Supplement for "Connected Trade Flows via Trade Cost: Spatial Autoregressive Framework"

Hanbat Jeong* and Jieun Lee†

Abstract

This document contains some technical proofs, additional MC, and empirical results for Jeong and Lee (2026). 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 $\dim(x_{ij}) = 1$:

$$y_{ij} = \underbrace{\exp\left(\beta_0^0 + \beta_1^0 x_{ij}\right)}_{=\mu_{ij}=\mathbb{E}(y_{ij}|x_{ij})} \cdot \xi_{ij} \Leftrightarrow y_{ij} = \mu_{ij} + u_{ij}, \text{ where } u_{ij} = \mu_{ij}(\xi_{ij} - 1), \tag{1.1}$$

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

*Department of Economics, Macquarie Business School, Macquarie University.
E-mail: *hanbat.jeong@mq.edu.au*

†Department of Economics, Emory University.
E-mail: *jieunlee.sophia@gmail.com*

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:

$$[\beta_0]: \quad \mathbb{E}(u_{ij}) = \mathbb{E}\left(y_{ij} - \exp\left(\beta_0^0 + \beta_1^0 x_{ij}\right)\right) = 0, \text{ and} \tag{1.2}$$

$$[\beta_1]: \quad \mathbb{E}(x_{ij}u_{ij}) = \mathbb{E}\left(x_{ij}\left(y_{ij} - \exp\left(\beta_0^0 + \beta_1^0 x_{ij}\right)\right)\right) = 0. \tag{1.3}$$

We now consider the log transformation of (1.1) to estimate β_0^0 and β_1^0:

$$\ln(y_{ij}) = \beta_0^0 + \beta_1^0 x_{ij} + v_{ij}, \tag{1.4}$$

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, $\ln\left(\mathbb{E}(y_{ij}|x_{ij})\right) \neq \mathbb{E}(\ln(y_{ij})|x_{ij})$). Santos Silva and Tenreyro (2006) point out that the gap $\mathbb{E}\left(\ln(y_{ij})\right) - \ln\left(\mathbb{E}(y_{ij})\right) < 0$ 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 $\ln\left(\mathbb{E}(y_{ij})\right) = \ln(1 \cdot 0.5 + 0 \cdot 0.5) \simeq -0.6931$. 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 $\mathbb{E}\left(\ln(y_{ij}+1)\right) = 0.5 \cdot \ln(1+1) + 0.5 \cdot \ln(0+1) \simeq 0.3466$. The gap is about 1.0397.

2. Suppose $y_{ij} = \exp\left(\tilde{y}_{ij}\right)$, where $\tilde{y}_{ij} \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$. Then, $\mathbb{E}\left(\ln(y_{ij})\right) - \ln\left(\mathbb{E}(y_{ij})\right) = \mu - \left(\mu + \frac{1}{2}\sigma^2\right) = -\frac{1}{2}\sigma^2$. 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 (1.1) to estimate the constant elasticities. To see this, suppose $y_{ij}^* := 100 \cdot y_{ij}$, i.e., $y_{ij}^* \overset{i.i.d.}{\sim} 100 \cdot \text{Bernoulli}(0.5)$. 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., $y_{*,ij} \overset{i.i.d.}{\sim} 0.01 \cdot \text{Bernoulli}(0.5)$. Observe that $\ln\left(\mathbb{E}(y_{*,ij})\right) = \ln(0.01 \cdot 1 \cdot 0.5 + 0.01 \cdot 0 \cdot 0.5) \simeq -5.2983$. Now consider $\mathbb{E}\ln(y_{*,ij})$, where some arbitrary constant (e.g., 1) is added for zero outcomes to be defined so that $\mathbb{E}\left(\ln(y_{*,ij}+1)\right) = \ln(0.01 \cdot 1 + 1) \cdot 0.5 + \ln(0.01 \cdot 0 + 1) \cdot 0.5 \simeq 0.0050$. 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 (1.1). 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 $\beta^0 = (\beta_0^0, \beta_1^0)'$.

Regarding (1.2), by the Maclaurin series expansion for $\mathbb{E}(\ln(\xi_{ij}))$, observe that

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

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, $\mathbb{E}\left(\ln(\xi_{ij})\right)$ 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 (1.4).

Regarding (1.3), observe that

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

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. $\mathbb{E}\left(x_{ij}v_{ij}\right) = 0$ 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, $\left.\frac{d\ln(y_{ij}+c)}{dy_{ij}}\right|_{y_{ij}=0} = \frac{1}{c}$. This quantity means the marginal change of the log-transformed outcome $\ln(y_{ij}+c)$ when $y_{ij} = 0$. Then,

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

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 (1.5).

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 (1.4). The asymptotic bias of $\hat{\beta}^+(c)$ can be characterized by the following difference:

$$\begin{aligned}\hat{\beta}^+(c) - \beta^0 = &\begin{bmatrix} 1 & \frac{1}{N}\sum_{i,j=1}^n x_{ij} \\ \frac{1}{N}\sum_{i,j=1}^n x_{ij} & \frac{1}{N}\sum_{i,j=1}^n x_{ij}^2 \end{bmatrix}^{-1} \cdot \begin{pmatrix} \frac{1}{N}\sum_{i,j=1}^n v_{ij} \\ \frac{1}{N}\sum_{i,j=1}^n x_{ij}v_{ij} \end{pmatrix} \\ &+ \begin{bmatrix} 1 & \frac{1}{N}\sum_{i,j=1}^n x_{ij} \\ \frac{1}{N}\sum_{i,j=1}^n x_{ij} & \frac{1}{N}\sum_{i,j=1}^n x_{ij}^2 \end{bmatrix}^{-1} \cdot \begin{pmatrix} \frac{1}{N}\sum_{i,j=1}^n \Delta_{y,ij}(c) \\ \frac{1}{N}\sum_{i,j=1}^n x_{ij}\Delta_{y,ij}(c) \end{pmatrix},\end{aligned} \tag{1.5}$$

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

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,
- $\mu_{ij} = \mathbb{E}(y_{ij}|\mathbf{x})$, where $\mathbf{x}$ 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,
- τ_{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

$$\mu_{ij} = \mu \cdot G_i^I \cdot G_j^E \cdot \tau_{ij}, \tag{1.6}$$

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 (1.6) is simplified by $\mu_{ij} = \mu \cdot G_i \cdot G_j \cdot \tau_{ij}$. If τ_{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 (1.6)) only consider the bilateral trade cost between two countries. For example, McCallum (1995) considers the following specification on τ_{ij}:

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

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:

$$\mu_{ij} = \frac{G_i \cdot G_j}{G^W} \cdot \left(\frac{\tau_{ij}}{\Pi_j \cdot P_i} \right)^{1-\varrho}, \tag{1.7}$$

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 $(P_1, \cdots, P_n)$,

$$\Pi_j = \left(\sum_{k=1}^{n} \frac{G_k}{G^W} \left(\frac{\tau_{kj}}{P_k} \right)^{1-\varrho} \right)^{\frac{1}{1-\varrho}}. \tag{1.8}$$

Hence, the outward resistance Π_j represents the overall trade cost from j since each $\frac{\tau_{kj}}{P_k}$ illustrates the normalized trade cost from j to k and Π_j consists of aggregated $\frac{\tau_{kj}}{P_k}$ for $k = 1, \cdots, n$ weighted by the GDP shares $\frac{G_k}{G^W}$ for $k = 1, \cdots, n$. For example, suppose that τ_{kj} (= trade cost from j to k) for some k decreases. A drop in τ_{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 $(\Pi_1, \cdots, \Pi_n)$,

$$P_i = \left(\sum_{k=1}^{n} \frac{G_k}{G^W} \left(\frac{\tau_{ik}}{\Pi_k} \right)^{1-\varrho} \right)^{\frac{1}{1-\varrho}}. \tag{1.9}$$

Like the outward resistance, the inward resistance P_i captures the overall trade cost to i. Like the outward resistance, if τ_{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 τ_{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 τ_{ij} increases under large ϱ, μ_{ij} in equation (1.7) significantly decreases. When $\varrho \to \infty$, $P_i \to \min_{k=1,\cdots,n}\{\frac{\tau_{ik}}{\Pi_k}\}$ and $\Pi_j \to \min_{k=1,\cdots,n}\{\frac{\tau_{kj}}{P_k}\}$. In the case of perfect substitutability, trade flows are dominated by the route with the lowest resistance (i.e., the smallest $\frac{\tau_{ik}}{\Pi_k}$ or $\frac{\tau_{kj}}{P_k}$). On the other hand, τ_{ij} does not play a role in μ_{ij} if $\varrho \to 1$. As $\varrho \to 1$, $P_i = \sum_{k=1}^{n} \frac{\tau_{ik}}{\Pi_k} \cdot \frac{G_k}{G^W}$ and $\Pi_j = \sum_{k=1}^{n} \frac{\tau_{kj}}{P_k} \cdot \frac{G_k}{G^W}$. 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 (1.7) is the true model, conventional gravity specification (e.g., equation (1.6)) omits the multilateral resistance terms. Since the multilateral resistance terms (1.8) and (1.9) contain $\{G_k, \tau_{ik}, \tau_{kj}\}_{k=1}^{n}$, the omitted terms in the traditional gravity equation are dependent on the original components G_i, G_j, and τ_{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 3). We will apply the backsolving procedure. Suppose that the trade cost factors $\{\tau_{ij}\}$ were determined in **Stage 2**.

