections between k and i via $W^{p_1+p_3}$ and (ii) $(p_2 + p_3)$-th order connections between j and l by $W^{p_2+p_3}$ such that $p_1 + p_2 + p_3 = p$.

For illustration purposes, we demonstrate the first-order effects (i.e., $p = 1$) and the second-order effects (i.e., $p = 2$). The first-order effects represent *direct* signals between two pairs, weighted by their spatial dependences: (i) If two pairs share the same origin (i.e., $j = l$), the signal between them would reflect the fact that they only vary by their destinations so that it is weighted by their destination-based dependence; (ii) Similarly, if two pairs share the same destination (i.e., $i = k$), the signal between them would be weighted by their origin-based dependence; (iii) Otherwise, two pairs both have distinguished origins and destinations. In this case, the signal between them would be weighted by the product of their dependences in the destination pair and the origin pair. In this manner, the first-order effects are specified as

$$s^{(1)}_{ij,kl} = \lambda_d \mathbf{1}(j = l) w_{ik} + \lambda_o \mathbb{I}(i = k) w_{jl} + \lambda_w w_{ik} w_{jl},$$

where $s^{(p)}_{ij,kl}$ $(p = 1, 2, \cdots)$ indicates the pth-order effects of $s_{ij,kl}$ (i.e., $s_{ij,kl} = \mathbf{1}(i = k, j = l) + \sum_{p=1}^{\infty} s^{(p)}_{ij \leftarrow kl}$). In general, the pth-order effects decompose $s_{ij \leftarrow kl}$ by p-step paths. One may observe that when $p = 2$, the signal with two pairs sharing the same origin (i.e., $j = l$) is decomposed as

$$
\begin{aligned}
& s^{(2)}_{ij,kl} \\
=& \left(e'_{n,j} \otimes e'_{n,i}\right) \mathbf{A}^2 \left(e_{n,l} \otimes e_{n,k}\right) \\
=& \left(e'_{n,j} \otimes e'_{n,i}\right) \begin{pmatrix} \lambda_d^2(I_n \otimes W^2) + \lambda_o^2(W^2 \otimes I_n) + \lambda_w^2(W^2 \otimes W^2) \\ +2\lambda_d\lambda_o(W \otimes W) + 2\lambda_d\lambda_w(W \otimes W^2) + 2\lambda_o\lambda_w(W^2 \otimes W) \end{pmatrix} \left(e_{n,l} \otimes e_{n,k}\right) \\
=& \lambda_d^2 \mathbb{I}(j = l)(W^2)_{ik} + \lambda_o^2 (W^2)_{jl} \mathbb{I}(i = k) + \lambda_w^2 (W^2)_{jl}(W^2)_{ik} + 2\lambda_d\lambda_o w_{jl} w_{ik} + 2\lambda_d\lambda_w w_{jl}(W^2)_{ik} + 2\lambda_o\lambda_w (W^2)_{jl} w_{ik}.
\end{aligned}
$$

2 Theoretical details in statistical analysis

2.1 First- and second-order conditions

Recall that the statistical objective function is:

$$\ell_N(\theta, \boldsymbol{\phi}) = \sum_{i,j=1}^{n} \left(-\mu_{ij}(\theta, \boldsymbol{\phi}) + y_{ij} \ln\left(\mu_{ij}(\theta, \boldsymbol{\phi})\right) - \ln\left(y_{ij}!\right)\right) - \frac{1}{2}\left(\sum_{j=1}^{n} \alpha_j - \sum_{i=1}^{n} \eta_i\right)^2, \tag{38}$$

where $\mu_{ij}(\theta, \boldsymbol{\phi}) = \exp\left(\tilde{\mu}_{ij}(\theta, \boldsymbol{\phi})\right)$ with $\tilde{\mu}_{ij}(\theta, \boldsymbol{\phi}) = \sum_{k,l=1}^{n} s_{ij,kl}(\lambda)\left(x'_{kl}\beta + \alpha_l + \eta_k\right)$. For $i, j = 1, \cdots, n$, let

$$\begin{aligned}
\xi_{ij}(\theta, \boldsymbol{\phi}) &= \frac{y_{ij}}{\mu_{ij}(\theta, \boldsymbol{\phi})} : \text{ multiplicative residual evaluated at } (\theta, \boldsymbol{\phi}), \\
u_{ij}(\theta, \boldsymbol{\phi}) &= \mu_{ij}(\theta, \boldsymbol{\phi})\left(\xi_{ij}(\theta, \boldsymbol{\phi}) - 1\right) = y_{ij} - \mu_{ij}(\theta, \boldsymbol{\phi}) : \text{ additive residual at } (\theta, \boldsymbol{\phi}), \text{ and} \\
z_{ij}\left(\beta, \eta_i, \alpha_j\right) &= x'_{ij}\beta + \alpha_j + \eta_i : \text{ exogenous component evaluated at } \left(\beta, \eta_i, \alpha_j\right).
\end{aligned}$$

For notational convenience, we further define $\boldsymbol{\theta} = \left(\theta', \boldsymbol{\phi}'\right)'$, $\mathbf{W}_d = I_n \otimes W$, $\mathbf{W}_o = W \otimes I_n$, and $\mathbf{W}_w = W \otimes W$.

For a general notation, we observe that

$$\partial_\theta \ell_N(\boldsymbol{\theta}) = \sum_{i,j=1}^{n} \left(\xi_{ij}(\boldsymbol{\theta}) - 1\right) \partial_\theta \mu_{ij}(\boldsymbol{\theta}) = \sum_{i,j=1}^{n} \partial_\theta \tilde{\mu}_{ij}(\boldsymbol{\theta}) u_{ij}(\boldsymbol{\theta})$$

since $\partial_\theta \mu_{ij}(\boldsymbol{\theta}) = \mu_{ij}(\boldsymbol{\theta}) \partial_\theta \tilde{\mu}_{ij}(\boldsymbol{\theta})$. This implies that the moment condition from (38) is

$$\mathbb{E}\left(\partial_{\boldsymbol{\theta}} \tilde{\mu}_{ij}(\boldsymbol{\theta}) u_{ij}(\boldsymbol{\theta})\right) = 0 \text{ if and only if } \boldsymbol{\theta} = \boldsymbol{\theta}^0.$$

On the other hand, the nonlinear two-stage least squares estimator is obtained by

$$\sum_{i,j=1}^{n} \left(y_{ij} - \exp(\tilde{\mu}_{ij}(\boldsymbol{\theta}))\right)^2.$$

The first-order condition is

$$2 \sum_{i,j=1}^{n} \exp\left(\tilde{\mu}_{ij}(\boldsymbol{\theta})\right) \partial_\theta \tilde{\mu}_{ij}(\boldsymbol{\theta}) u_{ij}(\boldsymbol{\theta}) = 0,$$

which implies the following moment condition.

$$\mathbb{E}\left(\underbrace{\exp\left(\tilde{\mu}_{ij}(\boldsymbol{\theta})\right)}_{\text{additional weight}} \partial_{\boldsymbol{\theta}} \tilde{\mu}_{ij}(\boldsymbol{\theta}) u_{ij}(\boldsymbol{\theta})\right) = 0 \text{ if and only if } \boldsymbol{\theta} = \boldsymbol{\theta}^0.$$

Whenever $\tilde{\mu}_{ij}(\boldsymbol{\theta}) > 0$, $\exp\left(\tilde{\mu}_{ij}(\boldsymbol{\theta})\right) > 1$. Moreover, $\exp\left(\tilde{\mu}_{ij}(\boldsymbol{\theta})\right)$ is huge for some ij. One can observe that inefficiency occurs since this method heavily depends on a relatively small number of observations (Silva and Tenreyro (2006, Sec. III A).

The detailed first-order conditions are reported below:

$$\partial_{\lambda_d}\ell_N(\boldsymbol{\theta}) = \sum_{i,j=1}^{n}\left(\sum_{k,l=1}^{n}\left[\mathbf{W}_d\mathbf{S}^{-2}(\lambda)\right]_{ij,kl} z_{kl}(\beta,\eta_k,\alpha_l)\right) u_{ij}(\boldsymbol{\theta}),$$

$$\partial_{\lambda_o}\ell_N(\boldsymbol{\theta}) = \sum_{i,j=1}^{n}\left(\sum_{k,l=1}^{n}\left[\mathbf{W}_o\mathbf{S}^{-2}(\lambda)\right]_{ij,kl} z_{kl}(\beta,\eta_k,\alpha_l)\right) u_{ij}(\boldsymbol{\theta}),$$

$$\partial_{\lambda_w}\ell_N(\boldsymbol{\theta}) = \sum_{i,j=1}^{n}\left(\sum_{k,l=1}^{n}\left[\mathbf{W}_w\mathbf{S}^{-2}(\lambda)\right]_{ij,kl} z_{kl}(\beta,\eta_k,\alpha_l)\right) u_{ij}(\boldsymbol{\theta}), \text{ and}$$

$$\partial_{\beta}\ell_N(\boldsymbol{\theta}) = \sum_{i,j=1}^{n}\left(\sum_{k,l=1}^{n} s_{ij,kl}(\lambda)x_{kl}\right) u_{ij}(\boldsymbol{\theta}),$$

where $[\mathbf{C}]_{ij,kl}$ denotes the $((j-1)n+i,(l-1)n+k)$-element of an N-dimensional square matrix $\mathbf{C}$. We verify that the penalty term does not play a role in the first-order conditions for the main parameters. For the fixed-effect components, observe

$$\partial_{\alpha_l}\ell_N(\boldsymbol{\theta}) = \sum_{i,j=1}^{n}\left(\sum_{k=1}^{n} s_{ij,kl}(\lambda)\right) u_{ij}(\boldsymbol{\theta}) - \underbrace{\left(\sum_{j=1}^{n}\alpha_j - \sum_{i=1}^{n}\eta_i\right)}_{=0}, \text{ and}$$

$$\partial_{\eta_k}\ell_N(\boldsymbol{\theta}) = \sum_{i,j=1}^{n}\left(\sum_{l=1}^{n} s_{ij,kl}(\lambda)\right) u_{ij}(\boldsymbol{\theta}) + \underbrace{\left(\sum_{j=1}^{n}\alpha_j - \sum_{i=1}^{n}\eta_i\right)}_{=0}.$$

By the restriction, note that $\sum_{j=1}^{n}\alpha_j - \sum_{i=1}^{n}\eta_i = 0$ holds. Using the vector notation, we have

$$\begin{pmatrix}\partial_\theta\ell_N(\boldsymbol{\theta})\\ \partial_{\boldsymbol{\phi}}\ell_N(\boldsymbol{\theta})\end{pmatrix} = \begin{pmatrix}\left[\mathbf{W}_d\mathbf{S}^{-2}(\lambda)\mathbf{Z}(\boldsymbol{\theta}), \mathbf{W}_o\mathbf{S}^{-2}(\lambda)\mathbf{Z}(\boldsymbol{\theta}), \mathbf{W}_w\mathbf{S}^{-2}(\lambda)\mathbf{Z}(\boldsymbol{\theta}), \mathbf{S}^{-1}(\lambda)\mathbf{X}\right]'\mathbf{u}(\boldsymbol{\theta})\\ \left(\mathbf{S}^{-1}(\lambda)\mathbf{D}\right)'\mathbf{u}(\boldsymbol{\theta})\end{pmatrix}.$$

Here,

- $\mathbf{Z}(\boldsymbol{\theta}) = \mathbf{X}\beta + \boldsymbol{\alpha}\otimes l_n + l_n\otimes\boldsymbol{\eta}$ with $\mathbf{X} = (x_{ij,k})$ being an $N\times K$ matrix of regressors, $\mathbf{Z} = \mathbf{Z}(\boldsymbol{\theta}^0)$,
- $\mathbf{D} = [\mathbf{I}_n\otimes l_n, l_n\otimes\mathbf{I}_n]$,
- $\mathbf{u}(\boldsymbol{\theta}) = (u_{11}(\boldsymbol{\theta}), u_{21}(\boldsymbol{\theta}), \cdots, u_{n1}(\boldsymbol{\theta}), \cdots, u_{1n}(\boldsymbol{\theta}), u_{2n}(\boldsymbol{\theta}), \cdots, u_{nn}(\boldsymbol{\theta}))'$, and $\mathbf{u} = \mathbf{u}(\boldsymbol{\theta}^0)$.