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:

$$\max_{\{c_{ij}\}_{j=1}^n} U_i = \left(\sum_{j=1}^n \chi_j^{\frac{1}{\varrho}} \cdot c_{ij}^{\frac{\varrho-1}{\varrho}}\right)^{\frac{\varrho}{\varrho-1}} \text{ subject to } \sum_{j=1}^n p_{ij}c_{ij} = G_i, \tag{1.10}$$

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 (1.10), 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

$$c_{ij}^* = \frac{C_i^{\frac{\varrho}{\varrho-1}} \chi_j}{(\lambda_i p_{ij})^{\varrho}}. \tag{1.11}$$

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

$$\min_{\{c_{ij}\}_{j=1}^n} \sum_{j=1}^n p_{ij}c_{ij} \text{ subject to } \left(\sum_{j=1}^n \chi_j^{\frac{1}{\varrho}} \cdot c_{ij}^{\frac{\varrho-1}{\varrho}}\right)^{\frac{\varrho}{\varrho-1}} = \overline{U}_i \tag{1.12}$$

for some $\overline{U}_i$. We set up the Lagrangian to solve (1.12):

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

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 (1.12) 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,

$$
\lambda_i^{**} = \left(\sum_{j=1}^{n} \chi_j p_{ij}^{1-\varrho} \right)^{\frac{1}{1-\varrho}}. \tag{1.13}
$$

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

$$
\begin{aligned}
E_i^{**} &= \sum_{j=1}^{n} p_{ij} \left(\lambda_i^{**} C_i^{\frac{1}{\varrho-1}} \chi_j^{\frac{1}{\varrho}} p_{ij}^{-1} \right)^{\varrho} \\
&= (\lambda_i^{**})^{\varrho} C_i^{\frac{\varrho}{\varrho-1}} \sum_{j=1}^{n} \chi_j p_{ij}^{1-\varrho} \\
&= \lambda_i^{**} \cdot \overline{U}_i
\end{aligned}
\tag{1.14}
$$

by the constraint $\overline{U}_i = C_i^{\frac{\varrho}{\varrho-1}}$ and (1.13). 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 (1.11) 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,

$$\lambda_i = C_i^{\frac{1}{\varrho-1}} P_i^{\frac{1-\varrho}{\varrho}} G_i^{-\frac{1}{\varrho}}. \tag{1.15}$$

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, (1.11) generates the demand function:

$$c_{ij}^* = C_i^{\frac{\rho}{\varrho-1}} \chi_j p_{ij}^{-\varrho} C_i^{-\frac{\rho}{\varrho-1}} P_i^{\varrho-1} G_i = \chi_j \left(\frac{p_{ij}}{P_i} \right)^{-\varrho} \frac{G_i}{P_i}. \tag{1.16}$$

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

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

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

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

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

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

Hence, we have

$$\mu_{ij}^{*} = p_{ij}c_{ij}^{*} = \chi_j p_{ij}^{1-\varrho} P_i^{-(1-\varrho)} G_i = \chi_j (p_j \tau_{ij})^{1-\varrho} P_i^{-(1-\varrho)} G_i. \tag{1.17}$$

The market-clearing condition imposes

$$G_j = \sum_{i=1}^{n} \mu_{ij}^{*} = \chi_j p_j^{1-\varrho} \sum_{i=1}^{n} \tau_{ij}^{1-\varrho} P_i^{-(1-\varrho)} G_i. \tag{1.18}$$

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

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

by the definition in (1.8). Hence, equation (1.17) becomes

$$\mu_{ij} = \frac{G_i G_j}{G^W} \left(\frac{\tau_{ij}}{\Pi_j P_i}\right)^{1-\varrho}. \tag{1.19}$$

Further, we can verify that

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

which is the same as (1.9).

Step 2 (solving Stage 2). The next step is to characterize the equilibrium negotiated trade cost factor τ_{ij}. Suppose that the countries' connectivity matrix W is given from **Stage 1**. Our specification on τ_{ij} is following:

$$\tau_{ij} = \left(\mu_{ij}^{\text{proxy}}\right)^{-1} \cdot \underbrace{D_{ij,1}^{\tilde{\beta}_1} \cdots D_{ij,K}^{\tilde{\beta}_K}}_{\equiv \tau_{ij}^{+}}. \tag{1.20}$$

τ_{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 (τ_{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 $\tilde{\beta}_1, \cdots, \tilde{\beta}_K$.

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

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

$$\mu_{ij}^{\text{proxy}} = \left(\prod_{k=1}^{n} \mu_{kj}^{w_{ik}}\right)^{\tilde{\lambda}_d} \left(\prod_{l=1}^{n} \mu_{il}^{w_{jl}}\right)^{\tilde{\lambda}_o} \left(\prod_{k,l=1}^{n} \mu_{kl}^{w_{ik}w_{jl}}\right)^{\tilde{\lambda}_w}, \tag{1.21}$$

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

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

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

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 τ_{ij} is reduced ($\tau_{ij} \leq \tau_{ij}^{+}$) by utilizing information about the trade cost. On the other hand, if $\tilde{\lambda}_d \simeq \tilde{\lambda}_o \simeq \tilde{\lambda}_w \simeq 0$, $\tau_{ij} \simeq \tau_{ij}^{+}$ since $\mu_{ij}^{\text{proxy}} \simeq 1$. We will provide the detailed interpretations of those geometric averages later.

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

$$\mu_{ij}^{+} = D_{ij,1}^{\beta_1} \cdots D_{ij,K}^{\beta_K}$$

denote the pure exogenous part of μ_{ij}. Note that equation (1.19) can be alternatively represented by

$$\begin{aligned}
\mu_{ij} &= \frac{G_i G_j}{G^W} \left(\frac{\tau_{ij}}{P_i \Pi_j}\right)^{1-\varrho} \\
&= \frac{G_i G_j}{G^W} \cdot P_i^{\varrho-1} \Pi_j^{\varrho-1} \cdot \left(\mu_{ij}^{\text{proxy}}\right)^{\varrho-1} \cdot \left(\tau_{ij}^{+}\right)^{1-\varrho} \\
&= \underbrace{(\overline{\mu}_{.j}^{i})^{\lambda_d} (\overline{\mu}_{i.}^{j})^{\lambda_o} (\overline{\mu}_{..}^{ij})^{\lambda_w}}_{\text{Part A}} \cdot \underbrace{P_i^{\varrho-1} \Pi_j^{\varrho-1}}_{\text{Part B}} \cdot \underbrace{G_i G_j \cdot (G^W)^{-1} \cdot \mu_{ij}^{+}}_{\text{Part C}}.
\end{aligned} \tag{1.22}$$

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 (1.22), 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 (1.22), 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). Further, we denote

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

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:

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

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 (1.22) can be rewritten as an implicit funtion form:

$$\mu^*_{ij} = (\overline{\mu}^{i*}_{.j})^{\lambda_d} (\overline{\mu}^{j*}_{i.})^{\lambda_o} (\overline{\mu}^{ij*}_{..})^{\lambda_w} \cdot \tilde{\alpha}_j(\boldsymbol{\mu}^*) \cdot \tilde{\eta}_i(\boldsymbol{\mu}^*) \cdot \mu^+_{ij}. \tag{1.23}$$

The superscript "$*$" in the equation above denotes the optimal flow. Since all the components in equation (1.23) are positive, we can have the following log-transformed vector notation:

$$\ln \boldsymbol{\mu}^* = \mathbf{A} \ln \boldsymbol{\mu}^* + \tilde{\mathbf{x}}(\boldsymbol{\mu}^*), \tag{1.24}$$

where $\mathbf{A} = \lambda_d(I_n \otimes W) + \lambda_o(W \otimes I_n) + \lambda_w(W \otimes W)$, and $\tilde{\mathbf{x}}(\boldsymbol{\mu}^*)$ is an $N \times 1$ vector having $\tilde{x}_{ij}(\boldsymbol{\mu}^*) = \ln(\tilde{\alpha}_j(\boldsymbol{\mu}^*) \cdot \tilde{\eta}_i(\boldsymbol{\mu}^*) \cdot \mu^+_{ij})$ 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 $\tilde{\mathbf{x}}(\boldsymbol{\mu}^*)$. If $\rho_{\text{spec}}(\mathbf{A}) < 1$, we have a unique solution to equation (1.24): $\ln \boldsymbol{\mu}^* = \mathbf{S}^{-1}\tilde{\mathbf{x}}(\boldsymbol{\mu}^*)$

where $\mathbf{S} = I_N - \mathbf{A}$. Then,

$$
\begin{aligned}
\mu_{ij}^* &= \prod_{k=1}^{n}\prod_{l=1}^{n} \exp\left(s_{ij,kl}\tilde{x}_{kl}(\boldsymbol{\mu}^*)\right) \\
&= \exp\left(\sum_{k=1}^{n}\sum_{l=1}^{n} s_{ij,kl}\tilde{x}_{kl}(\boldsymbol{\mu}^*)\right) \\
&= \exp\left(\sum_{k=1}^{n}\sum_{l=1}^{n} s_{ij,kl}\left(\sum_{m=1}^{K}\beta_m \ln(D_{kl,m}) + \alpha_l(\boldsymbol{\mu}^*) + \eta_k(\boldsymbol{\mu}^*)\right)\right)
\end{aligned}
\tag{1.25}
$$

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 (1.25) as the conditional expectation of y_{ij}.

If representation (1.25) 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 verify the uniqueness of $\boldsymbol{\mu}^*$ for counterfactual analysis. 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 (1.25) can be rewritten as

$$
\mu_{ij}^* = \mu_{ij}^{++} \cdot \left(\prod_{k=1}^{n}\prod_{l=1}^{n} \tilde{\alpha}_l^{s_{ij,kl}}(\boldsymbol{\mu}^*)\right) \cdot \left(\prod_{k=1}^{n}\prod_{l=1}^{n} \tilde{\eta}_k^{s_{ij,kl}}(\boldsymbol{\mu}^*)\right), \text{ for } i,j = 1, \cdots, n, \tag{1.26}
$$

where

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

and

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

Consequently, equation (1.26) can be simplified as the following additive form:

$$
\begin{aligned}
\ln(\mu_{ij}^*) =& \ln(\mu_{ij}^{++}) + \sum_{k=1}^{n}\sum_{l=1}^{n} s_{ij,kl}\left(\ln(\tilde{\alpha}_l(\boldsymbol{\mu}^*)) + \ln(\tilde{\eta}_k(\boldsymbol{\mu}^*))\right) \\
\Leftrightarrow \ln(\boldsymbol{\mu}^*) =& \Psi\left(\boldsymbol{\mu}^*, \boldsymbol{\mu}^{++}, \boldsymbol{\mu}^{+}\right),
\end{aligned}
\tag{1.27}
$$

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.

Step 3: partner selection in Stage 1. See the main draft.

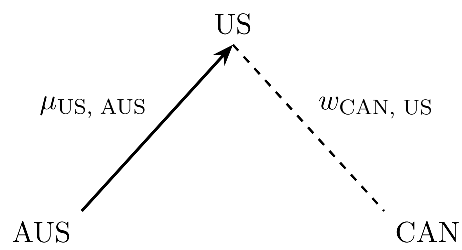
Interpretations. Each w_{ij} captures the strength of proximity (connectivity) between i and j. For intuition, consider a nearest-neighbor specification where $w_{ij} = 1$ if j is the nearest neighbor of i and $w_{ij} = 0$ otherwise. Under this extreme case,

- $\overline{\mu}^{i}_{\cdot j} = \mu_{kj}$ 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.