A general form of the second-order condition is

$$\partial_{\theta\theta}\ell_N(\boldsymbol{\theta}) = \sum_{i,j=1}^{n}\left(-\partial_\theta\tilde{\mu}_{ij}(\boldsymbol{\theta})\partial_\theta\tilde{\mu}_{ij}(\boldsymbol{\theta})'\mu_{ij}(\boldsymbol{\theta}) + u_{ij}(\boldsymbol{\theta})\partial_{\theta\theta}\tilde{\mu}_{ij}(\boldsymbol{\theta})\right),$$

and $\partial_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_N(\boldsymbol{\theta})$ has the following block diagonal structure:

$$\begin{aligned}\partial_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_N(\boldsymbol{\theta}) &= \begin{bmatrix}\partial_{\theta\theta}\ell_N(\boldsymbol{\theta}) & \partial_{\theta\boldsymbol{\alpha}}\ell_N(\boldsymbol{\theta}) & \partial_{\theta\boldsymbol{\eta}}\ell_N(\boldsymbol{\theta})\\ \partial_{\boldsymbol{\alpha}\theta}\ell_N(\boldsymbol{\theta}) & \partial_{\boldsymbol{\alpha}\boldsymbol{\alpha}}\ell_N(\boldsymbol{\theta}) & \partial_{\boldsymbol{\alpha}\boldsymbol{\eta}}\ell_N(\boldsymbol{\theta})\\ \partial_{\boldsymbol{\eta}\theta}\ell_N(\boldsymbol{\theta}) & \partial_{\boldsymbol{\eta}\boldsymbol{\alpha}}\ell_N(\boldsymbol{\theta}) & \partial_{\boldsymbol{\eta}\boldsymbol{\eta}}\ell_N(\boldsymbol{\theta})\end{bmatrix}\\ &= \begin{bmatrix}\partial_{\theta\theta}\ell_N(\boldsymbol{\theta}) & \partial_{\theta\boldsymbol{\phi}}\ell_N(\boldsymbol{\theta})\\ \partial_{\boldsymbol{\phi}\theta}\ell_N(\boldsymbol{\theta}) & \partial_{\boldsymbol{\phi}\boldsymbol{\phi}}\ell_N(\boldsymbol{\theta})\end{bmatrix}.\end{aligned}$$

First, here are the detailed elements of the first block, $\partial_{\theta\theta}\ell_N(\boldsymbol{\theta})$:

$$\begin{aligned}
\partial_{\lambda_d\lambda_d}\ell_N(\boldsymbol{\theta}) =& -\sum_{i,j=1}^{n}\mu_{ij}(\boldsymbol{\theta})\left(\sum_{k,l=1}^{n}\left[\mathbf{W}_d\mathbf{S}^{-2}(\lambda)\right]_{ij,kl} z_{kl}(\beta,\eta_k,\alpha_l)\right)^2\\
&+\sum_{i,j=1}^{n}\sum_{k,l=1}^{n} 2\left[\mathbf{W}_d^2\mathbf{S}^{-3}(\lambda)\right]_{ij,kl} z_{kl}(\beta,\eta_k,\alpha_l)u_{ij}(\boldsymbol{\theta}),\\
\partial_{\lambda_d\lambda_o}\ell_N(\boldsymbol{\theta}) =& -\sum_{i,j=1}^{n}\mu_{ij}(\boldsymbol{\theta})\sum_{k_1,l_1,k_2,l_2=1}^{n}\left[\mathbf{W}_d\mathbf{S}^{-2}(\lambda)\right]_{ij,k_1l_1}\left[\mathbf{W}_o\mathbf{S}^{-2}(\lambda)\right]_{ij,k_2l_2} z_{k_1l_1}(\beta,\eta_{k_1},\alpha_{l_1})z_{k_2l_2}(\beta,\eta_{k_2},\alpha_{l_2})\\
&+\sum_{i,j=1}^{n}\sum_{k,l=1}^{n} 2\left[\mathbf{W}_d\mathbf{W}_o\mathbf{S}^{-3}(\lambda)\right]_{ij,kl} z_{kl}(\beta,\eta_k,\alpha_l)u_{ij}(\boldsymbol{\theta}),\\
\partial_{\lambda_d\lambda_w}\ell_N(\boldsymbol{\theta}) =& -\sum_{i,j=1}^{n}\mu_{ij}(\boldsymbol{\theta})\sum_{k_1,l_1,k_2,l_2=1}^{n}\left[\mathbf{W}_d\mathbf{S}^{-2}(\lambda)\right]_{ij,k_1l_1}\left[\mathbf{W}_w\mathbf{S}^{-2}(\lambda)\right]_{ij,k_2l_2} z_{k_1l_1}(\beta,\eta_{k_1},\alpha_{l_1})z_{k_2l_2}(\beta,\eta_{k_2},\alpha_{l_2})\\
&+\sum_{i,j=1}^{n}\sum_{k,l=1}^{n} 2\left[\mathbf{W}_d\mathbf{W}_w\mathbf{S}^{-3}(\lambda)\right]_{ij,kl} z_{kl}(\beta,\eta_k,\alpha_l)u_{ij}(\boldsymbol{\theta}),\\
\partial_{\lambda_d\beta}\ell_N(\boldsymbol{\theta}) =& -\sum_{i,j=1}^{n}\mu_{ij}(\boldsymbol{\theta})\sum_{k_1,l_1,k_2,l_2=1}^{n}\left[\mathbf{W}_d\mathbf{S}^{-2}(\lambda)\right]_{ij,k_1l_1} s_{ij,k_2l_2}(\lambda)z_{k_1l_1}(\beta,\eta_{k_1},\alpha_{l_1})x_{k_2l_2}\\
&+\sum_{i,j=1}^{n}\sum_{k,l=1}^{n}\left[\mathbf{W}_d\mathbf{S}^{-2}(\lambda)\right]_{ij,kl} x_{kl}u_{ij}(\boldsymbol{\theta}),
\end{aligned}$$

$$\begin{aligned}
\partial_{\lambda_o\lambda_o}\ell_N(\boldsymbol{\theta}) =& - \sum_{i,j=1}^{n} \mu_{ij}(\boldsymbol{\theta}) \left(\sum_{k,l=1}^{n} \left[\mathbf{W}_o\mathbf{S}^{-2}(\lambda)\right]_{ij,kl} z_{kl}(\beta,\eta_k,\alpha_l) \right)^2 \\
&+ \sum_{i,j=1}^{n} \sum_{k,l=1}^{n} 2\left[\mathbf{W}_o^2\mathbf{S}^{-3}(\lambda)\right]_{ij,kl} z_{kl}(\beta,\eta_k,\alpha_l) u_{ij}(\boldsymbol{\theta}), \\
\partial_{\lambda_o\lambda_w}\ell_N(\boldsymbol{\theta}) =& - \sum_{i,j=1}^{n} \mu_{ij}(\boldsymbol{\theta}) \sum_{k_1,l_1,k_2,l_2=1}^{n} \left[\mathbf{W}_o\mathbf{S}^{-2}(\lambda)\right]_{ij,k_1l_1} \left[\mathbf{W}_w\mathbf{S}^{-2}(\lambda)\right]_{ij,k_2l_2} z_{k_1l_1}(\beta,\eta_{k_1},\alpha_{l_1}) z_{k_2l_2}(\beta,\eta_{k_2},\alpha_{l_2}) \\
&+ \sum_{i,j=1}^{n} \sum_{k,l=1}^{n} 2\left[\mathbf{W}_o\mathbf{W}_w\mathbf{S}^{-3}(\lambda)\right]_{ij,kl} z_{kl}(\beta,\eta_k,\alpha_l) u_{ij}(\boldsymbol{\theta}), \\
\partial_{\lambda_o\beta}\ell_N(\boldsymbol{\theta}) =& - \sum_{i,j=1}^{n} \mu_{ij}(\boldsymbol{\theta}) \sum_{k_1,l_1,k_2,l_2=1}^{n} \left[\mathbf{W}_o\mathbf{S}^{-2}(\lambda)\right]_{ij,k_1l_1} s_{ij,k_2l_2}(\lambda) z_{k_1l_1}(\beta,\eta_{k_1},\alpha_{l_1}) x_{k_2l_2} \\
&+ \sum_{i,j=1}^{n} \sum_{k,l=1}^{n} \left[\mathbf{W}_o\mathbf{S}^{-2}(\lambda)\right]_{ij,kl} x_{kl} u_{ij}(\boldsymbol{\theta}),
\end{aligned}$$

$$\begin{aligned}
\partial_{\lambda_w\lambda_w}\ell_N(\boldsymbol{\theta}) =& - \sum_{i,j=1}^{n} \mu_{ij}(\boldsymbol{\theta}) \left(\sum_{k,l=1}^{n} \left[\mathbf{W}_w\mathbf{S}^{-2}(\lambda)\right]_{ij,kl} z_{kl}(\beta,\eta_k,\alpha_l) \right)^2 \\
&+ \sum_{i,j=1}^{n} \sum_{k,l=1}^{n} 2\left[\mathbf{W}_w^2\mathbf{S}^{-3}(\lambda)\right]_{ij,kl} z_{kl}(\beta,\eta_k,\alpha_l) u_{ij}(\boldsymbol{\theta}), \\
\partial_{\lambda_w\beta}\ell_N(\boldsymbol{\theta}) =& - \sum_{i,j=1}^{n} \mu_{ij}(\boldsymbol{\theta}) \sum_{k_1,l_1,k_2,l_2=1}^{n} \left[\mathbf{W}_w\mathbf{S}^{-2}(\lambda)\right]_{ij,k_1l_1} s_{ij,k_2l_2}(\lambda) z_{k_1l_1}(\beta,\eta_{k_1},\alpha_{l_1}) x_{k_2l_2} \\
&+ \sum_{i,j=1}^{n} \sum_{k,l=1}^{n} \left[\mathbf{W}_w\mathbf{S}^{-2}(\lambda)\right]_{ij,kl} x_{kl} u_{ij}(\boldsymbol{\theta}),
\end{aligned}$$

and

$$\partial_{\beta\beta}\ell_N(\boldsymbol{\theta}) = - \sum_{i,j=1}^{n} \mu_{ij}(\boldsymbol{\theta}) \sum_{k_1,l_1,k_2,l_2=1}^{n} s_{ij,k_1l_1}(\lambda) s_{ij,k_2l_2}(\lambda) x'_{k_1l_1} x_{k_2l_2}.$$

Second, consider the second block, $\partial_{\theta\boldsymbol{\phi}}\ell_N(\boldsymbol{\theta})$:

$$\begin{aligned}
\partial_{\lambda_d\alpha_l}\ell_N(\boldsymbol{\theta}) =& \sum_{k=1}^{n}\sum_{i,j=1}^{n}\left[\mathbf{W}_d\mathbf{S}^{-2}(\lambda)\right]_{ij,kl}u_{ij}(\boldsymbol{\theta}) \\
&-\sum_{k=1}^{n}\sum_{i,j=1}^{n}\sum_{p,q=1}^{n}\mu_{ij}(\boldsymbol{\theta})\left[\mathbf{W}_d\mathbf{S}^{-2}(\lambda)\right]_{ij,pq}z_{pq}(\beta,\eta_p,\alpha_q)s_{ij,kl}(\lambda), \\
\partial_{\lambda_o\alpha_l}\ell_N(\boldsymbol{\theta}) =& \sum_{k=1}^{n}\sum_{i,j=1}^{n}\left[\mathbf{W}_o\mathbf{S}^{-2}(\lambda)\right]_{ij,kl}u_{ij}(\boldsymbol{\theta}) \\
&-\sum_{k=1}^{n}\sum_{i,j=1}^{n}\sum_{p,q=1}^{n}\mu_{ij}(\boldsymbol{\theta})\left[\mathbf{W}_o\mathbf{S}^{-2}(\lambda)\right]_{ij,pq}z_{pq}(\beta,\eta_p,\alpha_q)s_{ij,kl}(\lambda), \\
\partial_{\lambda_w\alpha_l}\ell_N(\boldsymbol{\theta}) =& \sum_{k=1}^{n}\sum_{i,j=1}^{n}\left[\mathbf{W}_w\mathbf{S}^{-2}(\lambda)\right]_{ij,kl}u_{ij}(\boldsymbol{\theta}) \\
&-\sum_{k=1}^{n}\sum_{i,j=1}^{n}\sum_{p,q=1}^{n}\mu_{ij}(\boldsymbol{\theta})\left[\mathbf{W}_w\mathbf{S}^{-2}(\lambda)\right]_{ij,pq}z_{pq}(\beta,\eta_p,\alpha_q)s_{ij,kl}(\lambda), \\
\partial_{\beta\alpha_l}\ell_N(\boldsymbol{\theta}) =& -\sum_{k=1}^{n}\sum_{i,j=1}^{n}\sum_{p,q=1}^{n}\mu_{ij}(\boldsymbol{\theta})s_{ij,kl}(\lambda)s_{ij,pq}(\lambda)x_{pq},
\end{aligned}$$

$$\begin{aligned}
\partial_{\lambda_d\eta_k}\ell_N(\boldsymbol{\theta}) =& \sum_{l=1}^{n}\sum_{i,j=1}^{n}\left[\mathbf{W}_d\mathbf{S}^{-2}(\lambda)\right]_{ij,kl}u_{ij}(\boldsymbol{\theta}) \\
&-\sum_{l=1}^{n}\sum_{i,j=1}^{n}\sum_{p,q=1}^{n}\mu_{ij}(\boldsymbol{\theta})\left[\mathbf{W}_d\mathbf{S}^{-2}(\lambda)\right]_{ij,pq}z_{pq}(\beta,\eta_p,\alpha_q)s_{ij,kl}(\lambda), \\
\partial_{\lambda_o\eta_k}\ell_N(\boldsymbol{\theta}) =& \sum_{l=1}^{n}\sum_{i,j=1}^{n}\left[\mathbf{W}_o\mathbf{S}^{-2}(\lambda)\right]_{ij,kl}u_{ij}(\boldsymbol{\theta}) \\
&-\sum_{l=1}^{n}\sum_{i,j=1}^{n}\sum_{p,q=1}^{n}\mu_{ij}(\boldsymbol{\theta})\left[\mathbf{W}_o\mathbf{S}^{-2}(\lambda)\right]_{ij,pq}z_{pq}(\beta,\eta_p,\alpha_q)s_{ij,kl}(\lambda), \\
\partial_{\lambda_w\eta_k}\ell_N(\boldsymbol{\theta}) =& \sum_{l=1}^{n}\sum_{i,j=1}^{n}\left[\mathbf{W}_w\mathbf{S}^{-2}(\lambda)\right]_{ij,kl}u_{ij}(\boldsymbol{\theta}) \\
&-\sum_{l=1}^{n}\sum_{i,j=1}^{n}\sum_{p,q=1}^{n}\mu_{ij}(\boldsymbol{\theta})\left[\mathbf{W}_w\mathbf{S}^{-2}(\lambda)\right]_{ij,pq}z_{pq}(\beta,\eta_p,\alpha_q)s_{ij,kl}(\lambda), \text{ and} \\
\partial_{\beta\eta_k}\ell_N(\boldsymbol{\theta}) =& -\sum_{l=1}^{n}\sum_{i,j=1}^{n}\sum_{p,q=1}^{n}\mu_{ij}(\boldsymbol{\theta})s_{ij,kl}(\lambda)s_{ij,pq}(\lambda)x_{pq}.
\end{aligned}$$

Third, consider the last block, $\partial_{\boldsymbol{\phi\phi}}\ell_N(\boldsymbol{\theta})$:

$$
\begin{aligned}
\partial_{\alpha_l\alpha_l}\ell_N(\boldsymbol{\theta}) &= -\sum_{i,j=1}^{n}\left(\sum_{k=1}^{n}\sum_{p=1}^{n} s_{ij,kl}(\lambda)s_{ij,pl}(\lambda)\right)\mu_{ij}(\boldsymbol{\theta}) - 1,\\
\partial_{\alpha_l\alpha_s}\ell_N(\boldsymbol{\theta}) &= -\sum_{i,j=1}^{n}\left(\sum_{k=1}^{n}\sum_{p=1}^{n} s_{ij,kl}(\lambda)s_{ij,ps}(\lambda)\right)\mu_{ij}(\boldsymbol{\theta}) - 1 \text{ if } l \neq s,\\
\partial_{\alpha_l\eta_k}\ell_N(\boldsymbol{\theta}) &= -\sum_{i,j=1}^{n}\left(\sum_{k=1}^{n}\sum_{q=1}^{n} s_{ij,kl}(\lambda)s_{ij,kq}(\lambda)\right)\mu_{ij}(\boldsymbol{\theta}) + 1,\\
\partial_{\eta_k\eta_k}\ell_N(\boldsymbol{\theta}) &= -\sum_{i,j=1}^{n}\left(\sum_{l=1}^{n}\sum_{q=1}^{n} s_{ij,kl}(\lambda)s_{ij,kq}(\lambda)\right)\mu_{ij}(\boldsymbol{\theta}) - 1, \text{ and}\\
\partial_{\eta_k\eta_t}\ell_N(\boldsymbol{\theta}) &= -\sum_{i,j=1}^{n}\left(\sum_{l=1}^{n}\sum_{q=1}^{n} s_{ij,kl}(\lambda)s_{ij,tq}(\lambda)\right)\mu_{ij}(\boldsymbol{\theta}) - 1 \text{ if } k \neq t.
\end{aligned}
$$

To have a vector/matrix notation, we define

$$
\begin{aligned}
&\boldsymbol{\mu}(\boldsymbol{\theta}) = \left(\exp\left(\tilde{\mu}_{11}(\boldsymbol{\theta})\right),\cdots,\exp\left(\tilde{\mu}_{n1}(\boldsymbol{\theta})\right),\cdots,\exp\left(\tilde{\mu}_{1n}(\boldsymbol{\theta})\right),\cdots,\exp\left(\tilde{\mu}_{nn}(\boldsymbol{\theta})\right)\right), \text{ and}\\
&\tilde{\boldsymbol{\mu}}(\boldsymbol{\theta}) = \left(\tilde{\mu}_{11}(\boldsymbol{\theta}),\cdots,\tilde{\mu}_{n1}(\boldsymbol{\theta}),\cdots,\tilde{\mu}_{1n}(\boldsymbol{\theta}),\cdots,\tilde{\mu}_{nn}(\boldsymbol{\theta})\right)
\end{aligned}
$$

Indeed, $\tilde{\boldsymbol{\mu}}(\boldsymbol{\theta}) = \mathbf{S}^{-1}(\lambda)\left(\mathbf{X}\beta + \boldsymbol{\alpha}\otimes l_n + l_n\otimes\boldsymbol{\eta}\right) = \mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta})$. First,

$$
\partial_{\theta\theta}\ell_N(\boldsymbol{\theta}) = -\left(\mathbf{S}^{-1}\mathbf{G}(\boldsymbol{\theta})\right)'\operatorname{Diag}\left(\boldsymbol{\mu}(\boldsymbol{\theta})\right)\left(\mathbf{S}^{-1}\mathbf{G}(\boldsymbol{\theta})\right) + \mathbf{H}^{\theta\theta}(\boldsymbol{\theta}),
$$

where $\mathbf{G}(\boldsymbol{\theta}) = \left[\mathbf{W}_d\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta}), \mathbf{W}_o\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta}), \mathbf{W}_w\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta}), \mathbf{X}\right]$, and $\mathbf{H}^{\theta\theta}(\boldsymbol{\theta}) = \begin{bmatrix}\mathbf{H}^{\lambda\lambda}(\boldsymbol{\theta}) & \mathbf{H}^{\beta\lambda\prime}(\boldsymbol{\theta})\\ \mathbf{H}^{\beta\lambda}(\boldsymbol{\theta}) & \mathbf{H}^{\beta\beta}(\boldsymbol{\theta})\end{bmatrix}$ with

$$
\mathbf{H}^{\lambda\lambda}(\boldsymbol{\theta}) = \begin{bmatrix}
\left(2\mathbf{W}_d^2\mathbf{S}^{-3}(\lambda)\mathbf{Z}(\boldsymbol{\theta})\right)'\mathbf{u}(\boldsymbol{\theta}) & \left(2\mathbf{W}_d\mathbf{W}_o\mathbf{S}^{-3}(\lambda)\mathbf{Z}(\boldsymbol{\theta})\right)'\mathbf{u}(\boldsymbol{\theta}) & \left(2\mathbf{W}_d\mathbf{W}_w\mathbf{S}^{-3}(\lambda)\mathbf{Z}(\boldsymbol{\theta})\right)'\mathbf{u}(\boldsymbol{\theta})\\
* & \left(2\mathbf{W}_o^2\mathbf{S}^{-3}(\lambda)\mathbf{Z}(\boldsymbol{\theta})\right)'\mathbf{u}(\boldsymbol{\theta}) & \left(2\mathbf{W}_o\mathbf{W}_w\mathbf{S}^{-3}(\lambda)\mathbf{Z}(\boldsymbol{\theta})\right)'\mathbf{u}(\boldsymbol{\theta})\\
* & * & \left(2\mathbf{W}_w^2\mathbf{S}^{-3}(\lambda)\mathbf{Z}(\boldsymbol{\theta})\right)'\mathbf{u}(\boldsymbol{\theta})
\end{bmatrix},
$$

$\mathbf{H}^{\beta\lambda}(\boldsymbol{\theta}) = \left[\left(\mathbf{W}_d\mathbf{S}^{-2}(\lambda)\mathbf{X}\right)'\mathbf{u}(\boldsymbol{\theta}) \quad \left(\mathbf{W}_o\mathbf{S}^{-2}(\lambda)\mathbf{X}\right)'\mathbf{u}(\boldsymbol{\theta}) \quad \left(\mathbf{W}_w\mathbf{S}^{-2}(\lambda)\mathbf{X}\right)'\mathbf{u}(\boldsymbol{\theta})\right]$, and $\mathbf{H}^{\beta\beta}(\boldsymbol{\theta}) = \mathbf{0}_{K\times K}$.

Second,

$$
\partial_{\boldsymbol{\phi}\theta}\ell_N(\boldsymbol{\theta}) = -\left(\mathbf{S}^{-1}(\lambda)\mathbf{D}\right)'\operatorname{Diag}(\boldsymbol{\mu}(\boldsymbol{\theta}))\left(\mathbf{S}^{-1}\mathbf{G}(\boldsymbol{\theta})\right) + \mathbf{H}^{\phi\theta}(\boldsymbol{\theta}),
$$

where

$$
\mathbf{H}^{\phi\theta}(\boldsymbol{\theta}) = \left[\left(\mathbf{W}_d\mathbf{S}^{-2}(\lambda)\mathbf{D}\right)'\mathbf{u}(\boldsymbol{\theta}) \quad \left(\mathbf{W}_o\mathbf{S}^{-2}(\lambda)\mathbf{D}\right)'\mathbf{u}(\boldsymbol{\theta}) \quad \left(\mathbf{W}_w\mathbf{S}^{-2}(\lambda)\mathbf{D}\right)'\mathbf{u}(\boldsymbol{\theta}) \quad \mathbf{0}_{2n\times K}\right].
$$

Last, note that

$$
\partial_{\boldsymbol{\phi\phi}}\ell_N(\boldsymbol{\theta}) = -\left(\mathbf{S}^{-1}(\lambda)\mathbf{D}\right)'\operatorname{Diag}(\boldsymbol{\mu}(\boldsymbol{\theta}))\left(\mathbf{S}^{-1}(\lambda)\mathbf{D}\right) + \mathbf{H}^{\boldsymbol{\phi\phi}},
$$

where

$$
\mathbf{H}^{\boldsymbol{\phi\phi}} = -\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}\otimes l_n l_n'.
$$

Note that $\mathbf{H}^{\phi\phi}$ does not depend on specific parameter values.

In sum,

$$\partial_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_N(\boldsymbol{\theta}) = \mathbf{H}^A(\boldsymbol{\theta}) + \mathbf{H}^B(\boldsymbol{\theta}), \tag{39}$$

where

$$\mathbf{H}^A(\boldsymbol{\theta}) = -\begin{bmatrix} \left(\mathbf{S}^{-1}(\lambda)\mathbf{G}(\boldsymbol{\theta})\right)' \\ \left(\mathbf{S}^{-1}(\lambda)\mathbf{D}\right)' \end{bmatrix} \mathrm{Diag}(\boldsymbol{\mu}(\boldsymbol{\theta})) \begin{bmatrix} \mathbf{S}^{-1}(\lambda)\mathbf{G}(\boldsymbol{\theta}) & \mathbf{S}^{-1}(\lambda)\mathbf{D} \end{bmatrix}$$

and

$$\mathbf{H}^B(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{H}^{\theta\theta}(\boldsymbol{\theta}) & \mathbf{H}^{\phi\theta\prime}(\boldsymbol{\theta}) \\ \mathbf{H}^{\phi\theta}(\boldsymbol{\theta}) & \mathbf{H}^{\phi\phi} \end{bmatrix}.$$

2.2 NED properties

Establishing consistency and asymptotic normality relies on the laws of large numbers (LLN) and the central limit theorem (CLT). Jenish and Prucha (2009) examine the pointwise LLN, uniform LLN, and CLT for spatial mixing processes. Jenish and Prucha (2012) extend the notion of near-epoch dependent (NED) processes in the time series context to spatial random fields.

This paper focuses on revealing the main statistics' NED properties on the α-mixing random fields. For this, we reproduce the following regularity assumptions for reader's convenience.

Assumption 2.1. Each cross-section unit (region) $i \in \{1, \cdots, n\}$ is located in a d-dimensional space $\mathcal{D}_n \subset \mathcal{D}$, where $\mathcal{D}$ denotes a set (irregular lattice) of all potential locations in $\mathbb{R}^d$. We assume $\lim_{n\to\infty} \#(\mathcal{D}_n) = \infty$ and $\min_{i\neq j} d(l(i), l(j)) \geq 1$, where $\#(\mathcal{D}_n)$ denotes the cardinality of $\mathcal{D}_n$, $l : i \mapsto l(i) \in \mathcal{D}$ stands for an injective location function, and $d(l(i), l(j))$ is a distance between i and j.

Assumption 2.2. Let Λ be the parameter space of λ. For each $\lambda \in \Lambda$, we define

$$\mathbf{A}(\lambda) = \lambda_d(I_n \otimes W) + \lambda_o(W \otimes I_n) + \lambda_w(W \otimes W)$$

and $\mathbf{A} = \mathbf{A}(\lambda^0)$.

(i) When the SAR weights are used, we assume that $\sup_n \sup_{\lambda\in\Lambda} \|\mathbf{A}(\lambda)\|_\infty < 1$.

(ii) For any $\lambda \in \Lambda$,

- If SAR weights are chosen, $\mathbf{S}^{-1}(\lambda)$ is uniformly bounded in the row sum norm.
- If SMA weights are selected, $I_N + \mathbf{A}(\lambda)$ is uniformly bounded in the row sum norm.
- If MESS weights are chosen, $e^{\mathbf{A}(\lambda)}$ is uniformly bounded in the row sum norm.

Assumption 2.3. We assume that the parameter space Θ of θ is compact, and $\theta^0 \in \Theta$.

Assumption 2.4. Let $C > 0$ be a generic finite constant.

(i) Assume $\{x_{ij}\}$, $\{\eta_i^0\}$, and $\{\alpha_j^0\}$ are random fields satisfying $\max_k \sup_{i,j,n} |x_{ij,k}| < C$, $\sup_{i,n} |\eta_i^0| < C$, and $\sup_{j,n} |\alpha_j^0| < C$.

(ii) Assume $\{\xi_{ij}\}$ is a random field satisfying $\sup_{i,j,n} \mathbb{E}|\xi_{ij}|^{4+c} < C$ for some $c > 0$.

(iii) Assume $\mathbb{E}\left(\xi_{ij}|\mathbf{z}\right) = 1$ for all $i, j = 1, \cdots, n$.

Lemma 2.1. For each (i,j), we define the additive error, $u_{ij} = \mu_{ij}(\xi_{ij}-1)$, to have $u_{ij} = y_{ij} - \mu_{ij}$. Under Assumption 2.4, we obtain $\mathbb{E}(u_{ij}|\mathbf{z}) = 0$ and $\sup_{i,j,n} \mathbb{E}|u_{ij}|^{4+c} < C$.

Assumption 2.1 illustrates the topological specification for the cross-section units' locations. The minimum distance assumption prevents cross-section units from having clustered locations, which possibly generate extreme spatial influences. Hence, it is more natural for regional analyses. Recall that each OD flow, y_{ij}, is generated by two locations, i and j. Hence, a pair (i,j) for y_{ij} is located in the product space $\mathcal{D} \times \mathcal{D} \subset \mathbb{R}^{2d}$. For notational simplicity, we denote (i,j) as ij in this supplement document. In consequence, the location of a pair can be defined by $l^p : ij \mapsto l^p(ij) \in \mathcal{D} \times \mathcal{D} \subset \mathbb{R}^{2d}$. As Jeong and Lee (2024), we employ the maximum metric to evaluate the distance between two pairs, ij and kl:

$$d^p\left(l^p(ij), l^p(kl)\right) = \max\{d(l(i), l(k)), d(l(j), l(l))\}. \tag{40}$$

For notational simplicity, $d^p_{ij,kl} = d^p\left(l^p(ij), l^p(kl)\right)$ and $d_{ik} = d(l(i), l(k))$. The distance between pairs in (40) is measured by a larger distance between the distance between origins and that between destinations. Using this device, we want to control $\mathbb{C}(y_{ij}, y_{kl})$: $\mathbb{C}(y_{ij}, y_{kl}) \to 0$ as $d^p_{ij,kl} \to \infty$. As an illustrative example, consider the covariance between y_{ij} and y_{kj} with $i \neq j$, which means the two flows share the same origin but different destinations. Even for their common origin j, this setting implies $\mathbb{C}(y_{ij}, y_{kj}) \to 0$ as $d_{ik} \to \infty$. Assumption 2.4 describes the properties of the components in $\{x_{ij}\}$, $\{\eta_i^0\}$ and $\{\alpha_j^0\}$, and the errors $\{\xi_{ij}\}$ for a simple asymptotic analysis.