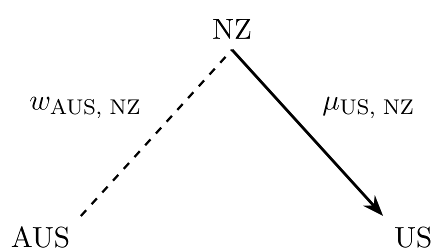
Figure 1: Common origin + Cross-destination linkage



   - 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



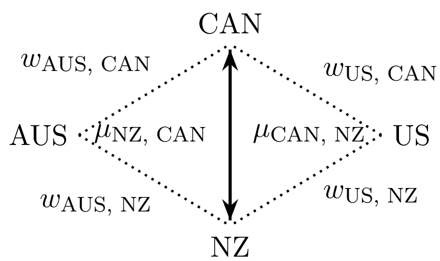
   - If $\lambda_o > 0$, $\mu_{\text{US,AUS}}$ and $\mu_{\text{US,NZ}}$ are positively associated. AUS and NZ coordinate their exports to the US through joint scheduling, consolidation, or hub sharing, lowering costs.

[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.

- 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 $\mu_{\text{NZ, CAN}}$).

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↔NZ) support AUS → 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 → US.

The nearest-neighbor example clarifies intuition, but in practice, W is row-normalized based on historical trade flows. Then $\overline{\mu}^i_{\cdot j} = \prod_k \mu_{kj}^{w_{ik}}$, $\overline{\mu}^j_{i\cdot} = \prod_l \mu_{il}^{w_{jl}}$, and $\overline{\mu}^{ij}_{\cdot\cdot} = \prod_{k,l} \mu_{kl}^{w_{ik}w_{jl}}$ 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, $\overline{\mu}^i_{\cdot j}$, $\overline{\mu}^j_{i\cdot}$ and $\overline{\mu}^{ij}_{\cdot\cdot}$ are pair-specific characteristics instead unit-specific ones.

Now taking the natural logarithm on (1.23), we obtain

$$
\begin{aligned}
\ln(\mu_{ij}) = &- \ln(G^W) + \lambda_d \ln(\overline{\mu}^i_{.j}) + \lambda_o \ln(\overline{\mu}^j_{i.}) + \lambda_w \ln(\overline{\mu}^{ij}_{..}) \\
&+ \ln(G_i) + \ln(G_j) + (\varrho - 1)\ln(\Pi_i) + (\varrho - 1)\ln(P_j) \\
&+ \sum_{k=1}^{K} \beta_k \ln(D_{ij,k}) + \sum_{l=1}^{L} \gamma_{l,o} \ln(E_{j,l}) + \sum_{l=1}^{L} \gamma_{l,d} \ln(E_{i,l}).
\end{aligned} \tag{1.28}
$$

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 (1.28) 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 $F_j(v) := \Pr[\vartheta_j(\omega) \leq v] = \exp[-A_j v^{-b}]$ for $v > 0$.

Let $w_j > 0$ be country j's wage. Let $\tau_{ij} \geq 1$ 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

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

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 τ_{ij} follows the conventional iceberg assumption. Krugman (1995) points out an advantage of this iceberg specification since this assumption implies:

$$
\frac{p_{ij}(\omega)}{p_{ij}(\omega')} = \frac{\frac{w_j}{\vartheta_j(\omega)}\tau_{ij}}{\frac{w_j}{\vartheta_j(\omega')}\tau_{ij}} = \frac{\vartheta_j(\omega')}{\vartheta_j(\omega)} \text{ for } \omega \neq \omega'. \tag{1.29}
$$

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, $\tau_{ii} = 1$ for all i, while $\tau_{ij} > 1$ 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, $\tau_{ij} \leq \tau_{ik} \cdot \tau_{kj}$ 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]

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

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:

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

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 in country j) does not depend on the destination. Since our framework does not specify a product-specific trade cost, our framework also satisfies (1.29). As an extension, if we specify a trade cost as a function of product-specific factors (i.e., $\tau_{ij}(\boldsymbol{\mu}, \omega) = \tau^{e}_{ij}(\boldsymbol{\mu}, \omega) \cdot \tau^{+}_{ij}$), (1.29) 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

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

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

Also, we define

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

Then, the CDF of p_{ij} is

$$\begin{aligned}
G_{ij}(p) &= \Pr[p_{ij}(\omega) \leq p] \\
&= \Pr\left[\frac{w_j \tau_{ij}}{\vartheta_j(\omega)} \leq p\right] \\
&= \Pr\left[\vartheta_j(\omega) \geq \frac{w_j \tau_{ij}}{p}\right] \\
&= 1 - F_j\left(\frac{w_j \tau_{ij}}{p}\right) \\
&= 1 - \exp\left(-A_j\left(\frac{w_j \tau_{ij}}{p}\right)^{-b}\right)
\end{aligned} \tag{1.30}$$

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

$$\begin{aligned}
G_i(p) &= \Pr[p_i(\omega) \leq p] \\
&= \Pr\left[\min_{j=1,\cdots,n} p_{ij}(\omega) \leq p\right] \\
&= 1 - \Pr\left[\min_{j=1,\cdots,n} p_{ij}(\omega) > p\right] \\
&= 1 - \Pr\left[\{p_{i1}(\omega) > p\} \cap \{p_{i2}(\omega) > p\} \cap \cdots \cap \{p_{in}(\omega) > p\}\right] \\
&= 1 - \prod_{j=1}^{n} \Pr[p_{ij}(\omega) > p] \\
&= 1 - \prod_{j=1}^{n} (1 - \Pr[p_{ij}(\omega) \leq p]) \\
&= 1 - \prod_{j=1}^{n} \exp\left(-A_j\left(\frac{w_j \tau_{ij}}{p}\right)^{-b}\right) \quad \text{by (1.30)} \\
&= 1 - \exp\left(-\sum_{j=1}^{n} A_j\left(\frac{w_j \tau_{ij}}{p}\right)^{-b}\right).
\end{aligned} \tag{1.31}$$

Equation (1.31) 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 (1.31) 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 (1.31) since τ_{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., $\tau_{ij}(\boldsymbol{\mu})$ 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 (1.30) and (1.31), 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?

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

$$\Pr\left[\vartheta_{i1}(\omega) \leq v_1, \cdots, \vartheta_{in}(\omega) \leq v_n\right] = \exp\left[-\left(\sum_{j=1}^{n}\left(A_{ij}v_j^{-b}\right)^{\frac{1}{1-\varpi}}\right)^{1-\varpi}\right]. \tag{1.32}$$

Here,

- A_{ij} is the scale parameter showing absolute advantage of countries;
- $b > 0$ is the shape parameter (leading to $\Pr\left[\vartheta_{ij}(\omega) \leq v\right] = \exp\left(-A_{ij}v^{-b}\right)$); 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, (1.32) is extended from a univariate Fréchet distribution, i.e.,

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

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 $\exp\left[-\ln(u_1, \cdots, u_n)\right]$ 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

$$C(u_1, \cdots, u_n) = \Pr\left[U_1 \leq u_1, \cdots, U_n \leq u_n\right],$$

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

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

where $F_i(x) = \Pr[X_i \leq x]$ 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**.

Observe

$$\begin{aligned}
s_{ij} = & \underbrace{\Pr[p_{ij}(\omega) < \min_{k\neq j} p_{ik}(\omega)]}_{\text{probability that country j's price to i is the lowest one}} \\
= & \int_0^\infty \underbrace{\int_0^\infty \mathbb{I}\{p_{ij}(\omega) < \min_{k\neq j} p_{ik}(\omega)\} dG^*_{ij}(p')}_{=\Pr[\min_{k\neq j} p_{ik}(\omega) > p] \text{ when } p_{ij}(\omega)=p} dG_{ij}(p) \text{ by the definition} \\
= & \int_0^\infty \Pr[\min_{k\neq j} p_{ik}(\omega)] dG_{ij}(p) \\
= & \int_0^\infty \Pr\left[\cap_{k\in\{1,\cdots,n\}\setminus\{j\}} \{p_{ik}(\omega) > p\}\right] dG_{ij}(p) \\
= & \int_0^\infty \left(\prod_{k\in\{1,\cdots,n\}\setminus\{j\}} (1 - G_{ik}(p))\right) dG_{ij}(p)
\end{aligned} \tag{1.33}$$

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

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

and

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

since $\frac{d}{dp}\left(1 - \exp\left(-A_j \left(\frac{w_j\tau_{ij}}{p}\right)^{-b}\right)\right) = -\exp\left(-A_j \left(\frac{w_j\tau_{ij}}{p}\right)^{-b}\right) \cdot -bp^{b-1}A_j(w_j\tau_{ij})^{-b}$. From (1.33), we have

$$\begin{aligned}
s_{ij} = & \int_0^\infty \left(\prod_{k\in\{1,\cdots,n\}\setminus\{j\}} (1 - G_{ik}(p))\right) dG_{ij}(p) \\
= & \int_0^\infty \exp\left(-\sum_{j=1}^n A_j \left(\frac{w_j\tau_{ij}}{p}\right)^{-b}\right) \cdot bp^{b-1} A_j (w_j\tau_{ij})^{-b} dp \\
= & A_j (w_j\tau_{ij})^{-b} \cdot \int_0^\infty bp^{b-1} \cdot \exp\left(-p^b \Upsilon_i\right) dp \\
= & \frac{A_j (w_j\tau_{ij})^{-b}}{\Upsilon_i}
\end{aligned} \tag{1.34}$$

where $\Upsilon_i := \sum_{j=1}^n A_j (w_j\tau_{ij})^{-b}$. 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,

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

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

$$\mu_{ij} = G_i \times s_{ij} = G_i \times \frac{A_j(w_j\tau_{ij})^{-b}}{\sum_{k=1}^{n} A_k(w_k\tau_{ik})^{-b}} = \underbrace{\frac{G_i}{\sum_{k=1}^{n} A_k(w_k\tau_{ik})^{-b}}}_{\text{country } i\text{-specific factor}} \times \underbrace{A_j w_j^{-b}}_{\text{country } j\text{-specific factor}} \times \tau_{ij}^{-b}, \tag{1.35}$$

which is the Eaton and Kortum (2002) gravity equation. In contrast to Anderson and van Wincoop (2003), it is not possible to endogenize τ_{ij} in the same way since equation (1.35) 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:

$$\ln y_{ij} = \lambda_d \sum_{k=1}^{n} w_{ik} \ln y_{kj} + \lambda_o \sum_{l=1}^{n} w_{jl} \ln y_{il} + \lambda_w \sum_{k=1}^{n}\sum_{l=1}^{n} w_{ik}w_{jl} \ln y_{kl} + x_{ij}'\beta + v_{ij}, \tag{1.36}$$