Lemma 2.1 illustrates that the key properties of $\{u_{ij}\}$ are implied by those of $\{\xi_{ij}\}$.

Proof of Lemma 2.1. First, observe $\mathbb{E}\left(\epsilon_{ij}|\mathbf{z}\right) = \mathbb{E}\left(\mu_{ij}(\xi_{ij}-1)|\mathbf{z}\right) = \mu_{ij}\left(\mathbb{E}(\xi_{ij}|\mathbf{z}) - 1\right) = 0$.

Second, by Assumptions 2.2 (ii), 2.3 and 2.4 (i),

$$\tilde{\mu}_{ij} = \sum_{k,l=1}^{n} s_{ij,kl}(x'_{kl}\beta^0 + \alpha_l^0 + \eta_k^0) \leq \|\mathbf{S}^{-1}\|_\infty \cdot \sup_{i,j,n} |x_{ij}\beta^0 + \alpha_j^0 + \eta_i^0| < \infty.$$

This implies $\mu_{ij} = \exp\left(\tilde{\mu}_{ij}\right)$ is bounded. Then, $|\mu_{ij}| \leq C$. It implies $|\mu_{ij}(\xi_{ij}-1)|^p \leq C^p \cdot |\xi_{ij}-1|^p$ a.s. for any $p \geq 1$. Suppose $\mathbb{E}|\xi_{ij}|^p < \infty$ for an arbitrary $p \geq 1$. We need to show $\mathbb{E}|\xi_{ij}-1|^p < \infty$. Since $|\xi_{ij}-1| \leq |\xi_{ij}| + 1$ and the c_r-inequality (i.e.,$(a+b)^p \leq 2^{p-1}(a^p + b^p)$), we have

$$|\xi_{ij}-1|^p \leq 2^{p-1}\left(|\xi_{ij}|^p + 1\right).$$

It implies $\mathbb{E}|\xi_{ij}-1|^p \leq 2^{p-1}\left(\mathbb{E}|\xi_{ij}|^p + 1\right) < \infty$ by monotonicity of $\mathbb{E}(\cdot)$. Consequently, $\mathbb{E}|u_{ij}|^p \leq C^p \cdot 2^{p-1}\left(\mathbb{E}|\xi_{ij}|^p + 1\right) < \infty$ for any $p \geq 1$. This completes the proof.$\blacksquare$

Assumption 2.5 (Covariance structure)**.** Let $\mathbf{u} = (u_{11}, \cdots, u_{n1}, \cdots, u_{1n}, \cdots, u_{nn})'$ be an $N \times 1$ vector of additive errors.

(i) We assume

$$\mathbf{u} = \mathbf{B}\mathbf{H}\boldsymbol{\epsilon}, \tag{41}$$

where $\mathbf{B}$ denotes some $N \times N$ matrix, $\mathbf{H} = \operatorname{diag}\left(\sigma^*_{11}, \cdots, \sigma^*_{n1}, \cdots, \sigma^*_{1n}, \cdots, \sigma^*_{nn}\right)$, and $\boldsymbol{\epsilon} = (\epsilon_{11}, \cdots, \epsilon_{n1}, \cdots, \epsilon_{1n}, \cdots, \epsilon_{nn})$ is an $N \times 1$ vector of innovations.

(ii) For each ij, we assume ϵ_{ij} is indenpendently drawn from a distribution with $\mathbb{E}(\epsilon_{ij}) = 0$ and $\mathbb{V}(\epsilon_{ij}) = 1$ with $\sup_{i,j,n} \mathbb{E}|\epsilon_{ij}|^{4+c} < \infty$.

(iii) Note that $\epsilon^*_{ij} = \sigma^*_{ij}\epsilon_{ij}$ denotes an unnormalized error. We assume $0 < \inf_{i,j,n} \sigma^*_{ij} \leq \sup_{i,j,n} \sigma^*_{ij} < \infty$.

Assumption 2.6 (Presence of dominant units)**.** Let $b_{kl,ij}$ be the $((l-1)n+k,(j-1)n+i)$-element of $\mathbf{B}$.

(i) For $i=1,\cdots,n$, denote i's outdegree by d_i such that

$$d_i = c_i \cdot n^{\delta_i},$$

where c_i is a positive constant and $\delta_i \in [0,1)$.

(ii) We assume $d_{ij} = \sum_{k=1}^{n}\sum_{l=1}^{n} b_{kl,ij} = 1 + d_i + d_j + d_i d_j$. Then, $d_{ij} = O\left(n^{\delta_i+\delta_j}\right)$.

The example below describes a possible example satisfying Assumption 2.5. Suppose ξ^*_{ij} is independently drawn from a distribution with $\mathbb{E}\left(\xi^*_{ij}\right)=1$ and $\mathbb{V}\left(\xi^*_{ij}\right)=\sigma^2_{\xi^*}>0$.

If the original distribution of $\{\xi^*_{ij}\}$ is the log-normal distribution (i.e., $\xi^*_{ij}=\exp\left(\mu_{\xi^{**}}+\sigma_{\xi^{**}}\cdot Z_{ij}\right)$, where Z_{ij} is a standard normal variable), we have

$$\begin{aligned}
\mathbb{E}\left(\xi^*_{ij}\right) &= \exp\left(\mu_{\xi^{**}}+\frac{1}{2}\sigma^2_{\xi^{**}}\right) = \exp(0)=1, \text{ and}\\
\mathbb{V}\left(\xi^*_{ij}\right) &= \left(\exp(\sigma^2_{\xi^{**}})-1\right)\cdot\exp\left(2\mu_{\xi^{**}}+\sigma^2_{\xi^{**}}\right)\right) = \exp(\sigma^2_{\xi^{**}})-1
\end{aligned}$$

since $\mu_{\xi^{**}} = -\frac{1}{2}\sigma^2_{\xi^{**}}$ is required to satisfy $\mathbb{E}\left(\xi^*_{ij}\right)=1$.

If the original distribution of $\{\xi^*_{ij}\}$ is the gamma distribution (i.e., $\xi^*_{ij}\sim \mathit{Gamma}\left(\theta_{\text{shape}},\theta_{\text{rate}}\right)$), we then have

$$\begin{aligned}
\mathbb{E}\left(\xi^*_{ij}\right) &= \frac{\theta_{\text{shape}}}{\theta_{\text{rate}}} = 1, \text{ and}\\
\mathbb{V}\left(\xi^*_{ij}\right) &= \frac{\theta_{\text{shape}}}{\theta^2_{\text{rate}}} = \frac{1}{\theta_{\text{rate}}}
\end{aligned}$$

since $\theta_{\text{shape}}=\theta_{\text{rate}}$ is needed to have $\mathbb{E}\left(\xi^*_{ij}\right)=1$.

From the i.i.d. multiplicative errors $\{\xi^*_{ij}\}$, we define

$$\epsilon^*_{ij} = \mu_{ij}(\xi^*_{ij}-1)$$

for $i,j=1,\cdots,n$. Then, $\mathbb{E}\left(\epsilon^*_{ij}\right)=\mathbb{E}\left(\underbrace{\mathbb{E}(\epsilon^*_{ij}|\mathbf{z})}_{=0}\right)=0$ and

$$\begin{aligned}
\mathbb{V}(\epsilon^*_{ij}) &= \mathbb{E}\left(\mathbb{V}(\epsilon^*_{ij}|\mathbf{z})\right)+\mathbb{V}\left(\mathbb{E}(\epsilon^*_{ij}|\mathbf{z})\right) \quad (\because \text{by the law of total variance})\\
&= \mathbb{E}\left(\mathbb{V}(\epsilon^*_{ij}|\mathbf{z})\right) \quad (\because \mathbb{E}(\epsilon^*_{ij}|\mathbf{z})=0)\\
&= \bar{\mu}_{ij,2}\cdot\sigma^2_{\xi^*},
\end{aligned}$$

where $\bar{\mu}_{ij,2}=\mathbb{E}(\mu^2_{ij})$. Further, $\{\epsilon^*_{ij}\}$ is mutually independent. Consequently, we can set

$$\epsilon^*_{ij} = \sigma^*_{ij}\cdot\epsilon_{ij} = \sigma_{\xi^*}\sqrt{\bar{\mu}_{ij,2}}\cdot\epsilon_{ij},$$

that is, $\sigma^*_{ij}=\sigma_{\xi^*}\sqrt{\bar{\mu}_{ij,2}}$.

As an example, we consider an SMA process with $W = (w_{ij})$ for $\mathbf{u}$:

$$\mathbf{B} = I_N + (I_n \otimes W) + (W \otimes I_n) + (W \otimes W).$$

Under this specification,

$$u_{ij} = \sigma^*_{ij}\epsilon_{ij} + \sum_{k=1}^{n} w_{ik}\sigma^*_{kj}\epsilon_{kj} + \sum_{l=1}^{n} w_{jl}\sigma^*_{il}\epsilon_{il} + \sum_{k,l=1}^{n} w_{ik}w_{jl}\sigma^*_{kl}\epsilon_{kl}.$$

For analytical simplicity, we suppose $w_{ij} \geq 0$ for all ij. Then, we have

$$\begin{aligned}
\sum_{i,j=1}^{n}\sum_{k,l=1}^{n} \mathbb{E}(u_{ij}u_{kl}) &= \sum_{i,j=1}^{n}\sum_{k,l=1}^{n} \left(e'_{n,j} \otimes e'_{n,i}\right) \mathbf{B}\mathbf{H}^2\mathbf{B}' \left(e_{n,l} \otimes e_{n,k}\right) \\
&= \sum_{i,j=1}^{n}\sum_{k,l=1}^{n}\sum_{p,q=1}^{n} \sigma^{*2}_{pq} b_{ij,pq} b_{kl,pq} \\
&\leq \left(\sup_{i,j,n} \sigma^{*2}_{ij}\right) \cdot \sum_{p,q=1}^{n} \left(\sum_{i,j=1}^{n} b_{ij,pq}\right) \left(\sum_{k,l=1}^{n} b_{kl,pq}\right) \\
&= \left(\sup_{i,j,n} \sigma^{*2}_{ij}\right) \cdot \sum_{p,q=1}^{n} \left(\underbrace{\sum_{i,j=1}^{n} b_{ij,pq}}_{=d_{pq}}\right)^2 \\
&= \left(\sup_{i,j,n} \sigma^{*2}_{ij}\right) \cdot \sum_{i,j=1}^{n} d_{ij}^2.
\end{aligned}$$

For each ij, observe

$$\begin{aligned}
d_{ij} &= \sum_{k,l=1}^{n} b_{kl,ij} \\
&= \sum_{k,l=1}^{n} \left(e'_{n,l} \otimes e'_{n,k}\right) \left(I_N + (I_n \otimes W) + (W \otimes I_n) + (W \otimes W)\right) \left(e_{n,j} \otimes e_{n,i}\right) \\
&= \sum_{k,l=1}^{n} \left(\mathbf{1}(i=k, j=l) + \mathbf{1}(j=l) \cdot w_{kl} + w_{lj} \cdot \mathbf{1}(i=k) + w_{ki} \cdot w_{lj}\right) \\
&= 1 + \sum_{k=1}^{n} w_{ki} + \sum_{l=1}^{n} w_{lj} + \sum_{k=1}^{n} w_{ki} \cdot \sum_{l=1}^{n} w_{lj} \\
&= 1 + d_i + d_j + d_i \cdot d_j.
\end{aligned}$$

[Upper bound? Lower bound?] [Additional properties of u_{ij}]

3. Let $\xi_{ij} = 1 + \frac{u_{ij}}{\mu_{ij}}$ for $i, j = 1, \cdots, n$. Then, we can observe $\mathbb{E}\left(\xi_{ij}|\mathbf{z}\right) = 1$ since $\mathbb{E}\left(\xi_{ij}|\mathbf{z}\right) = 1 + \frac{1}{\mu_{ij}}\mathbb{E}(u_{ij}|\mathbf{z}) = 1$. The lemma below shows the NED properties of $\{y_{ij}\}$. [Description] [baseline stochastic space]

Lemma 2.2. Assume Assumptions 2.1, 2.2, and 2.4 hold. [To be updated]

(i) Then, we have uniform L_p-boundedness of $\{y_{ij}\}$. That is, $\sup_{n,i,j} \|y_{ij}\|_{L_{4+c}} < \infty$.

(ii)

Proof of Lemma 2.2 (i) We need to show $\sup_{i,j,n} \|\mu_{ij} \cdot \xi_{ij}\|_{L_{4+c}} < \infty$.

2.3 Asymptotic distribution

Variance structure

This section provides details on deriving the asymptotic distribution of the PPMLE.

Linear model. Before introducing the details, an intuition of deriving the variance structure can be delivered through a linear model:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{D}\boldsymbol{\phi} + \mathbf{u},$$

where

- $\mathbf{y} = (y_{11}, y_{21}, \cdots, y_{n1}, \cdots, y_{1n}, y_{2n}, \cdots, y_{nn})'$,
- $\mathbf{X} = (x_{ij,k})$ is an $N \times K$ matrix of regressors,
- $\mathbf{D} = [\mathbf{I}_n \otimes l_n, l_n \otimes \mathbf{I}_n]$ is an $N \times 2n$ matrix of dummy variables, and
- $\mathbf{u} = (u_{11}, u_{21}, \cdots, u_{n1}, \cdots, u_{1n}, u_{2n}, \cdots, u_{nn})'$ is an N-dimensional vector of disturbances.

Then, the log-likelihood function is

$$\ell_N(\beta, \boldsymbol{\phi}) = -\frac{1}{2}\left(\mathbf{y} - \mathbf{X}\beta - \mathbf{D}\boldsymbol{\phi}\right)'\left(\mathbf{y} - \mathbf{X}\beta - \mathbf{D}\boldsymbol{\phi}\right) - \frac{1}{2}\left(v'\boldsymbol{\phi}\right)^2,$$

where $v = (l_n', -l_n')'$. The first-order conditions are

$$\begin{aligned}
[\beta] :& \mathbf{X}'\left(\mathbf{y} - \mathbf{X}\beta - \mathbf{D}\boldsymbol{\phi}\right) = \mathbf{0}, \\
[\boldsymbol{\phi}] :& \mathbf{D}'\left(\mathbf{y} - \mathbf{X}\beta - \mathbf{D}\boldsymbol{\phi}\right) - v\underbrace{v'\boldsymbol{\phi}}_{=0} = \mathbf{0}.
\end{aligned}$$

Let $\boldsymbol{\theta} = \left(\beta', \boldsymbol{\phi}'\right)'$ for notational convenience. The second-order derivatives are

$$\partial_{\boldsymbol{\theta\theta}}\ell_N(\beta, \boldsymbol{\phi}) = -\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{D} \\ \mathbf{D}'\mathbf{X} & \mathbf{D}'\mathbf{D} \end{bmatrix} - \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \otimes l_n l_n' \end{bmatrix}.$$

Note that $-\mathbf{D}'\mathbf{D} - \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \otimes l_n l_n' = -\begin{bmatrix} n\mathbf{I}_n & l_n l_n' \\ l_n l_n' & n\mathbf{I}_n \end{bmatrix} - \begin{bmatrix} l_n l_n' & -l_n l_n' \\ -l_n l_n' & l_n l_n' \end{bmatrix} = -\begin{bmatrix} n\mathbf{I}_n + l_n l_n' & \mathbf{0} \\ \mathbf{0} & n\mathbf{I}_n + l_n l_n' \end{bmatrix}$. For additional analysis, $\widetilde{\mathbf{D}'\mathbf{D}} := \begin{bmatrix} n\mathbf{I}_n + l_n l_n' & \mathbf{0} \\ \mathbf{0} & n\mathbf{I}_n + l_n l_n' \end{bmatrix}$. Since $\operatorname{rank}(\mathbf{D}'\mathbf{D}) = 2n - 1$, the presence of the penalty term leads to having full rank for the $\mathbf{D}'\mathbf{D}$ part.

Consequently, the quadratic expansion of $\ell_N(\beta, \boldsymbol{\phi})$ is

$$\begin{aligned}
\mathbf{0} = \partial_{\boldsymbol{\theta}} \ell_N(\hat{\boldsymbol{\theta}}) &= \partial_{\boldsymbol{\theta}} \ell_N(\boldsymbol{\theta}^0) + \partial_{\boldsymbol{\theta}\boldsymbol{\theta}} \ell_N(\boldsymbol{\theta}^0) \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right) \\
\Leftrightarrow \begin{pmatrix} \hat{\beta} - \beta^0 \\ \hat{\boldsymbol{\phi}} - \boldsymbol{\phi}^0 \end{pmatrix} &= \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{D} \\ \mathbf{D}'\mathbf{X} & \widetilde{\mathbf{D}'\mathbf{D}} \end{bmatrix}^{-1} \cdot \begin{pmatrix} \mathbf{X}'\mathbf{u} \\ \mathbf{D}'\mathbf{u} \end{pmatrix}
\end{aligned}$$

if $\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{D} \\ \mathbf{D}'\mathbf{X} & \widetilde{\mathbf{D}'\mathbf{D}} \end{bmatrix}$ is invertible. Note that the above expansion holds as equality since the second-order derivatives do not rely on $\boldsymbol{\theta}$. For convenience, we define

$$\mathbf{Q} \equiv \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{12}' & \mathbf{Q}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{D} \\ \mathbf{D}'\mathbf{X} & \widetilde{\mathbf{D}'\mathbf{D}} \end{bmatrix}^{-1}.$$

Note that

$$\begin{aligned}
\mathbf{Q}_{11} &= \left(\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{D}\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1} \mathbf{D}'\mathbf{X}\right)^{-1}, \\
\mathbf{Q}_{12} &= -\mathbf{Q}_{11}\mathbf{X}'\mathbf{D}\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}, \\
\mathbf{Q}_{21} = \mathbf{Q}_{12}' &= -\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1} \mathbf{D}'\mathbf{X}\mathbf{Q}_{11}, \text{ and} \\
\mathbf{Q}_{22} &= \left(\widetilde{\mathbf{D}'\mathbf{D}} - \mathbf{D}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{D}\right)^{-1}.
\end{aligned}$$

We are interested in obtaining the asymptotic distribution of $\sqrt{N}(\hat{\beta} - \beta^0)$. We define $\boldsymbol{\Gamma} = \begin{bmatrix} N\mathbf{I}_K & \mathbf{0} \\ \mathbf{0} & n\mathbf{I}_{2n} \end{bmatrix}$ to have

$$\boldsymbol{\Gamma}^{-\frac{1}{2}} \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{D} \\ \mathbf{D}'\mathbf{X} & \widetilde{\mathbf{D}'\mathbf{D}} \end{bmatrix} \boldsymbol{\Gamma}^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{N}\mathbf{X}'\mathbf{X} & \frac{1}{n\sqrt{n}}\mathbf{X}'\mathbf{D} \\ \frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{X} & \frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}} \end{bmatrix} = O_p(1)$$

and its positive definiteness for large n. Let $\boldsymbol{\Sigma}_N = \begin{bmatrix} \frac{1}{N}\mathbf{X}'\mathbf{X} & \frac{1}{n\sqrt{n}}\mathbf{X}'\mathbf{D} \\ \frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{X} & \frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}} \end{bmatrix}$. Observe that $\boldsymbol{\Sigma}_N$ does not depend on both β and $\boldsymbol{\phi}$.

In consequence, the approximated variance of $\begin{pmatrix} \sqrt{N}\left(\hat{\beta} - \beta^0\right) \\ \sqrt{n}\left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}^0\right) \end{pmatrix}$ is[8]

$$\begin{bmatrix} \frac{1}{N}\mathbf{X}'\mathbf{X} & \frac{1}{n\sqrt{n}}\mathbf{X}'\mathbf{D} \\ \frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{X} & \frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{N}\mathbf{X}'\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{X} & \frac{1}{n\sqrt{n}}\mathbf{X}'\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{D} \\ \frac{1}{n\sqrt{n}}\mathbf{D}'\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{X} & \frac{1}{n}\mathbf{D}'\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{D} \end{bmatrix} \begin{bmatrix} \frac{1}{N}\mathbf{X}'\mathbf{X} & \frac{1}{n\sqrt{n}}\mathbf{X}'\mathbf{D} \\ \frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{X} & \frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}} \end{bmatrix}^{-1},$$

[8]When the likelihood is correctly specified, by the likelihood equation, the approximated variance is $\begin{bmatrix} \frac{1}{N}\mathbf{X}'\mathbf{X} & \frac{1}{n\sqrt{n}}\mathbf{X}'\mathbf{D} \\ \frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{X} & \frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}} \end{bmatrix}^{-1}$.

since $\mathbf{\Sigma}_N = \begin{bmatrix} \frac{1}{N}\mathbf{X}'\mathbf{X} & \frac{1}{n\sqrt{n}}\mathbf{X}'\mathbf{D} \\ \frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{X} & \frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}} \end{bmatrix}$.

To evaluate the sandwich-form matrix above, we will employ the following lemma.

Lemma 2.3. (i) $\begin{bmatrix} X & Y \\ Y' & Z \end{bmatrix} \begin{bmatrix} P & Q \\ Q' & R \end{bmatrix} \begin{bmatrix} X & Y \\ Y' & Z \end{bmatrix} = \begin{bmatrix} XPX + YQ'X + XQY' + YRY' & XPY + YQ'Y + XQZ + YRZ \\ Y'PX + ZQ'X + Y'QY' + ZRY' & Y'PY + ZQ'Y + Y'QZ + ZRZ \end{bmatrix}$. Then, the main parameter part of the variance matrix is the first block, $XPX + YQ'X + XQY' + YRY'$.

(ii) If $\begin{bmatrix} X & Y \\ Y' & Z \end{bmatrix} = \begin{bmatrix} A & B \\ B' & C \end{bmatrix}^{-1}$, note that $X = \left(A - BC^{-1}B'\right)^{-1}$, $Y = -\left(A - BC^{-1}B'\right)^{-1} BC^{-1}$ and $Z = C^{-1} + C^{-1}B'\left(A - BC^{-1}B'\right)^{-1} BC^{-1}$ by the inverse of the partitioned matrix formula. Then, the main parameter part of the variance matrix is

$$
\begin{aligned}
&\left(A - BC^{-1}B'\right)^{-1} P \left(A - BC^{-1}B'\right)^{-1} \\
&- \left(A - BC^{-1}B'\right)^{-1} BC^{-1}Q' \left(A - BC^{-1}B'\right)^{-1} \\
&- \left(A - BC^{-1}B'\right)^{-1} QC^{-1}B' \left(A - BC^{-1}B'\right)^{-1} \\
&+ \left(A - BC^{-1}B'\right)^{-1} BC^{-1}RC^{-1}B' \left(A - BC^{-1}B'\right)^{-1} \\
&= \left(A - BC^{-1}B'\right)^{-1} \left(P - BC^{-1}Q' - QC^{-1}B' + BC^{-1}RC^{-1}B'\right) \left(A - BC^{-1}B'\right)^{-1},
\end{aligned}
$$

which implies a sandwich form.

If the likelihood is correctly specified, $\begin{bmatrix} X & Y \\ Y' & Z \end{bmatrix}^{-1} = \begin{bmatrix} P & Q \\ Q' & R \end{bmatrix}$. Then, the main parameter part of the variance matrix is simplified by $\left(A - BC^{-1}B'\right)^{-1}$ and can be consistently estimated. The form of $A - BC^{-1}B'$ is

$$
\begin{aligned}
\mathbf{\Sigma}_{\beta,N} &= -\frac{1}{N}\partial_{\beta\beta}\ell_N(\beta,\boldsymbol{\phi}) - \left(-\frac{1}{n\sqrt{n}}\partial_{\beta\boldsymbol{\phi}}\ell_N(\beta,\boldsymbol{\phi})\right)\left(-\frac{1}{n}\partial_{\boldsymbol{\phi}\boldsymbol{\phi}}\ell_N(\beta,\boldsymbol{\phi})\right)^{-1}\left(-\frac{1}{n\sqrt{n}}\partial_{\beta\boldsymbol{\phi}}\ell_N(\beta,\boldsymbol{\phi})\right)' \\
&= \frac{1}{N}\mathbf{X}'\mathbf{X} - \frac{1}{\sqrt{N}}\left(\frac{1}{\sqrt{N}}\mathbf{X}'\mathbf{D}\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\frac{1}{\sqrt{N}}\mathbf{D}'\mathbf{X}\right) \\
&= \frac{1}{N}\mathbf{X}'\mathbf{M}_{\mathbf{D}}\mathbf{X},
\end{aligned}
$$

where $\mathbf{M}_{\mathbf{D}} = I_N - \mathbf{D}\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\mathbf{D}'$.

On the other hand, the form of $P - BC^{-1}Q' - QC^{-1}B' + BC^{-1}RC^{-1}B'$ is

$$
\begin{aligned}
\mathbf{\Omega}_{\beta,N} =& \frac{1}{N}\mathbf{X}'\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{X} - \frac{1}{n\sqrt{n}}\mathbf{X}'\mathbf{D}\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\frac{1}{n\sqrt{n}}\mathbf{D}'\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{X} - \frac{1}{n\sqrt{n}}\mathbf{X}'\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{D}\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{X} \\
&+ \frac{1}{n\sqrt{n}}\mathbf{X}'\mathbf{D}\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\frac{1}{n}\mathbf{D}'\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{D}\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{X} \\
=& \frac{1}{N}\mathbf{X}'\left(\mathbb{E}(\mathbf{u}\mathbf{u}') - \mathbf{D}\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\mathbf{D}'\mathbb{E}(\mathbf{u}\mathbf{u}') - \mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{D}\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\mathbf{D}' + \mathbf{D}\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\mathbf{D}'\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{D}\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\mathbf{D}'\right)\mathbf{X} \\
=& \frac{1}{N}\mathbf{X}'(I_N - \mathbf{D}\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\mathbf{D}')\mathbb{E}(\mathbf{u}\mathbf{u}')(I_N - \mathbf{D}\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\mathbf{D}')\mathbf{X} \\
& \frac{1}{N}\mathbf{X}'\mathbf{M}_{\mathbf{D}}\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{M}_{\mathbf{D}}\mathbf{X}.
\end{aligned}
$$