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 constructed from a symmetric matrix $\widetilde{W} = (\widetilde{w}_{ij})$. That is, $W = \text{Diag}^{\text{sum}}(\widetilde{W})^{-1}\widetilde{W}$ with $\text{Diag}^{\text{sum}}(\widetilde{W}) = \text{diag}\left(\sum_{j=1}^{n} \widetilde{w}_{1j}, \cdots, \sum_{j=1}^{n} \widetilde{w}_{nj}\right)$. Assume $\sum_{j=1}^{n} \widetilde{w}_{ij} > 0$ for all $i = 1, \cdots, n$, so that $\text{Diag}^{\text{sum}}(\widetilde{W})^{\pm 1/2}$ is well-defined. Define another symmetrically normalized matrix

$$\widetilde{\widetilde{W}} \equiv \text{Diag}^{\text{sum}}(\widetilde{W})^{-1/2}\widetilde{W}\text{Diag}^{\text{sum}}(\widetilde{W})^{-1/2} = \text{Diag}^{\text{sum}}(\widetilde{W})^{1/2}W\text{Diag}^{\text{sum}}(\widetilde{W})^{-1/2}. \tag{1.37}$$

Since $\widetilde{\widetilde{W}}$ is symmetric, by the spectral theorem, there exists an orthogonal matrix $\widetilde{Q}$ and a real diagonal matrix $D = \text{diag}(\varphi_1, \ldots, \varphi_n)$ such that $\widetilde{\widetilde{W}} = \widetilde{Q}D\widetilde{Q}'$. Also, note that $W = \text{Diag}^{\text{sum}}(\widetilde{W})^{-1/2}\widetilde{\widetilde{W}}\text{Diag}^{\text{sum}}(\widetilde{W})^{1/2}$ by (1.37). Hence,

$$W = \text{Diag}^{\text{sum}}(\widetilde{W})^{-1/2}\widetilde{\widetilde{W}}\text{Diag}^{\text{sum}}(\widetilde{W})^{1/2} = \text{Diag}^{\text{sum}}(\widetilde{W})^{-1/2}\widetilde{Q}D\widetilde{Q}'\text{Diag}^{\text{sum}}(\widetilde{W})^{1/2} = QDQ^{-1}$$

by letting $Q \equiv \text{Diag}^{\text{sum}}(\widetilde{W})^{-1/2}\widetilde{Q}$, so that $Q^{-1} = \widetilde{Q}'\text{Diag}^{\text{sum}}(\widetilde{W})^{1/2}$. In particular, W is diagonalizable with real eigenvalues $\varphi_1, \ldots, \varphi_n$, and these are exactly the eigenvalues of the symmetric matrix $\widetilde{\widetilde{W}}$.

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

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

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 (2.1) 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 $j \in \mathcal{V}$; $w_{ij} = 0$, otherwise. Define $z = (z_1, \cdots, z_n)'$,

[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 $\sum_{j=1}^n w_{ij} = 1$.

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 (2.1). 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]

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

$$\rho_{\text{spec}}(\mathbf{A}) = \max\{b(1,1), b(1, \varphi_{\min}), b(\varphi_{\min}, 1), b(\varphi_{\min}, \varphi_{\min})\} < 1, \tag{1.38}$$

[7] To intuitively explain the two extreme cases, consider the structure of Wy in equation (2.1). 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 n is large.

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,

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

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

$$
\begin{aligned}
s_{ij,kl} &= \left(e_{n,j}' \otimes e_{n,i}'\right) \left(I_N + \sum_{r=1}^{\infty} \left(\lambda_d (I_n \otimes W) + \lambda_o (W \otimes I_n) + \lambda_w (W \otimes W)\right)^r \right) (e_{n,l} \otimes e_{n,k}) \\
&= \mathbb{I}(j=l, i=k) + \sum_{p=1}^{\infty} \left(e_{n,j}' \otimes e_{n,i}'\right) \mathbf{A}^r \left(e_{n,l} \otimes e_{n,k}\right) \\
&= \mathbb{I}(j=l, i=k) + \sum_{r=1}^{\infty} \sum_{r_1+r_2+r_3=r} \frac{r!}{r_1! r_2! r_3!} \lambda_d^{r_1} \lambda_o^{r_2} \lambda_w^{r_3} \left(W^{r_1+r_3}\right)_{ik} \left(W^{r_2+r_3}\right)_{jl}.
\end{aligned}
$$

Then, the r-th order effect contains (i) the (r_1+r_3)-th order connections between k and i via $W^{r_1+r_3}$ and (ii) (r_2+r_3)-th order connections between j and l by $W^{r_2+r_3}$ such that $r_1+r_2+r_3=r$.

Put differently, note that

$$s_{ij,kl} = \mathbb{I}(i = k, j = l) + \sum_{r=1}^{\infty} s^{(r)}_{ij\leftarrow kl},$$

where $s^{(r)}_{ij,kl}$ $(r = 1, 2, \cdots)$ indicates the rth-order effects of $s_{ij,kl}$. In general, the rth-order effects decompose $s_{ij,kl}$ by r-step paths. For illustration purposes, we demonstrate the first-order effects (i.e., $r = 1$) and the second-order effects (i.e., $r = 2$). The first-order effects are specified as

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

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.

One may observe that when $r = 2$, the signal with two pairs ($kl \mapsto ij$) is decomposed as

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

In other words, $s^{(2)}_{ij,kl} = \sum_{p=1}^{n} \sum_{q=1}^{n} (\mathbf{A})_{ij,pq} (\mathbf{A})_{pq,kl}$, which means $s^{(2)}_{ij,kl}$ represents the effect from kl to ij through pq (i.e., $kl \mapsto pq \mapsto ij$). The representation above shows that there are six possible channels.

- $\lambda_d^2 \mathbb{I}(j = l)(W^2)_{ik}$: This shows $kj \mapsto ij$ since $l = j$ (same origin). Hence, this term consists of the second-order effect from the destination k in the origin pair kl to the destination i in the destination pair ij. That is, $(w^2)_{ik} = \sum_{p=1}^{n} w_{ip} w_{pk}$ illustrates $k \mapsto p \mapsto i$ for $p = 1, \cdots, n$.

- $\lambda_o^2 (W^2)_{jl} \mathbb{I}(i = k)$: This term characterizes the force $il \mapsto ij$ (same destination). It consists of the effect from the origin l in the origin pair il to another origin j in the destination pair ij ($l \mapsto q \mapsto j$ for $q = 1, \cdots, n$).

- $\lambda_w^2 (W^2)_{jl} (W^2)_{ik}$: This term consists of the second-order effect from origin l to another origin j ($l \mapsto q \mapsto j$ for $q = 1, \cdots, n$), and those from destination k to another destination i ($k \mapsto p \mapsto i$ for $p = 1, \cdots, n$).

- $2\lambda_d \lambda_o w_{jl} w_{ik}$: This second-order effect is characterized by two first-order effects: $l \mapsto j$ (origin l to another origin j) and $k \to i$ (destination k to another destination i). Indeed, this effect is a combination of $kj \mapsto ij$ and $il \mapsto ij$.

- $2\lambda_d \lambda_w w_{jl} (W^2)_{ik}$: This channel is a combination of $kj \mapsto ij$ and $kl \mapsto ij$. Hence, the resulting term consists of $l \mapsto j$ (first-order effect) and $k \mapsto p \mapsto i$ (second-order effect) for $p = 1, \cdots, n$.

- $2\lambda_o \lambda_w (W^2)_{jl} w_{ik}$: This term is generated by a combination of $il \mapsto ij$ and $kl \mapsto ij$. The resulting term consists of $k \mapsto i$ (first-order effect) and $l \mapsto q \mapsto j$ (second-order effect).

Using the same way, the third-order effect can be represented by

$$s_{ij,kl}^{(3)} = \sum_{p_1=1}^{n} \sum_{q_1=1}^{n} \sum_{p_2=1}^{n} \sum_{q_2=1}^{n} (\mathbf{A})_{ij,p_1 q_1} (\mathbf{A})_{p_1 q_1, p_2 q_2} (\mathbf{A})_{p_2 q_2, kl}.$$

This representation illustrates the chain $kl \mapsto p_2 q_2 \mapsto p_1 q_1 \mapsto ij$ for $p_1, p_2, q_1, q_2 = 1, \cdots, n$. Then, there are 10 channels. In general, there are $\binom{r+2}{r}$ channels for the r-th order effect ($r = 1, 2, \cdots$). At each order r, each term corresponds to a nonnegative integer triple (r_1, r_2, r_3) with $r_1 + r_2 + r_3 = r$ hence there are $\binom{r+2}{r}$ channels.

2 Theoretical details in statistical analysis

2.1 First- and second-order conditions

Recall that the statistical objective function is:

$$\ell_N(\theta, \boldsymbol{\phi}) = \sum_{i,j=1}^{n} \left(-\mu_{ij}(\theta, \boldsymbol{\phi}) + y_{ij} \ln \left(\mu_{ij}(\theta, \boldsymbol{\phi}) \right) - \ln \left(y_{ij}! \right) \right) - \frac{1}{2} \left(\sum_{j=1}^{n} \alpha_j - \sum_{i=1}^{n} \eta_i \right)^2, \tag{2.1}$$

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

$$\xi_{ij}(\theta, \boldsymbol{\phi}) = \frac{y_{ij}}{\mu_{ij}(\theta, \boldsymbol{\phi})} : \text{ multiplicative residual evaluated at } (\theta, \boldsymbol{\phi}),$$

$$u_{ij}(\theta, \boldsymbol{\phi}) = \mu_{ij}(\theta, \boldsymbol{\phi})\left(\xi_{ij}(\theta, \boldsymbol{\phi}) - 1\right) = y_{ij} - \mu_{ij}(\theta, \boldsymbol{\phi}) : \text{ additive residual at } (\theta, \boldsymbol{\phi}), \text{ and}$$

$$z_{ij}\left(\beta, \eta_i, \alpha_j\right) = x'_{ij}\beta + \alpha_j + \eta_i : \text{ exogenous component evaluated at } \left(\beta, \eta_i, \alpha_j\right).$$

For notational convenience, we further define $\boldsymbol{\theta} = (\theta', \boldsymbol{\phi}')'$, $\mathbf{W}_d = I_n \otimes W$, $\mathbf{W}_o = W \otimes I_n$, 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 (2.1) 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 $\partial_{\boldsymbol{\theta\theta}}\ell_N(\boldsymbol{\theta})$ has the following block diagonal structure:

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

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, $\partial_{\phi\phi}\ell_N(\boldsymbol{\theta})$:

$$\partial_{\alpha_l\alpha_l}\ell_N(\boldsymbol{\theta}) = -\sum_{i,j=1}^{n}\left(\sum_{k=1}^{n}\sum_{p=1}^{n} s_{ij,kl}(\lambda)s_{ij,pl}(\lambda)\right)\mu_{ij}(\boldsymbol{\theta}) - 1,$$

$$\partial_{\alpha_l\alpha_s}\ell_N(\boldsymbol{\theta}) = -\sum_{i,j=1}^{n}\left(\sum_{k=1}^{n}\sum_{p=1}^{n} s_{ij,kl}(\lambda)s_{ij,ps}(\lambda)\right)\mu_{ij}(\boldsymbol{\theta}) - 1 \text{ if } l \neq s,$$

$$\partial_{\alpha_l\eta_k}\ell_N(\boldsymbol{\theta}) = -\sum_{i,j=1}^{n}\left(\sum_{k=1}^{n}\sum_{q=1}^{n} s_{ij,kl}(\lambda)s_{ij,kq}(\lambda)\right)\mu_{ij}(\boldsymbol{\theta}) + 1,$$

$$\partial_{\eta_k\eta_k}\ell_N(\boldsymbol{\theta}) = -\sum_{i,j=1}^{n}\left(\sum_{l=1}^{n}\sum_{q=1}^{n} s_{ij,kl}(\lambda)s_{ij,kq}(\lambda)\right)\mu_{ij}(\boldsymbol{\theta}) - 1, \text{ and}$$

$$\partial_{\eta_k\eta_t}\ell_N(\boldsymbol{\theta}) = -\sum_{i,j=1}^{n}\left(\sum_{l=1}^{n}\sum_{q=1}^{n} s_{ij,kl}(\lambda)s_{ij,tq}(\lambda)\right)\mu_{ij}(\boldsymbol{\theta}) - 1 \text{ if } k \neq t.$$

To have a vector/matrix notation, we define

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

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,

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

where $\mathbf{G}(\boldsymbol{\theta}) = [\mathbf{W}_d\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta}), \mathbf{W}_o\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta}), \mathbf{W}_w\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta}), \mathbf{X}]$, 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,

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

where

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

Last, note that

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

where

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

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

In sum,

$$\partial_{\boldsymbol{\theta\theta}} \ell_N(\boldsymbol{\theta}) = \mathbf{H}^A(\boldsymbol{\theta}) + \mathbf{H}^B(\boldsymbol{\theta}), \tag{2.2}$$

where

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

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 $i \in \{1, \cdots, n\}$ is located in $\mathcal{D}_n \subset \mathcal{D}$, 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 2.2. We posit that W is constructed by row–normalizing a symmetric base matrix $\widetilde{W}$ (e.g., geographic/logistical affinity), $W = \text{Diag}^{\text{sum}}(\widetilde{W})^{-1}\widetilde{W}$, allowing W itself to be asymmetric after normalization.

Assumption 2.3. (i) For each ij, we assume

$$\tau_{ij}^{+} = D_{ij,1}^{\tilde{\beta}_1} \cdots D_{ij,K}^{\tilde{\beta}_K},$$

where $D_{ij,k}$ $(k = 1, \cdots, K)$ represents a bilateral characteristic affecting τ_{ij}. $\tilde{\beta}_1, \cdots, \tilde{\beta}_K$ are parameters. We assume that the baseline cost τ_{ij}^{+} satisfies the triangle inequality: for arbitrary three countries i, j, and k, $\tau_{ij}^{+} \leq \tau_{ik}^{+} \cdot \tau_{kj}^{+}$.

(ii) If i chooses $k \in \{1, \cdots, n\} \setminus \{i\}$ with probability w_{ik} and j selects $l \in \{1, \cdots, n\} \setminus \{j\}$ with probability w_{jl} as partners (hubs), the trade cost from j to i through k and l is

$$\widetilde{\tau}_{ij}\left(\boldsymbol{\mu}; k, l\right) = \mu_{kj}^{-\tilde{\lambda}_d} \mu_{il}^{-\tilde{\lambda}_o} \mu_{kl}^{-\tilde{\lambda}_w} \cdot \tau_{ij}^{+},$$

where $\tilde{\lambda}_d$, $\tilde{\lambda}_o$ and $\tilde{\lambda}_w$ are coefficients and $\boldsymbol{\mu} = (\mu_{11}, \mu_{21}, \cdots, \mu_{n1}, \cdots, \mu_{1n}, \mu_{2n}, \cdots, \mu_{nn})'$.

(iii) i's and j's partner choices are independent, so the probability of using the route (k, l) is $w_{ik} w_{jl}$.

(iv) Then, the overall trade cost from j to i is defined as

$$\tau_{ij}(\boldsymbol{\mu}) = \exp\left(\mathbb{E}_W\left[\ln(\widetilde{\tau}_{ij}\left(\boldsymbol{\mu}; k, l\right))\right]\right), \text{ where } \mathbb{E}_W(\cdot) = \sum_{k,l=1}^{n} w_{ik} w_{jl}(\cdot).$$

Assumption 2.4. (i) We assume

$$\max\{\lambda_d + \lambda_o + \lambda_w, \lambda_d\varphi_{\min} + \lambda_o + \lambda_w\varphi_{\min}, \lambda_d + \lambda_o\varphi_{\min} + \lambda_w\varphi_{\min}, \lambda_d\varphi_{\min} + \lambda_o\varphi_{\min} + \lambda_w\varphi_{\min}^2\} < 1, \tag{2.3}$$

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,

$$\begin{aligned} \alpha_j(\boldsymbol{\mu}) &= -\frac{1}{2}\ln(G^W) + \ln\left(G_j\right) + \ln\left(\Pi_j^{\varrho-1}(\boldsymbol{\mu})\right) \text{ for } j = 1, \cdots, n \text{ and} \\ \eta_i(\boldsymbol{\mu}) &= -\frac{1}{2}\ln(G^W) + \ln\left(G_i\right) + \ln\left(P_i^{\varrho-1}(\boldsymbol{\mu})\right), \text{ for } i = 1, \cdots, n, \end{aligned} \tag{2.4}$$

where $\Pi_j(\boldsymbol{\mu}) = \left(\sum_{i=1}^n \frac{G_r}{G^W}\left(\frac{\tau_{ij}(\boldsymbol{\mu})}{P_i(\boldsymbol{\mu})}\right)^{1-\varrho}\right)^{\frac{1}{1-\varrho}}$, $P_i(\boldsymbol{\mu}) = \left(\sum_{j=1}^n \frac{G_j}{G^W}\left(\frac{\tau_{ij}(\boldsymbol{\mu})}{\Pi_j(\boldsymbol{\mu})}\right)^{1-\varrho}\right)^{\frac{1}{1-\varrho}}$, and $G^W = \sum_{i=1}^n G_i$.

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 $\sup_n \sup_{\lambda\in\Lambda} \|\mathbf{A}(\lambda)\|_\infty < 1$.

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, $\boldsymbol{\Phi} \subset [-C, C]^{2n}$ for some finite constant $C > 0$.

(i) For each $(\theta, \boldsymbol{\phi}) \in \boldsymbol{\Theta}$, define $\mathbf{J}_N^{\phi\phi}(\boldsymbol{\theta}) = \frac{1}{N}\Big(\mathbf{D}'\mathbf{S}^{-1\prime}(\lambda)\operatorname{Diag}\big(\boldsymbol{\mu}(\boldsymbol{\theta})\big)\mathbf{S}^{-1}(\lambda)\mathbf{D} - \mathbf{H}^{\phi\phi}\Big)$. Assume $\liminf_{n\to\infty} \inf_{\theta\in\Theta_\theta} \inf_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} \varphi_{\min}\Big(\mathbf{J}_N^{\phi\phi}(\theta, \boldsymbol{\phi})\Big) > 0$. Then, for each $\theta \in \Theta_\theta$ and for n sufficiently large, $\widehat{\boldsymbol{\phi}}(\theta) = \arg\max_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} \ell_N(\theta, \boldsymbol{\phi})$ is unique.

(ii) For each $(\theta, \boldsymbol{\phi}) \in \boldsymbol{\Theta}$, define

$$\mathbf{J}_N^{\theta\theta}(\boldsymbol{\theta}) = \frac{1}{N}\Big(\mathbf{G}(\boldsymbol{\theta})'\mathbf{S}^{-1\prime}(\lambda)\operatorname{Diag}\big(\boldsymbol{\mu}(\boldsymbol{\theta})\big)\mathbf{S}^{-1}(\lambda)\mathbf{G}(\boldsymbol{\theta}) - \mathbf{H}^{\theta\theta}(\boldsymbol{\theta})\Big),$$
$$\mathbf{J}_N^{\theta\phi}(\boldsymbol{\theta}) = \frac{1}{N}\Big(\mathbf{G}(\boldsymbol{\theta})'\mathbf{S}^{-1\prime}(\lambda)\operatorname{Diag}\big(\boldsymbol{\mu}(\boldsymbol{\theta})\big)\mathbf{S}^{-1}(\lambda)\mathbf{D} - \mathbf{H}^{\phi\theta}(\boldsymbol{\theta})'\Big), \text{ and } \mathbf{J}_N^{\phi\theta}(\boldsymbol{\theta}) = \big(\mathbf{J}_N^{\theta\phi}(\boldsymbol{\theta})\big)'.$$

Here, $\mathbf{G}(\boldsymbol{\theta}) = \Big[\mathbf{W}_d\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta}), \mathbf{W}_o\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta}), \mathbf{W}_w\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta}), \mathbf{X}\Big]$. For each $\theta \in \Theta_\theta$, let

$$\widehat{\mathbf{J}}_N^{\theta\theta}(\theta) = \mathbf{J}_N^{\theta\theta}(\theta, \widehat{\boldsymbol{\phi}}(\theta)), \quad \widehat{\mathbf{J}}_N^{\theta\phi}(\theta) = \mathbf{J}_N^{\theta\phi}(\theta, \widehat{\boldsymbol{\phi}}(\theta)), \quad \widehat{\mathbf{J}}_N^{\phi\phi}(\theta) = \mathbf{J}_N^{\phi\phi}(\theta, \widehat{\boldsymbol{\phi}}(\theta)).$$

Assume $\liminf_{n\to\infty} \inf_{\theta\in\Theta_\theta} \varphi_{\min}\Big(\widehat{\mathbf{H}}(\theta)\Big) > 0$, where $\widehat{\mathbf{H}}(\theta) = \widehat{\mathbf{J}}_N^{\theta\theta}(\theta) - \widehat{\mathbf{J}}_N^{\theta\phi}(\theta)\Big[\widehat{\mathbf{J}}_N^{\phi\phi}(\theta)\Big]^{-1}\widehat{\mathbf{J}}_N^{\phi\theta}(\theta)$.