The fixed-effect parameter part is $Y'PY + ZQ'Y + Y'QZ + ZRZ$:

$$\begin{aligned}
&C^{-1}B'\left(A-BC^{-1}B'\right)^{-1}P\left(A-BC^{-1}B'\right)^{-1}BC^{-1}\\
&-C^{-1}Q'\left(A-BC^{-1}B'\right)^{-1}BC^{-1}-C^{-1}B'\left(A-BC^{-1}B'\right)^{-1}BC^{-1}Q'\left(A-BC^{-1}B'\right)^{-1}BC^{-1}\\
&-C^{-1}B'\left(A-BC^{-1}B'\right)^{-1}QC^{-1}-C^{-1}B'\left(A-BC^{-1}B'\right)^{-1}QC^{-1}B'\left(A-BC^{-1}B'\right)^{-1}BC^{-1}\\
&+C^{-1}RC^{-1}+C^{-1}RC^{-1}B'\left(A-BC^{-1}B'\right)^{-1}BC^{-1}+C^{-1}B'\left(A-BC^{-1}B'\right)^{-1}BC^{-1}RC^{-1}\\
&+C^{-1}B'\left(A-BC^{-1}B'\right)^{-1}BC^{-1}RC^{-1}B'\left(A-BC^{-1}B'\right)^{-1}BC^{-1}\\
=&C^{-1}\begin{pmatrix} R-Q'\left(A-BC^{-1}B'\right)^{-1}B-B'\left(A-BC^{-1}B'\right)^{-1}Q\\ +RC^{-1}B'\left(A-BC^{-1}B'\right)^{-1}B+B'\left(A-BC^{-1}B'\right)^{-1}BC^{-1}R\end{pmatrix}C^{-1}\\
&+C^{-1}B'\left(A-BC^{-1}B'\right)^{-1}\left(P-BC^{-1}Q'-QC^{-1}B'+BC^{-1}RC^{-1}B'\right)\left(A-BC^{-1}B'\right)^{-1}BC^{-1}.
\end{aligned}$$

Hence, the approximated variance of $\boldsymbol{\phi}$ is:

$$\begin{aligned}
&\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\begin{pmatrix}\frac{1}{n}\mathbf{D}'\mathbb{E}(\mathbf{uu}')\mathbf{D}-\frac{1}{n\sqrt{n}}\mathbf{D}'\mathbb{E}(\mathbf{uu}')\mathbf{X}\boldsymbol{\Sigma}_{\beta,N}^{-1}\frac{1}{n\sqrt{n}}\mathbf{X}'\mathbf{D}-\frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{X}\boldsymbol{\Sigma}_{\beta,N}^{-1}\frac{1}{n\sqrt{n}}\mathbf{X}'\mathbb{E}(\mathbf{uu}')\mathbf{D}\\ +\frac{1}{n}\mathbf{D}'\mathbb{E}(\mathbf{uu}')\mathbf{D}\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{X}\boldsymbol{\Sigma}_{\beta,N}^{-1}\frac{1}{n\sqrt{n}}\mathbf{X}'\mathbf{D}\\ +\frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{X}\boldsymbol{\Sigma}_{\beta,N}^{-1}\frac{1}{n\sqrt{n}}\mathbf{X}'\mathbf{D}\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\frac{1}{n}\mathbf{D}'\mathbb{E}(\mathbf{uu}')\mathbf{D}\end{pmatrix}\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\\
&+\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{X}\boldsymbol{\Sigma}_{\beta,N}^{-1}\boldsymbol{\Omega}_{\beta,N}\boldsymbol{\Sigma}_{\beta,N}^{-1}\frac{1}{n\sqrt{n}}\mathbf{X}'\mathbf{D}\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\\
=&n\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\mathbf{D}'\left(\left(I_N-\mathbf{M}_{\mathbf{D}}\mathbf{X}(\mathbf{X}'\mathbf{M}_{\mathbf{D}}\mathbf{X})^{-1}\mathbf{X}'\right)'\mathbb{E}(\mathbf{uu}')\left(I_N-\mathbf{M}_{\mathbf{D}}\mathbf{X}(\mathbf{X}'\mathbf{M}_{\mathbf{D}}\mathbf{X})^{-1}\mathbf{X}'\right)\right)\mathbf{D}\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}
\end{aligned}$$

since $\boldsymbol{\Sigma}_{\beta,N}=\frac{1}{N}\mathbf{X}'\mathbf{M}_{\mathbf{D}}\mathbf{X}$ and $\boldsymbol{\Omega}_{\beta,N}=\frac{1}{N}\mathbf{X}'\mathbf{M}_{\mathbf{D}}\mathbb{E}(\mathbf{uu}')\mathbf{M}_{\mathbf{D}}\mathbf{X}$.

Our model. Two notable features of our model exist. Due to our model's nonlinearity, the second-order derivatives depend on θ and $\boldsymbol{\phi}$. Consequently, estimating $\boldsymbol{\Sigma}_N$ (the scaled expected negative Hessian) requires consistent estimates for θ^0 and $\boldsymbol{\phi}^0$. Assuming such consistent estimates are available, our main target is to estimate

$$\begin{aligned}
\boldsymbol{\Sigma}_N\equiv-\mathbb{E}\left(\boldsymbol{\Gamma}^{-\frac{1}{2}}\partial_{\boldsymbol{\theta\theta}}\ell_N(\boldsymbol{\theta}^0)\boldsymbol{\Gamma}^{-\frac{1}{2}}|\mathbf{z}\right)&=\begin{bmatrix}\frac{1}{\sqrt{N}}\mathbf{G}'\\ \frac{1}{\sqrt{n}}\mathbf{D}'\end{bmatrix}\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\begin{bmatrix}\frac{1}{\sqrt{N}}\mathbf{G} & \frac{1}{\sqrt{n}}\mathbf{D}\end{bmatrix}+\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \frac{1}{n}\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}\otimes l_nl_n'\end{bmatrix}\\
&=\begin{bmatrix}\frac{1}{N}\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{G} & \frac{1}{n\sqrt{n}}\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{D}\\ \frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{G} & \frac{1}{n}\mathbf{D}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{D}+\frac{1}{n}\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}\otimes l_nl_n'\end{bmatrix},
\end{aligned}$$

where

- $\mathbf{G}=\mathbf{G}(\boldsymbol{\theta}^0)=\left[\mathbf{W}_d\mathbf{S}^{-1}\mathbf{Z},\mathbf{W}_o\mathbf{S}^{-1}\mathbf{Z},\mathbf{W}_w\mathbf{S}^{-1}\mathbf{Z},\mathbf{X}\right]$,
- $\boldsymbol{\mu}=\boldsymbol{\mu}(\boldsymbol{\theta}^0)=(\exp(\tilde{\mu}_{11}),\cdots,\exp(\tilde{\mu}_{n1}),\cdots,\exp(\tilde{\mu}_{1n}),\cdots,\exp(\tilde{\mu}_{nn}))$,
- $\tilde{\boldsymbol{\mu}}=\tilde{\boldsymbol{\mu}}(\boldsymbol{\theta}^0)=(\tilde{\mu}_{11},\cdots,\tilde{\mu}_{n1},\cdots,\tilde{\mu}_{1n},\cdots,\tilde{\mu}_{nn})$.

Here, $\tilde{\boldsymbol{\mu}} = \mathbf{S}^{-1}\left(\mathbf{X}\beta^0 + \boldsymbol{\alpha}^0 \otimes l_n + l_n \otimes \boldsymbol{\eta}^0\right) = \mathbf{S}^{-1}\mathbf{Z}$. The relation above holds since

$$-\mathbb{E}\left(\boldsymbol{\Gamma}^{-\frac{1}{2}}\mathbf{H}_N^{\boldsymbol{\theta\theta}}(\boldsymbol{\theta}^0)\boldsymbol{\Gamma}^{-\frac{1}{2}}|\mathbf{z}\right) = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \frac{1}{n}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \otimes l_n l_n' \end{bmatrix}.$$

Let $\widetilde{\mathbf{D}'\mathbf{D}} := \mathbf{D}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{D} + \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \otimes l_n l_n'$. Hence, the form of $A - BC^{-1}B'$ is

$$\boldsymbol{\Sigma}_{\theta,N} = \frac{1}{N}\mathbf{G}'\mathbf{S}^{-1\prime}\left(\mathrm{Diag}(\boldsymbol{\mu}) - \mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{D}(\widetilde{\mathbf{D}'\mathbf{D}})^{-1}\mathbf{D}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\right)\mathbf{S}^{-1}\mathbf{G}.$$

Let $\mathbf{P_D} = \mathbf{S}^{-1}\mathbf{D}(\widetilde{\mathbf{D}'\mathbf{D}})^{-1}\mathbf{D}'\mathbf{S}^{-1\prime}$ be the projection-like matrix and $\mathbf{M_D} = I_N - \mathbf{P_D}\mathrm{Diag}(\boldsymbol{\mu})$. Then,

$$\boldsymbol{\Sigma}_{\theta,N} = \frac{1}{N}\mathbf{G}'\mathbf{S}^{-1\prime}\left(\mathrm{Diag}(\boldsymbol{\mu}) - \mathrm{Diag}(\boldsymbol{\mu})\mathbf{P_D}\mathrm{Diag}(\boldsymbol{\mu})\right)\mathbf{S}^{-1}\mathbf{G} = \frac{1}{N}\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{M_D}\mathbf{S}^{-1}\mathbf{G}.$$

Our next step is to obtain the form of $P - BC^{-1}Q' - QC^{-1}B' + BC^{-1}RC^{-1}B'$. For this, note that

$$\mathbb{V}\left(\left.\begin{pmatrix} \frac{1}{\sqrt{N}}\left(\mathbf{S}^{-1}\mathbf{G}\right)'\mathbf{u} \\ \frac{1}{\sqrt{n}}\left(\mathbf{S}^{-1}\mathbf{D}\right)'\mathbf{u} \end{pmatrix}\right|\mathbf{z}\right) = \begin{bmatrix} \frac{1}{N}\mathbf{G}'\mathbf{S}^{-1\prime}\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{S}^{-1}\mathbf{G} & \frac{1}{n^{\frac{3}{2}}}\mathbf{G}'\mathbf{S}^{-1\prime}\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{S}^{-1}\mathbf{D} \\ \frac{1}{n^{\frac{3}{2}}}\mathbf{D}'\mathbf{S}^{-1\prime}\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{S}^{-1}\mathbf{G} & \frac{1}{n}\mathbf{D}'\mathbf{S}^{-1\prime}\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{S}^{-1}\mathbf{D} \end{bmatrix}.$$

Then, the the form of $P - BC^{-1}Q' - QC^{-1}B' + BC^{-1}RC^{-1}B'$ is

$$\begin{aligned}
\boldsymbol{\Omega}_{\theta,N} =& \frac{1}{N}\mathbf{G}'\mathbf{S}^{-1\prime}\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{S}^{-1}\mathbf{G} - \frac{1}{n\sqrt{n}}\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{D}\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{S}^{-1\prime}\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{S}^{-1}\mathbf{G} \\
&- \frac{1}{n\sqrt{n}}\mathbf{G}'\mathbf{S}^{-1\prime}\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{S}^{-1}\mathbf{D}\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{G} \\
&+ \frac{1}{n\sqrt{n}}\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{D}\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\frac{1}{n}\mathbf{D}'\mathbf{S}^{-1\prime}\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{S}^{-1}\mathbf{D}\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{G} \\
=& \frac{1}{N}\mathbf{G}'\mathbf{S}^{-1\prime}\begin{pmatrix} \mathbb{E}\left(\mathbf{u}\mathbf{u}'\right) - \mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{D}\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\mathbf{D}'\mathbf{S}^{-1\prime}\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right) - \mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{S}^{-1}\mathbf{D}\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\mathbf{D}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu}) \\ +\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{D}\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\mathbf{D}'\mathbf{S}^{-1\prime}\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{S}^{-1}\mathbf{D}\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\mathbf{D}'\mathbf{S}^{-1}\mathrm{Diag}(\boldsymbol{\mu}) \end{pmatrix}\mathbf{S}^{-1}\mathbf{G} \\
=& \frac{1}{N}\mathbf{G}'\mathbf{S}^{-1\prime}\left(\left(I_N - \mathbf{P_D}\mathrm{Diag}(\boldsymbol{\mu})\right)'\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\left(I_N - \mathbf{P_D}\mathrm{Diag}(\boldsymbol{\mu})\right)\right)\mathbf{S}^{-1}\mathbf{G} \\
=& \frac{1}{N}\mathbf{G}'\mathbf{S}^{-1\prime}\mathbf{M}_{\mathbf{D}}'\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{M_D}\mathbf{S}^{-1}\mathbf{G}.
\end{aligned}$$

The approximated variance of $\boldsymbol{\phi}$ can be obtained by the following expansion:

$$\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\begin{pmatrix} \frac{1}{n}\mathbf{D}'\mathbf{S}^{-1\prime}\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{S}^{-1}\mathbf{D}-\frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{S}^{-1\prime}\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{S}^{-1}\mathbf{G}\boldsymbol{\Sigma}_{\theta,N}^{-1}\frac{1}{n\sqrt{n}}\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{D} \\ -\frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\frac{1}{n\sqrt{n}}\mathbf{S}^{-1}\mathbf{G}\boldsymbol{\Sigma}_{\theta,N}^{-1}\mathbf{G}'\mathbf{S}^{-1\prime}\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{S}^{-1}\mathbf{D} \\ +\frac{1}{n}\mathbf{D}'\mathbf{S}^{-1\prime}\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{S}^{-1}\mathbf{D}\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{G}\boldsymbol{\Sigma}_{\theta,N}^{-1}\frac{1}{n\sqrt{n}}\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{D} \\ +\frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{G}\boldsymbol{\Sigma}_{\theta,N}^{-1}\frac{1}{n\sqrt{n}}\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{D}\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\frac{1}{n}\mathbf{D}'\mathbf{S}^{-1\prime}\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{S}^{-1}\mathbf{D} \end{pmatrix}\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}$$

$$+\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{G}\boldsymbol{\Sigma}_{\theta,N}^{-1}\boldsymbol{\Omega}_{\theta,N}\boldsymbol{\Sigma}_{\theta,N}^{-1}\frac{1}{n\sqrt{n}}\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{D}\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}$$

$$=n\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\begin{pmatrix} \mathbf{D}'\mathbf{S}^{-1\prime}\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{S}^{-1}\mathbf{D} \\ -\mathbf{D}'\mathbf{S}^{-1\prime}\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{S}^{-1}\mathbf{G}\left(\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{M}_{\mathbf{D}}\mathbf{S}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{D} \\ -\mathbf{D}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{G}\left(\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{M}_{\mathbf{D}}\mathbf{S}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{S}^{-1\prime}\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{S}^{-1}\mathbf{D} \\ +\mathbf{D}'\mathbf{S}^{-1\prime}\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{P}_{\mathbf{D}}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{G} \\ \times\left(\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{M}_{\mathbf{D}}\mathbf{S}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{D} \\ +\mathbf{D}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{G}\left(\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{M}_{\mathbf{D}}\mathbf{S}^{-1}\mathbf{G}\right)^{-1} \\ \times\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{P}_{\mathbf{D}}\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{S}^{-1}\mathbf{D} \end{pmatrix}\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}$$

$$+\,n\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\mathbf{D}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{G}\left(\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{M}_{\mathbf{D}}\mathbf{S}^{-1}\mathbf{G}\right)^{-1}$$

$$\times\,\mathbf{G}'\mathbf{S}^{-1\prime}\mathbf{M}_{\mathbf{D}}'\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{M}_{\mathbf{D}}\mathbf{S}^{-1}\mathbf{G}\times\left(\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{M}_{\mathbf{D}}\mathbf{S}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{D}\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}.$$

Hence, the approximated variance of $\boldsymbol{\phi}$ is

$$\mathbf{V}_{\boldsymbol{\phi},N}=n\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}\mathbf{D}'\mathbf{S}^{-1\prime}\mathbf{M}_{\boldsymbol{\phi}}'\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{M}_{\boldsymbol{\phi}}\mathbf{S}^{-1}\mathbf{D}\left(\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1},$$

where

$$\mathbf{M}_{\boldsymbol{\phi}}=I_N-\mathbf{M}_{\mathbf{D}}\mathbf{S}^{-1}\mathbf{G}\left(\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{M}_{\mathbf{D}}\mathbf{S}^{-1}\mathbf{G}\right)^{-1}\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu}).$$

For the above, note that

- $C^{-1}=\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1}$,
- $R=\frac{1}{n}\mathbf{D}'\mathbf{S}^{-1\prime}\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{S}^{-1}\mathbf{D}$,
- $Q'=\frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{S}^{-1\prime}\mathbb{E}(\mathbf{u}\mathbf{u}')\mathbf{S}^{-1}\mathbf{G}$,
- $B=\frac{1}{n\sqrt{n}}\mathbf{G}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{D}$ and $B'=\frac{1}{n\sqrt{n}}\mathbf{D}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{G}$.

Step 1: Asymptotic expansion of $\hat{\theta}$

As a first step, we need to check the regularity conditions for the asymptotic expansion of $\hat{\theta}$ (Assumption B.1 in Fernandez-Val and Weidner (2016)). Note that the conditions (i) $\frac{\dim(\boldsymbol{\phi}_{2n})}{\sqrt{N}}=\frac{2n}{n}=2>0$ and (ii) smoothness of $\ell_N(\theta,\boldsymbol{\phi})$ in Assumption B.1 in Fernandez-Val and Weidner (2016) are satisfied. The third condition corresponds to the conditions (iv), (v), and (vi) in Assumption B.1 of Fernandez-Val and Weidner (2016).

The last regularity condition is strict concavity of $\ell_N(\boldsymbol{\theta})$. Due to network influences generated by the model, this is not trivial compared to usual PPMLE estimation.

Step 1 (Strict concavity). : Lemma 2.4 illustrates the conditions for strict concavity of $\ell_N(\boldsymbol{\theta})$.

Lemma 2.4. From (39), recall that $\partial_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_N(\boldsymbol{\theta}) = -\mathbf{H}^A(\boldsymbol{\theta}) - \mathbf{H}^B(\boldsymbol{\theta})$, where

$$\mathbf{H}^A(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{G}'(\boldsymbol{\theta}) \\ \mathbf{D}' \end{bmatrix} \mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu}(\boldsymbol{\theta}))\mathbf{S}^{-1} \begin{bmatrix} \mathbf{G}(\boldsymbol{\theta}) & \mathbf{D} \end{bmatrix}$$

and

$$\mathbf{H}^B(\boldsymbol{\theta}) = -\begin{bmatrix} \mathbf{H}^{\theta\theta}(\boldsymbol{\theta}) & \mathbf{H}^{\phi\theta\prime}(\boldsymbol{\theta}) \\ \mathbf{H}^{\phi\theta}(\boldsymbol{\theta}) & \mathbf{H}^{\phi\phi} \end{bmatrix}.$$

Let $\widetilde{\boldsymbol{\Theta}} = \widetilde{\Theta}_\lambda \times \widetilde{\Theta}_\beta \times \widetilde{\boldsymbol{\Theta}}_{\boldsymbol{\alpha}} \times \widetilde{\boldsymbol{\Theta}}_{\boldsymbol{\eta}}$ be parameter space containing possible values of $\boldsymbol{\theta}$. Here, $\widetilde{\Theta}_\lambda$, $\widetilde{\Theta}_\beta$, $\widetilde{\boldsymbol{\Theta}}_{\boldsymbol{\alpha}}$, and $\widetilde{\boldsymbol{\Theta}}_{\boldsymbol{\eta}}$ denote sub-parameter spaces for λ, β, $\boldsymbol{\alpha}$, and $\boldsymbol{\eta}$, respectively.

(i) Then, $\mathbf{H}^A(\boldsymbol{\theta})$ is positive definite for all possible values $\boldsymbol{\theta}$ in $\widetilde{\boldsymbol{\Theta}}$.

(ii) Let $\boldsymbol{\Theta} = \boldsymbol{\Theta}_\lambda \times \boldsymbol{\Theta}_\beta \times \boldsymbol{\Theta}_{\boldsymbol{\alpha}} \times \boldsymbol{\Theta}_{\boldsymbol{\eta}}$ be a parameter space satisfying $\inf_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left(\varphi_{\min}\left(\mathbf{H}^A(\boldsymbol{\theta})\right) + \varphi_{\min}\left(\mathbf{H}^B(\boldsymbol{\theta})\right)\right) > 0$, and assume $\Theta_\lambda \subseteq \widetilde{\Theta}_\lambda$, $\Theta_\beta \subseteq \widetilde{\Theta}_\beta$, $\Theta_{\boldsymbol{\alpha}} \subseteq \widetilde{\Theta}_{\boldsymbol{\alpha}}$ and $\Theta_{\boldsymbol{\eta}} \subseteq \widetilde{\Theta}_{\boldsymbol{\eta}}$.

Then, $\ell_N(\boldsymbol{\theta})$ is strict concave for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Here, $\varphi_{\min}(M)$ denotes the minimum eigenvalue of M.

Proof of Lemma 2.4. First, by construction, observe $\mathrm{Diag}(\boldsymbol{\mu}(\boldsymbol{\theta}))$ is a diagonal matrix with strictly positive elements for any $\boldsymbol{\theta} \in \tilde{\boldsymbol{\Theta}}$. By Assumption 2.2, $\mathbf{S}(\lambda)$ is invertible when $\lambda \subseteq \Theta_\lambda \in \tilde{\Theta}_\lambda$. Hence, $\mathbf{S}^{-1}(\lambda)$ is of full rank for $\lambda \in \Theta_\lambda$. Since $\begin{bmatrix} \mathbf{G}(\boldsymbol{\theta}) & \mathbf{D} \end{bmatrix}$ is a nonzero matrix, we verify $\mathbf{H}^A(\boldsymbol{\theta})$ is positive definite. In consequence, the major part of $\partial_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_N(\boldsymbol{\theta})$ is negative definite.

Second, it suffices to show $\mathbf{H}^A(\boldsymbol{\theta}) + \mathbf{H}^B(\boldsymbol{\theta})$ is positive definite since $\ell_N(\boldsymbol{\theta})$ is infinitely differentiable. Since $\mathbf{H}^A(\boldsymbol{\theta})$ and $\mathbf{H}^B(\boldsymbol{\theta})$ are symmetric, their all eigenvalues are real-valued. By Lemma A.5 in Ahn and Horenstein (2013) and our assumption, we have

$$\varphi_{\min}\left(\mathbf{H}^A(\boldsymbol{\theta}) + \mathbf{H}^B(\boldsymbol{\theta})\right) \geq \varphi_{\min}\left(\mathbf{H}^A(\boldsymbol{\theta})\right) + \varphi_{\min}\left(\mathbf{H}^B(\boldsymbol{\theta})\right) > 0.$$

Since the minimum eigenvalue of $\mathbf{H}^A(\boldsymbol{\theta}) + \mathbf{H}^B(\boldsymbol{\theta})$ is negative, $\mathbf{H}^A(\boldsymbol{\theta}) + \mathbf{H}^B(\boldsymbol{\theta})$ is positive definite. Then, we complete the proof.■

Lemma 2.4 specifies the parameter space $\boldsymbol{\Theta}$ guaranteeing strict concavity of $\ell_N(\boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. Note that the main part of $\partial_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_N(\boldsymbol{\theta})$ is $\mathbf{H}^A(\boldsymbol{\theta})$, and $\mathbf{H}^B(\boldsymbol{\theta})$ is a new term generated by the spatial interaction term and penalty term for the identification of fixed effects. Since $\mathbf{H}^A(\boldsymbol{\theta})$ is positive definite if $\mathbf{S}(\lambda)$ is invertible, $\varphi_{\min}\left(\mathbf{H}^A(\boldsymbol{\theta})\right)$ is positive and far from zero. On the other hand, $\mathbf{H}^B(\boldsymbol{\theta})$ might be indefinite. Lemma 2.4 means that strict concavity of $\ell_N(\boldsymbol{\theta})$ is achievable if the minimum eigenvalue of the minor part $\mathbf{H}^B(\boldsymbol{\theta})$ does not dominate $\varphi_{\max}\left(\mathbf{H}^A(\boldsymbol{\theta})\right)$.

Since the condition in Lemma 2.4 guarantees for strict concavity of $\ell_N(\boldsymbol{\theta})$, there is a unique solution to the optimization problem, $\hat{\boldsymbol{\theta}} = \mathrm{argmax}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\ell_N(\boldsymbol{\theta})$. Hence, first, this condition directly links to identification conditions for $\boldsymbol{\theta}^0$, i.e., $\boldsymbol{\theta}^0$ is a unique solution to $\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \ell_\infty(\boldsymbol{\theta})$, where $\ell_\infty(\boldsymbol{\theta}) \equiv \mathrm{plim}_{n\to\infty}\frac{1}{N}\ell_N(\boldsymbol{\theta})$ for each $\boldsymbol{\theta}$. Further, this condition can be restrictive since it requires strict concavity of $\ell_N(\boldsymbol{\theta})$ for all possible $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. This is because $\boldsymbol{\Theta}$ grows corresponding to n. Hence, we want to find some conditions, which are milder than the condition in Lemma 2.4. For this purpose, let $\Theta_\theta = \Theta_\lambda \times \Theta_\beta$ and $\boldsymbol{\Theta}_{\boldsymbol{\phi}} = \boldsymbol{\Theta}_{\boldsymbol{\alpha}} \times \boldsymbol{\Theta}_{\boldsymbol{\eta}}$.

Lemma 2.5. (i) Assume $\mathrm{liminf}_{n\to\infty} \inf_{\boldsymbol{\phi}\in\boldsymbol{\Theta}_{\boldsymbol{\phi}}} \varphi_{\min}\left(\frac{1}{n}\mathbf{D}'\mathbf{S}^{-1\prime}(\lambda)\mathrm{Diag}\left(\boldsymbol{\mu}(\boldsymbol{\theta})\right)\mathbf{S}^{-1}(\lambda)\mathbf{D} - \frac{1}{n}\mathbf{H}^{\phi\phi}\right) > 0$ for each $\theta \in \Theta_\theta$. Then, $\hat{\boldsymbol{\phi}}(\theta) = \mathrm{argmax}_{\boldsymbol{\phi}\in\boldsymbol{\Theta}_{\boldsymbol{\phi}}}\ell_N\left(\theta, \boldsymbol{\phi}\right)$ is unique for each $\theta \in \Theta_\theta$ and for a sufficiently large n.

(ii) For each $\theta \in \Theta_\theta$, let

$$\begin{aligned}\widehat{\mathbf{H}}(\theta) =& \frac{1}{N}\widehat{\mathbf{G}}'(\theta)\mathbf{S}^{-1\prime}(\lambda)\mathrm{Diag}\left(\widehat{\boldsymbol{\mu}}(\theta)\right)\mathbf{S}^{-1}(\lambda)\widehat{\mathbf{G}}(\theta) - \frac{1}{N}\widehat{\mathbf{H}}^{\theta\theta}(\theta) - \frac{1}{N}\left(\widehat{\mathbf{G}}'(\theta)\mathbf{S}^{-1\prime}(\lambda)\mathrm{Diag}\left(\widehat{\boldsymbol{\mu}}(\theta)\right)\mathbf{S}^{-1}(\lambda)\mathbf{D} - \widehat{\mathbf{H}}^{\phi\theta\prime}(\theta)\right) \\ & \cdot \left(\mathbf{D}\mathbf{S}^{-1\prime}(\lambda)\mathrm{Diag}\left(\widehat{\boldsymbol{\mu}}(\theta)\right)\mathbf{S}^{-1}(\lambda)\mathbf{D} - \mathbf{H}^{\phi\phi}\right)^{-1} \cdot \left(\mathbf{D}'\mathbf{S}^{-1\prime}(\lambda)\mathrm{Diag}\left(\widehat{\boldsymbol{\mu}}(\theta)\right)\mathbf{S}^{-1}(\lambda)\widehat{\mathbf{G}}(\theta) - \widehat{\mathbf{H}}^{\phi\theta}(\theta)\right)\end{aligned}$$

where $\widehat{\mathbf{G}}(\theta) = \mathbf{G}\left(\theta, \hat{\boldsymbol{\phi}}(\theta)\right)$, $\widehat{\boldsymbol{\mu}}(\theta) = \boldsymbol{\mu}(\theta, \hat{\boldsymbol{\phi}}(\theta))$, $\widehat{\mathbf{H}}^{\theta\theta}(\theta) = \mathbf{H}(\theta, \hat{\boldsymbol{\phi}}(\theta))$, and $\widehat{\mathbf{H}}^{\phi\theta}(\theta) = \mathbf{H}^{\phi\theta}(\theta, \hat{\boldsymbol{\phi}}(\theta))$ for each $\theta \in \Theta_\theta$.