Then, for n sufficiently large, $\widehat{\theta} = \arg\max_{\theta\in\Theta_\theta} \ell_N^c(\theta)$ is unique.

Assumption 2.7. (i) $\{x_{ij}\}$, $\{\eta_i^0\}$, 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) $\{\xi_{ij}\}$ is a random field satisfying $\sup_{i,j,n} \mathbb{E}|\xi_{ij}|^{2+c} < C$ for some $c > 0$.

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

Lemma 2.1. For each ij, we define the additive error, $u_{ij} = \mu_{ij}^0(\xi_{ij} - 1)$, to have $u_{ij} = y_{ij} - \mu_{ij}^0$. Under Assumption 2.7, we obtain $\mathbb{E}(u_{ij}|\mathbf{x}) = 0$ and $\sup_{i,j,n} \mathbb{E}|u_{ij}|^{2+c} < C$.

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 ij for y_{ij} is located in the product space $\mathcal{D} \times \mathcal{D} \subset \mathbb{R}^{2d}$. 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:

$$d^p\left(l^p(ij), l^p(kl)\right) = \max\{d(l(i), l(k)), d(l(j), l(l))\}. \tag{2.5}$$

For notational simplicity, we denote $d^p_{ij,kl} = d^p\left(l^p(ij), l^p(kl)\right)$ for two pairs ij and ik in $\mathcal{D} \times \mathbf{D}$, and $d_{ij} = d(l(i), l(j))$ for i and j in $\mathcal{D}$. The distance between pairs in (2.5) is measured by the larger distance between the origins and the destinations. Using this device, we want to control $\mathsf{Cov}(y_{ij}, y_{kl})$: $\mathsf{Cov}(y_{ij}, y_{kl}) \to 0$ 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 $\mathsf{Cov}(y_{ij}, y_{kj}) \to 0$ as $d_{ik} \to \infty$. Note that this metric specification is intended solely for simple asymptotic analysis, not for practical use. Assumption 2.7 describes the properties of the components in $\{x_{ij}\}$, $\{\eta^0_i\}$ and $\{\alpha^0_j\}$, 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 $\mathbb{E}\left(u_{ij}|\mathbf{x}\right) = \mathbb{E}\left(\mu^0_{ij}(\xi_{ij}-1)|\mathbf{x}\right) = \mu^0_{ij}\left(\mathbb{E}(\xi_{ij}|\mathbf{x}) - 1\right) = 0$.

Second, by Assumptions 2.5, 2.6 and 2.7 (i),

$$\tilde{\mu}^0_{ij} = \sum_{k,l=1}^{n} s_{ij,kl}(x'_{kl}\beta^0 + \alpha^0_l + \eta^0_k) \leq \|\mathbf{S}^{-1}\|_\infty \cdot \sup_{i,j,n} |x_{ij}\beta^0 + \alpha^0_j + \eta^0_i| < \infty.$$

This implies $\mu^0_{ij} = \exp\left(\tilde{\mu}^0_{ij}\right)$ is uniformly bounded, i.e., $\sup_{i,j,n}|\mu^0_{ij}| \leq C$. It implies $|\mu^0_{ij}(\xi_{ij} - 1)|^p \leq C^p \cdot |\xi_{ij}-1|^p$ 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.■

The lemma below shows the NED properties of $\{y_{ij}\}$.

Lemma 2.2. Assume Assumptions 2.1, 2.6, and 2.7 hold.

(i) We have uniform L_p-boundedness of $\{y_{ij}\}$. That is, $\sup_{n,i,j}\|y_{ij}\|_{L_{2+c}} < \infty$.

(ii) Let $\mathcal{Y} = \{y_{ij} : ij \in \mathcal{D}_n \times \mathcal{D}_n, n \geq 1\}$ and $\Xi = \{(x_{ij}, \xi_{ij}) : ij \in \mathcal{D}_n \times \mathcal{D}_n, n \geq 1\}$. Assume

Ξ is an α-mixing random field with spatial α-mixing coefficient $\alpha(u,v,r) \leq (u+v)^{\tau}\widehat{\alpha}(r)$ for some $\tau \geq 0$ and for some $0 < \tilde{\eta} < 2 + \frac{\eta}{2}$, $\widehat{\alpha}(r)$ satisfies $\sum_{r=1}^{\infty} r^{2d(\tau_*+1)-1}\widehat{\alpha}(r)^{\frac{\tilde{\eta}}{4+2\tilde{\eta}}} < \infty$. In addition, we assume $0 \leq w_{ij} \leq C \cdot d_{ij}^{-a}$ for some $C > 0$ and $a > 2d$.

Then, $\mathcal{Y}$ is uniformly L_2-NED on Ξ. That is,

$$\|y_{ij} - \mathbb{E}\left(y_{ij}|\mathcal{F}_{ij}(s)\right)\|_{L_2} \leq C \cdot s^{2d-a} \text{ for some } C > 0.$$

Here, $\mathcal{F}_{ij}(s) = \sigma\left(x_{kl}, \xi_{kl} : d_{ij,kl}^{p} \leq s\right)$ for $s \geq 0$.

Proof of Lemma 2.2 (i) We need to show $\sup_{i,j,n} \|\mu_{ij}^0 \cdot \xi_{ij}\|_{L_{2+c}} < \infty$. In the proof of Lemma 2.1, we already have $\sup_{i,j,n} |\mu_{ij}^0| < \infty$. Hence,

$$\sup_{i,j,n} \|\mu_{ij}^0 \cdot \xi_{ij}\|_{L_{2+c}} \leq \left(\sup_{i,j,n} |\mu_{ij}^0|\right) \cdot \sup_{i,j,n} \|\xi_{ij}\|_{L_{2+c}} < \infty$$

by Assumption 2.7 (ii).

(ii) For this, we will proceed with the following steps:

Step 1: As a first step, we will show $\{\tilde{\mu}_{ij}^0\}$ is uniformly L_2-NED on Ξ. Note that $\tilde{\mu}_{ij}^0$ is generated by $\{x_{kl}, \xi_{kl}\}_{k,l=1}^n$ (indeed, $\{\xi_{kl}\}_{k,l=1}^n$ does not play a role here). Consider two possible bases $\{\dot{x}_{kl}, \dot{\xi}_{kl}\}_{k,l=1}^n$ and $\{\ddot{x}_{kl}, \ddot{\xi}_{kl}\}_{k,l=1}^n$. Then, the difference between the two resulting $\tilde{\mu}_{ij}^0$ is:

$$\begin{aligned}
&\tilde{\mu}_{ij}^0\left(\{\dot{x}_{kl}\}_{k,l=1}^n\right) - \tilde{\mu}_{ij}^0\left(\{\ddot{x}_{kl}\}_{k,l=1}^n\right) \\
&= \sum_{k,l=1}^{n} s_{ij,kl}\left(\sum_{m=1}^{K} \beta_m^0\left(\dot{x}_{kl} - \ddot{x}_{kl}\right) + \left(\alpha_l^0(\{\dot{x}_{kl}\}_{k,l=1}^n) - \alpha_l^0(\{\ddot{x}_{kl}\}_{k,l=1}^n)\right) + \left(\eta_k^0(\{\dot{x}_{kl}\}_{k,l=1}^n) - \eta_k^0(\{\ddot{x}_{kl}\}_{k,l=1}^n)\right)\right).
\end{aligned} \tag{2.6}$$

Here, for example, $\alpha_l^0(\{\dot{x}_{kl}\}_{k,l=1}^n)$ denotes the fixed effect component α_l^0 generated by $\{\dot{x}_{kl}\}_{k,l=1}^n$. To characterize an upper bound of (2.6), for any kl observe that

$$\begin{aligned}
s_{ij,kl} &\leq \overline{s}_{ij,kl}, \\
\dot{x}_{kl} - \ddot{x}_{kl} &\leq 2\sup_{i,j,n}\max_{m=1,\cdots,K}|x_{ij,m}| \\
\alpha_l^0(\{\dot{x}_{kl}\}_{k,l=1}^n) - \alpha_l^0(\{\ddot{x}_{kl}\}_{k,l=1}^n) &\leq 2\sup_{j,n}|\alpha_j^0|, \text{ and} \\
\eta_k^0(\{\dot{x}_{kl}\}_{k,l=1}^n) - \eta_k^0(\{\ddot{x}_{kl}\}_{k,l=1}^n) &\leq 2\sup_{i,n}|\eta_i^0|,
\end{aligned}$$

where $\overline{s}_{ij,kl}$ denotes the $((j-1)n+i,(l-1)n+k)$-element of $(I_N - |\mathbf{A}|)^{-1}$. Here, the $((j-1)n+i,(l-1)n+k)$-element of $|\mathbf{A}|$ is $|\lambda_d^0\mathbb{I}(j=l)w_{ik}+\lambda_o^0\mathbb{I}(i=k)w_{jl}+\lambda_w^0 w_{ik}w_{jl}|$.

Using (2.6), we then measure the difference between $\tilde{\mu}_{ij}^0$ and $\mathbb{E}\left(\tilde{\mu}_{ij}^0|\mathcal{F}_{ij}(s)\right)$ for $s>0$. For this, note that $\mathbb{E}\left(\tilde{\mu}_{ij}^0|\mathcal{F}_{ij}(s)\right)$ is an approximation using x_{kl} when $d_{ij,kl}^p \le s$. Then, for a given $s>0$,

$$\begin{aligned}
&\|\tilde{\mu}_{ij}^0-\mathbb{E}\left(\tilde{\mu}_{ij}^0|\mathcal{F}_{ij}(s)\right)\|_{L_2} \\
\le& 2\left(K\cdot\sup_{i,j,n}\max_{m=1,\cdots,K}|x_{ij,m}|\cdot\max_{m=1,\cdots,K}|\beta_m^0|+\sup_{j,n}|\alpha_j^0|+\sup_{i,n}|\eta_i^0|\right)\cdot\left(\sum_{k,l:d_{ij,kl}^p>s}\overline{s}_{ij,kl}\right).
\end{aligned} \tag{2.7}$$

By Assumption 2.7 (i), $\sup_{i,j,n}|x_{i,j,m}|<\infty$ for all $m=1,\cdots,K$, $\sup_{j,n}|\alpha_j^0|<\infty$ and $\sup_{i,n}|\eta_i^0|<\infty$. From (2.7), hence, it suffices to examine $\sum_{k,l:d_{ij,kl}^p>s}\overline{s}_{ij,kl}$. Under the setting in Lemma 2.2 (ii), $\sum_{k,l:d_{ij,kl}^p>s}\overline{s}_{ij,kl}\le C\cdot s^{2d-a}$ for some $C>0$ by Lemma B.1 in Jeong and Lee (2024). Hence, we have $\|\tilde{\mu}_{ij}^0-\mathbb{E}\left(\tilde{\mu}_{ij}^0|\mathcal{F}_{ij}(s)\right)\|_{L_2}\le C\cdot s^{2d-a}$ for some $C>0$.