Assume $\mathrm{liminf}_{n\to\infty} \inf_{\theta\in\Theta_\theta} \varphi_{\min}\left(\widehat{\mathbf{H}}(\theta)\right) > 0$. Then, $\hat{\theta} = \mathrm{argmax}_{\theta\in\Theta_\theta} \ell_N^{\mathrm{c}}(\theta)$ is unique for a sufficiently large n.

Proof of Lemma 2.5. (i) Fix $\theta \in \Theta_\theta$ and consider $\mathrm{argmax}_{\boldsymbol{\phi}\in\boldsymbol{\Theta}_\phi} \ell_N(\theta, \boldsymbol{\phi})$. The first-order condition of this problem is $\partial_{\boldsymbol{\phi}}\ell_N\left(\theta, \hat{\boldsymbol{\phi}}(\theta)\right) = 0$, where $\hat{\boldsymbol{\phi}}(\theta)$ is a solution to $\max_{\boldsymbol{\phi}\in\boldsymbol{\Theta}_\phi} \ell_N(\theta, \boldsymbol{\phi})$. To achieve uniqueness of $\hat{\boldsymbol{\phi}}(\theta)$, a sufficient condition is $\partial_{\boldsymbol{\phi}\boldsymbol{\phi}}\ell_N(\theta, \boldsymbol{\phi}) < 0$ for all $\boldsymbol{\phi} \in \boldsymbol{\Theta}_\phi$. Since $\frac{1}{n}\partial_{\boldsymbol{\phi}\boldsymbol{\phi}}\ell_N(\theta, \boldsymbol{\phi}) = -\frac{1}{n}\mathbf{D}'\mathbf{S}^{-1\prime}(\lambda)\mathrm{Diag}\left(\boldsymbol{\mu}(\boldsymbol{\theta})\right)\mathbf{S}^{-1}(\lambda)\mathbf{D} + \frac{1}{n}\mathbf{H}^{\phi\phi}$ and $-\frac{1}{n}\mathbf{D}'\mathbf{S}^{-1\prime}(\lambda)\mathrm{Diag}\left(\boldsymbol{\mu}(\boldsymbol{\theta})\right)\mathbf{S}^{-1}(\lambda)\mathbf{D} + \frac{1}{n}\mathbf{H}^{\phi\phi} = O(1)$, the uniqueness can be achieve when the condition in Lemma 2.5 (i) is satisfied.

(ii) Suppose that $\hat{\boldsymbol{\phi}}(\theta)$ is unique for each $\theta \in \Theta_\theta$. Then, the next step is to find a condition for the uniqueness of $\hat{\theta} = \mathrm{argmax}_{\theta\in\Theta_\theta} \ell_N^{\mathrm{c}}(\theta)$. Note that $\hat{\theta}$ satisfies

$$\begin{aligned} 0 = \partial_\theta \ell_N^{\mathrm{c}}(\theta) &= \partial_\theta \ell_N\left(\theta, \hat{\boldsymbol{\phi}}(\theta)\right) + \partial_{\boldsymbol{\phi}}\ell_N\left(\theta, \hat{\boldsymbol{\phi}}(\theta)\right)\partial_\theta\hat{\boldsymbol{\phi}}(\theta) \\ &= \partial_\theta \ell_N\left(\theta, \hat{\boldsymbol{\phi}}(\theta)\right)\end{aligned}$$

since $\partial_{\boldsymbol{\phi}}\ell_N\left(\theta, \hat{\boldsymbol{\phi}}(\theta)\right) = 0$ for all $\theta \in \Theta_\theta$.

Then, a sufficient condition for the uniqueness of $\hat{\theta}$ is $\partial_{\theta\theta}\ell_N^{\mathrm{c}}(\theta) < 0$ for all $\theta \in \Theta_\theta$. Observe that

$$\begin{aligned}\frac{1}{N}\partial_{\theta\theta}\ell_N^{\mathrm{c}}(\theta) =& \frac{1}{N}\partial_\theta\left(\partial_\theta\ell_N(\theta, \hat{\boldsymbol{\phi}}(\theta)) + \partial_{\boldsymbol{\phi}}\ell_N(\theta, \hat{\boldsymbol{\phi}}(\theta))\right)\partial_\theta\hat{\boldsymbol{\phi}}(\theta) \\ =& \frac{1}{N}\partial_{\theta\theta}\ell_N(\theta, \hat{\boldsymbol{\phi}}(\theta)) - \frac{1}{n}\left(\frac{1}{n}\partial_{\theta\boldsymbol{\phi}}\ell_N(\theta, \hat{\boldsymbol{\phi}}(\theta))\right) \cdot \left(\frac{1}{n}\partial_{\boldsymbol{\phi}\boldsymbol{\phi}}\ell_N(\theta, \hat{\boldsymbol{\phi}}(\theta))\right)^{-1} \cdot \left(\frac{1}{n}\partial_{\boldsymbol{\phi}\theta}\ell_N(\theta, \hat{\boldsymbol{\phi}}(\theta))\right) \\ =& -\frac{1}{N}\mathbf{G}'(\theta, \hat{\boldsymbol{\phi}}(\theta))\mathbf{S}^{-1\prime}(\lambda)\mathrm{Diag}\left(\boldsymbol{\mu}(\theta, \hat{\boldsymbol{\phi}}(\theta))\right)\mathbf{S}^{-1}(\lambda)\mathbf{G}(\theta, \hat{\boldsymbol{\phi}}(\theta)) + \frac{1}{N}\mathbf{H}^{\theta\theta}(\theta, \hat{\boldsymbol{\phi}}(\theta)) \\ & - \frac{1}{n}\left(-\frac{1}{n}\mathbf{G}'(\theta, \hat{\boldsymbol{\phi}}(\theta))\mathbf{S}^{-1\prime}(\lambda)\mathrm{Diag}\left(\boldsymbol{\mu}(\theta, \hat{\boldsymbol{\phi}}(\theta))\right)\mathbf{S}^{-1}(\lambda)\mathbf{D} + \frac{1}{n}\mathbf{H}^{\phi\theta\prime}(\theta, \hat{\boldsymbol{\phi}}(\theta))\right) \\ & \cdot \left(-\frac{1}{n}\mathbf{D}\mathbf{S}^{-1\prime}(\lambda)\mathrm{Diag}\left(\boldsymbol{\mu}(\theta, \hat{\boldsymbol{\phi}}(\theta))\right)\mathbf{S}^{-1}(\lambda)\mathbf{D} + \frac{1}{n}\mathbf{H}^{\phi\phi}\right)^{-1} \\ & \cdot \left(-\frac{1}{n}\mathbf{D}'\mathbf{S}^{-1\prime}(\lambda)\mathrm{Diag}\left(\boldsymbol{\mu}(\theta, \hat{\boldsymbol{\phi}}(\theta))\right)\mathbf{S}^{-1}(\lambda)\mathbf{G}(\theta, \hat{\boldsymbol{\phi}}(\theta)) + \frac{1}{n}\mathbf{H}^{\phi\theta}(\theta, \hat{\boldsymbol{\phi}}(\theta))\right) \\ =& -\widehat{\mathbf{H}}(\theta).\end{aligned}$$

Hence, if the condition in Lemma 2.5 (i) is satisfied, $\hat{\theta}$ is unique.■

Lemma 2.5 characterizes the identification conditions for the first-stage parameters $\boldsymbol{\theta}^0$ when n is large. Condition (i) leads to the uniqueness of $\boldsymbol{\phi}(\theta) = \mathrm{argmax}_{\boldsymbol{\phi}\in\boldsymbol{\Theta}_\phi} \ell_\infty(\theta, \boldsymbol{\phi})$ for each $\theta \in \Theta_\theta$. Condition (ii) implies that the uniqueness of $\theta^0 = \mathrm{argmax}_{\theta\in\Theta_\theta} \ell_\infty(\theta, \boldsymbol{\phi}(\theta))$, and consequently, $\boldsymbol{\phi}^0 = \boldsymbol{\phi}(\theta^0)$.

Step 2 (Convergence of the fixed-effect estimators): Based on the established regularity conditions, our next step is to show convergence of the fixed-effect estimators. For each $\theta \in \Theta_\theta$, we let

$$\hat{\boldsymbol{\phi}}(\theta) = (\hat{\boldsymbol{\alpha}}(\theta)', \hat{\boldsymbol{\eta}}(\theta)')' = \operatorname{argmax}_{\boldsymbol{\phi} \in \boldsymbol{\Theta}_{\boldsymbol{\phi}}} \ell_N(\theta, \boldsymbol{\phi}).$$

Lemma 2.6. Assume $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

Variance estimation

The assumption below describes the properties of $\mathsf{K}(\cdot)$ and the inside argument of $\mathsf{K}(\cdot)$, which gives weights for the different covariances as a function of distances between two pairs.

Assumption 2.7. (i) The kernel $\mathsf{K} : \mathbb{R}_+ \cup \{0\} \to [-1, 1]$, with $\mathsf{K}(0) = 1$, $\mathsf{K}(x) = \mathsf{K}(-x)$, $\mathsf{K}(x) = 0$ for $|x| > 1$, satisfies

$$|\mathsf{K}(x) - 1| \leq c_{\mathsf{K}} \cdot |x|^{\rho_{\mathsf{K}}} \text{ for } |x| \leq 1,$$

for some $0 < c_{\mathsf{K}} < \infty$ and $\rho_{\mathsf{K}} \geq 1$.

(ii) $d^{p*}_{ij,kl} = d^{p}_{ij,kl} + v_{ij,kl}$, where $v_{ij,kl} = v_{kl,ij}$ represents the independently generated bounded measurement errors. That is, $|v_{ij,kl}| \leq c_v$, with $0 < c_v < \infty$, and $\{v_{ij,kl}\}$ is independent of $\{\xi_{ij}\}$ and any component of $\mathbf{z}$.

(iii) $d_N \to \infty$ as $n \to \infty$. Let $\deg^*_{ij} = \sum_{k=1}^n \sum_{l=1}^n \mathbf{1}\{d^{p*}_{ij,kl} \leq d_N\}$ and $\overline{\deg}^* = \max_{i,j=1,\cdots,n} \deg^*_{ij}$. [To be added]

2.4 Estimation and inference on γ_d and γ_o

From (7), we can recognize α^0_l and η^0_k in equation (9) should be fixed-effect components since they are complex functions of $\{x_{kl}\}^n_{k,l=1}$ and $\{x^u_k\}^n_{k=1}$.

By equation (7), we assume that $\{\alpha^0_l\}^n_{l=1}$ and $\{\eta^0_k\}^n_{k=1}$ are generated by the following processes:

$$\begin{aligned} \alpha^0_l &= \alpha^0_c + \ln(G_l) + \ln\left(\Pi^{\varrho-1}_l(\boldsymbol{\mu}^0)\right) + x^u_l \gamma^0_o + \epsilon^\alpha_l \text{ for } l = 1, \cdots, n, \text{ and} \\ \eta^0_k &= \eta^0_c + \ln(G_k) + \ln\left(P^{\varrho-1}_k(\boldsymbol{\mu}^0)\right) + x^u_k \gamma^0_d + \epsilon^\eta_k \text{ for } k = 1, \cdots, n, \end{aligned} \tag{42}$$

where α^0_c and η^0_c are constants (both constants represent $-\frac{1}{2}\ln(G^W)$), γ^0_o and γ^0_d represent vectors of the true parameters, and ϵ^α_l and ϵ^η_k are error components.

3 Additional simulation analysis

[To be added]

4 Empirical Application

Countries with large column sums: USA, JPN (Phase 1, 2), USA, DEU (Phase 3), USA, CHN, DEU (Phase 4).

References

Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica*, 81(3):1203–1227.

Anderson, J. and van Wincoop, E. (2003). Gravity with gravitas: a solution to the border puzzle. *American Economic Review*, 93:170–192.

Bramoullé, Y., Kranton, R., and D'Amours, M. (2014). Strategic interaction and networks. *American Economic Review*, 104(3):898–930.

Chung, F. R. (1997). *Spectral graph theory*, volume 92. American Mathematical Soc.

Dixit, A. K. and Stiglitz, J. E. (1993). Monopolistic competition and optimum product diversity: Reply. *The American Economic Review*, 83(1):302–304.

Eaton, J. and Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5):1741–1779.

Fernandez-Val, I. and Weidner, M. (2016). Individual and time effects in nonlinear panel models with large n, t. *Journal of Econometrics*, 192:291–312.

Head, K. and Mayer, T. (2014). Chapter 3 - gravity equations: Workhorse,toolkit, and cookbook. In Gopinath, G., Helpman, E., and Rogoff, K., editors, *Handbook of International Economics*, volume 4 of *Handbook of International Economics*, pages 131–195. Elsevier.

Jenish, N. and Prucha, I. (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics*, 150:86–98.

Jenish, N. and Prucha, I. (2012). On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics*, 170:178–190.

Jeong, H. and Lee, L. (2024). Maximum likelihood estimation of a spatial autoregressive model for origin-destination flow variables. *Journal of Econometrics*, 242:105790.

Krugman, P. (1995). *Handbook of International Economics - Increasing returns, imperfect competition and the positive theory of international trade*, volume 3. Elsevier.

LeSage, J. and Pace, R. (2008). Spatial econometric modeling of origin-destination flows. *Journal of Regional Science*, 48:941–967.

Lind, N. and Ramondo, N. (2023). Trade with correlation. *American Economic Review*, 113(2):317–53.

McCallum, J. (1995). National borders matter: Canada-u.s. regional trade patterns. *American Economic Review*, 85:615–623.

Mullahy, J. and Norton, E. (2024). Why transform Y? A the pitfalls of transformed regressions with a mass at zero. *Oxford Bulletin of Economics and Statistics*, 86:417 – 447.

Santos Silva, J. and Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics*, 88:641–658.

Tinbergen, J. (1962). *Shaping the World Economy.* Net York: The Twentieth Century Fund.