Step 2: Second, we will show $\{\mu_{ij}^0\}$ (note: $\mu_{ij}^0=\exp\left(\tilde{\mu}_{ij}^0\right)$) is uniformly L_2-NED on Ξ. Observe that

$$\begin{aligned}
&\left|\mu_{ij}^0\left(\{\dot{x}_{kl}\}_{k,l=1}^n\right)-\mu_{ij}^0\left(\{\ddot{x}_{kl}\}_{k,l=1}^n\right)\right| \\
=&\left|\exp\left(\tilde{\mu}_{ij}^0\left(\{\dot{x}_{kl}\}_{k,l=1}^n\right)\right)-\exp\left(\tilde{\mu}_{ij}^0\left(\{\ddot{x}_{kl}\}_{k,l=1}^n\right)\right)\right| \\
\le&\max\{\exp\left(\tilde{\mu}_{ij}^0\left(\{\dot{x}_{kl}\}_{k,l=1}^n\right)\right),\exp\left(\tilde{\mu}_{ij}^0\left(\{\ddot{x}_{kl}\}_{k,l=1}^n\right)\right)\}\cdot\left|\tilde{\mu}_{ij}^0\left(\{\dot{x}_{kl}\}_{k,l=1}^n\right)-\tilde{\mu}_{ij}^0\left(\{\ddot{x}_{kl}\}_{k,l=1}^n\right)\right|
\end{aligned}$$

by the mean value theorem. Even though $\exp(\cdot)$ is not a Lipschitz function, we can apply Proposition 2 in Jenish and Prucha (2012) since $\max\{\exp\left(\tilde{\mu}_{ij}^0\left(\{\dot{x}_{kl}\}_{k,l=1}^n\right)\right),\exp\left(\tilde{\mu}_{ij}^0\left(\{\ddot{x}_{kl}\}_{k,l=1}^n\right)\right)\}<\infty$ (local Lipschitz). Then, we have $\|\mu_{ij}^0-\mathbb{E}\left(\mu_{ij}^0|\mathcal{F}_{ij}(s)\right)\|_{L_2}\le C\cdot s^{2d-a}$ for some $C>0$.

Step 3: Last, we want to show $\{y_{ij}\}$ (note: $y_{ij}=\mu_{ij}^0\cdot\xi_{ij}$) is uniformly L_2-NED on Ξ. By the Cauchy-Schwarz inequality[8], we have

$$\begin{aligned}
\|y_{ij}-\mathbb{E}\left(y_{ij}|\mathcal{F}_{ij}(s)\right)\|_{L_2}&=\|\xi_{ij}\left(\mu_{ij}^0-\mathbb{E}\left(\mu_{ij}^0|\mathcal{F}_{ij}(s)\right)\right)\|_{L_2} \\
&\le\|\xi_{ij}\|_{L_2}\cdot\|\mu_{ij}^0-\mathbb{E}\left(\mu_{ij}^0|\mathcal{F}_{ij}(s)\right)\|_{L_2}\le C\cdot s^{2d-a}
\end{aligned}$$

for some $C>0$. The first equality above holds since $\{\omega\in\Omega:\omega=\xi_{ij}^{-1}(z)\text{ for }z\in\text{Range}(\xi_{ij}(\cdot))\}\in\mathcal{F}_{ij}(s)$ for any $s>0$.■

[8]For random variables X and Y, $\|XY\|_{L_2}\le\|X\|_{L_1}^{1/2}\cdot\|Y\|_{L_1}^{1/2}=\left(\int X^2dP\right)^{1/2}\cdot\left(\int Y^2dP\right)^{1/2}=\|X\|_{L_2}\cdot\|Y\|_{L_2}$.

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

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

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

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

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

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

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[9]

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

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

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

Lemma 2.3. We obtain the two results:

(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

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

estimated. The form of $A - BC^{-1}B'$ is

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

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

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

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:

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

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

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

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

where

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

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

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

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

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

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 $\mathbf{M_D} = I_N - \mathbf{P_D}\text{Diag}(\boldsymbol{\mu})$. Then,

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

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

$$\mathsf{Var}\left(\left(\begin{matrix}\frac{1}{\sqrt{N}}\left(\mathbf{S}^{-1}\mathbf{G}\right)'\mathbf{u} \\ \frac{1}{\sqrt{n}}\left(\mathbf{S}^{-1}\mathbf{D}\right)'\mathbf{u}\end{matrix}\right)\middle|\,\mathbf{x}\right) = \begin{bmatrix}\frac{1}{N}\mathbf{G}'\mathbf{S}^{-1\prime}\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{S}^{-1}\mathbf{G} & \frac{1}{n^{\frac{3}{2}}}\mathbf{G}'\mathbf{S}^{-1\prime}\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{S}^{-1}\mathbf{D} \\ \frac{1}{n^{\frac{3}{2}}}\mathbf{D}'\mathbf{S}^{-1\prime}\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{S}^{-1}\mathbf{G} & \frac{1}{n}\mathbf{D}'\mathbf{S}^{-1\prime}\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{S}^{-1}\mathbf{D}\end{bmatrix}.$$

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}\text{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}\text{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{G} \\
&+\frac{1}{n\sqrt{n}}\mathbf{G}'\mathbf{S}^{-1\prime}\text{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}\text{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{G} \\
=&\frac{1}{N}\mathbf{G}'\mathbf{S}^{-1\prime}\left(\begin{matrix}\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right) - \text{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}\text{Diag}(\boldsymbol{\mu}) \\ +\text{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}\text{Diag}(\boldsymbol{\mu})\end{matrix}\right) \\
&\times\mathbf{S}^{-1}\mathbf{G} \\
=&\frac{1}{N}\mathbf{G}'\mathbf{S}^{-1\prime}\left(\left(I_N - \mathbf{P_D}\text{Diag}(\boldsymbol{\mu})\right)'\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\left(I_N - \mathbf{P_D}\text{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:

$$
\begin{aligned}
&\left(\frac{1}{n}\widetilde{\mathbf{D}'\mathbf{D}}\right)^{-1} \\
&\times\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} \\
&\times\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}.
\end{aligned}
$$

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) $\frac{\dim(\phi_{2n})}{\sqrt{N}} = \frac{2n}{n} = 2 > 0$ 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.

trict concavity. Lemma 2.4 illustrates the conditions for strict concavity of $\ell_N(\boldsymbol{\theta})$.

Lemma 2.4. From (2.2), recall that $\partial_{\boldsymbol{\theta}\boldsymbol{\theta}}\ell_N(\boldsymbol{\theta}) = -\mathbf{H}^A(\boldsymbol{\theta}) - \mathbf{H}^B(\boldsymbol{\theta})$, where

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

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 $\text{Diag}(\boldsymbol{\mu}(\boldsymbol{\theta}))$ is a diagonal matrix with strictly positive elements for any $\boldsymbol{\theta} \in \tilde{\boldsymbol{\Theta}}$. By Assumption 2.5, $\mathbf{S}(\lambda)$ 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, $\hat{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\ell_N(\boldsymbol{\theta})$. 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 $\ell_\infty(\boldsymbol{\theta}) \equiv \text{plim}_{n\to\infty}\frac{1}{N}\ell_N(\boldsymbol{\theta})$ 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 $\text{liminf}_{n\to\infty} \inf_{\boldsymbol{\phi}\in\boldsymbol{\Theta}_{\boldsymbol{\phi}}} \varphi_{\min}\left(\frac{1}{n}\mathbf{D}'\mathbf{S}^{-1\prime}(\lambda)\text{Diag}\left(\boldsymbol{\mu}(\boldsymbol{\theta})\right)\mathbf{S}^{-1}(\lambda)\mathbf{D} - \frac{1}{n}\mathbf{H}^{\boldsymbol{\phi}\boldsymbol{\phi}}\right) > 0$ for each $\theta \in \Theta_\theta$. Then, $\hat{\boldsymbol{\phi}}(\theta) = \text{argmax}_{\boldsymbol{\phi}\in\boldsymbol{\Theta}_{\boldsymbol{\phi}}}\ell_N(\theta, \boldsymbol{\phi})$ is unique for each $\theta \in \Theta_\theta$ and for a sufficiently large n.

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

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

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

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

Proof of Lemma 2.5. (i) Fix $\theta \in \Theta_\theta$ and consider $\text{argmax}_{\boldsymbol{\phi}\in\boldsymbol{\Theta}_{\boldsymbol{\phi}}}\ell_N(\theta, \boldsymbol{\phi})$. 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 $\max_{\boldsymbol{\phi}\in\boldsymbol{\Theta}_{\boldsymbol{\phi}}} \ell_N(\theta, \boldsymbol{\phi})$. 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 $\boldsymbol{\phi} \in \boldsymbol{\Theta}_{\boldsymbol{\phi}}$. Since $\frac{1}{n}\partial_{\boldsymbol{\phi}\boldsymbol{\phi}}\ell_N(\theta, \boldsymbol{\phi}) = -\frac{1}{n}\mathbf{D}'\mathbf{S}^{-1\prime}(\lambda)\text{Diag}\left(\boldsymbol{\mu}(\boldsymbol{\theta})\right)\mathbf{S}^{-1}(\lambda)\mathbf{D}+\frac{1}{n}\mathbf{H}^{\boldsymbol{\phi}\boldsymbol{\phi}}$ and $-\frac{1}{n}\mathbf{D}'\mathbf{S}^{-1\prime}(\lambda)\text{Diag}\left(\boldsymbol{\mu}(\boldsymbol{\theta})\right)\mathbf{S}^{-1}(\lambda)\mathbf{D}+\frac{1}{n}\mathbf{H}^{\boldsymbol{\phi}\boldsymbol{\phi}} = O(1)$, 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 $\hat{\theta} = \text{argmax}_{\theta\in\Theta_\theta} \ell_N^c(\theta)$. Note that $\hat{\theta}$ satisfies

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

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 $\partial_{\theta\theta}\ell_N^c(\theta) < 0$ for all $\theta \in \Theta_\theta$. Observe that

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

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

Define the scaled log-likelihood as

$$
\widetilde{\ell}_N(\boldsymbol{\theta}) \equiv \frac{1}{N}\,\ell_N(\boldsymbol{\theta}) \text{ and } \ell_\infty(\boldsymbol{\theta}) \equiv \text{plim}_{n\to\infty}\,\widetilde{\ell}_N(\boldsymbol{\theta}) \text{ for } \boldsymbol{\theta} \in \boldsymbol{\Theta}
$$

whenever the limit exists. We say that $\boldsymbol{\theta}^0$ is *identified in large samples* if $\ell_\infty(\boldsymbol{\theta}) < \ell_\infty(\boldsymbol{\theta}^0)$ for all $\boldsymbol{\theta} \neq \boldsymbol{\theta}^0$ in $\boldsymbol{\Theta}$.

Lemma 2.6 (Large-sample identification)**.** Suppose Assumptions 2.1–2.5, 2.6, and 2.7 hold. Then:

(i) For each $\theta \in \Theta_\theta$, there exists a unique $\boldsymbol{\phi}(\theta) = \arg\max_{\boldsymbol{\phi}\in\boldsymbol{\Phi}}\,\ell_\infty(\theta, \boldsymbol{\phi})$.

(ii) The profiled criterion $\ell_\infty^c(\theta) \equiv \ell_\infty\Big(\theta, \boldsymbol{\phi}(\theta)\Big)$ has a unique maximizer $\theta^0 = \arg\max_{\theta\in\Theta_\theta}\,\ell_\infty^c(\theta)$, and we define $\boldsymbol{\phi}^0 = \boldsymbol{\phi}(\theta^0)$.

Hence $(\theta^0, \boldsymbol{\phi}^0)$ is identified in large samples in the sense that

$$\ell_\infty(\theta, \boldsymbol{\phi}) < \ell_\infty(\theta^0, \boldsymbol{\phi}^0) \text{ for all } (\theta, \boldsymbol{\phi}) \neq (\theta^0, \boldsymbol{\phi}^0) \text{ in } \boldsymbol{\Theta}.$$

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 $\theta \in \Theta_\theta$, recall that

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

Observe that the dimension of $\hat{\boldsymbol{\phi}}(\theta)$ is $2n$, growing with increasing n. Then, we need to evaluate the magnitudes of a $2n$- dimensional vector (e.g., $\hat{\boldsymbol{\phi}}(\theta) - \boldsymbol{\phi}^0$), a $2n \times 2n$ matrix (e.g., $-\frac{1}{n}\partial_{\boldsymbol{\phi}\boldsymbol{\phi}}\ell_N$), a $2n \times 2n \times 2n$ tensor (e.g., $\frac{1}{n}\partial_{\boldsymbol{\phi}\boldsymbol{\phi}\boldsymbol{\phi}}\ell_N$). For this, we utilize the (induced) q-norm $\|\cdot\|_q$ for $2 \leq q \leq \infty$.[10] Here are examples for this measure (details can be found in Fernandez-Val and Weidner (2016)):

- For an n-dimensional vector $a = (a_1, \cdots, a_n)'$, $\|a\|_q = (\sum_{i=1}^n |a_i|^q)^{\frac{1}{q}}$.
- For an $n \times n$ matrix $A = (a_{ij}) = (a_{.1}, \cdots a_{.n})$, $\|A\|_q = \max_{\{x \in \mathbb{R}^n : \|x\|_q = 1\}} \|Ax\|_q = \max_{\{x \in \mathbb{R}^n : \|x\|_q = 1\}} \|\sum_{i=1}^n x_i \cdot a_{.i}\|_q$. Note that the row-vector representation $\|A'\|_q = \max_{\{x \in \mathbb{R}^n : \|x\|_q = 1\}} \|A'x\|_q$ is also possible and generally $\|A\|_q \neq \|A'\|_q$. In detail, $\|A\|_q = \|A'\|_q$ only if $q = 2$ or A is symmetric. Since we focus on evaluating symmetric matrices, we do not need to have separate definitions.
- Consider an $n \times n \times n$ tensor $A = (a_{ijk})$ and consider i as the focal index. Then, A can be interpreted as a bilinear map

  $$A : (x, y) \mapsto z = (z_1, \cdots, z_n)' \text{ for } x, y \in \mathbb{R}^n$$

  such that $z_i = \sum_{j=1}^n \sum_{k=1}^n a_{i,jk} x_j y_k$. Then, the induced q-norm of A (by the first index) is

  $$\|A\|_q = \|A\|_{q,(1)} = \max_{\{x, y \in \mathbb{R}^n : \|x\|_q = 1, \|y\|_q = 1\}} \left\{ \left(\sum_{j=1}^n \sum_{k=1}^n a_{1,jk} x_j y_k, \cdots, \sum_{j=1}^n \sum_{k=1}^n a_{n,jk} x_j y_k \right)' \right\}.$$

  In general, index ordering matters, as in the case of the matrix q-norm. Since we focus on fully symmetric tensors across indices, treating the first index as fixed is reasonable.

[10] For finite-dimensional vector/matrix/tensor (e.g., $\frac{1}{\sqrt{N}}\partial_\theta \ell_N$), on the other hand, the Euclidean norm $\|\cdot\|$ is employed.

Using the q-norm, we obtain the following results.

Lemma 2.7. $\mathbb{E}\left(-\frac{1}{n}\partial_{\phi\phi}\ell_N\right) > 0$ and $\left\|\mathbb{E}\left(-\frac{1}{n}\partial_{\phi\phi}\ell_N\right)^{-1}\right\|_q = O_p(1)$.

Lemma 2.8. Suppose $q > 4$. Under the regularity conditions we have, the following relations hold.

(i-1) $\|\frac{1}{n}\partial_{\phi}\ell_N\|_q = O_p\left(n^{-\frac{1}{2}+\frac{1}{q}}\right)$ and $\|\frac{1}{\sqrt{N}}\partial_{\theta}\ell_N\| = O_p\left(1\right)$.
(i-2) $\left\|-\frac{1}{n}\partial_{\phi\phi}\ell_N - \mathbb{E}\left(-\frac{1}{n}\partial_{\phi\phi}\ell_N\right)\right\|_q = o_p(1)$.
(i-3) $\left\|\frac{1}{\sqrt{N}}\partial_{\theta\phi}\ell_N\right\|_q = O_p\left(n^{\frac{1}{q}}\right)$.
(i-4) $\left\|\frac{1}{\sqrt{N}}\partial_{\theta\theta}\ell_N\right\|_q = O_p\left(\sqrt{N}\right)$.
(ii-1) $\left\|-\frac{1}{n}\partial_{\phi\phi}\ell_N - \mathbb{E}\left(-\frac{1}{n}\partial_{\phi\phi}\ell_N\right)\right\| = o_p\left(n^{-\frac{1}{4}}\right)$.
(ii-2) $\left\|\frac{1}{\sqrt{N}}\partial_{\theta\theta}\ell_N - \mathbb{E}\left(\frac{1}{\sqrt{N}}\partial_{\theta\theta}\ell_N\right)\right\| = o_p\left(\sqrt{N}\right)$.
(ii-3) $\left\|\frac{1}{\sqrt{N}}\partial_{\theta\phi\phi}\ell_N - \mathbb{E}\left(\frac{1}{\sqrt{N}}\partial_{\theta\phi\phi}\ell_N\right)\right\| = o_p\left(n^{-\frac{1}{4}}\right)$.
(ii-4)

[To be written]

Let $r_\theta > 0$ and $r_\phi > 0$ with ...

Lemma 2.9. Assume $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

[Expansion]

$$\frac{1}{\sqrt{N}}\partial_{\theta}\ell_N\left(\theta, \widehat{\boldsymbol{\phi}}(\theta)\right) = \mathcal{U}^{(0)} + \mathcal{U}^{(1,a,1)} + \mathcal{U}^{(1,a,2)} + \mathcal{U}^{(1,b)} - \boldsymbol{\Sigma}_{\theta,N}\sqrt{N}\left(\theta - \theta^0\right) + \mathcal{R}(\theta),$$

where

$$\mathcal{U}^{(0)} = \frac{1}{\sqrt{N}}\partial_{\theta}\ell_N + \mathbb{E}\left(\frac{1}{\sqrt{N}}\partial_{\theta\phi}\ell_N\right)\cdot\mathbb{E}\left(-\frac{1}{\sqrt{N}}\partial_{\phi\phi}\ell_N\right)^{-1}\cdot\frac{1}{\sqrt{N}}\partial_{\phi\theta}\ell_N,$$

$$\mathcal{U}^{(1,a,1)} = \left\{\frac{1}{\sqrt{N}}\partial_{\theta\phi}\ell_N - \mathbb{E}\left(\frac{1}{\sqrt{N}}\partial_{\theta\phi}\ell_N\right)\right\}\cdot\mathbb{E}\left(-\frac{1}{\sqrt{N}}\partial_{\phi\phi}\ell_N\right)^{-1}\cdot\frac{1}{\sqrt{N}}\partial_{\phi}\ell_N,$$

$$\begin{aligned}\mathcal{U}^{(1,a,2)} = &-\mathbb{E}\left(\frac{1}{\sqrt{N}}\partial_{\theta\phi}\ell_N\right)\cdot\mathbb{E}\left(-\frac{1}{\sqrt{N}}\partial_{\phi\phi}\ell_N\right)^{-1}\cdot\left\{-\frac{1}{\sqrt{N}}\partial_{\phi\phi}\ell_N - \mathbb{E}\left(-\frac{1}{\sqrt{N}}\partial_{\phi\phi}\ell_N\right)\right\}\\ &\cdot\mathbb{E}\left(-\frac{1}{\sqrt{N}}\partial_{\phi\phi}\ell_N\right)^{-1}\cdot\frac{1}{\sqrt{N}}\partial_{\phi}\ell_N,\end{aligned}$$

and

$$\mathcal{U}^{(1,b)} = \frac{1}{2}\sum_{g=1}^{2n}\left(\mathbb{E}\left(\frac{1}{\sqrt{N}}\partial_{\theta\phi\phi_g}\ell_N\right) + \mathbb{E}\left(\frac{1}{\sqrt{N}}\partial_{\theta\phi}\ell_N\right)\cdot\mathbb{E}\left(-\frac{1}{\sqrt{N}}\partial_{\phi\phi}\ell_N\right)^{-1}\cdot\mathbb{E}\left(\frac{1}{\sqrt{N}}\partial_{\phi\phi\phi_g}\ell_N\right)\right)$$
$$\cdot\left[\mathbb{E}\left(-\frac{1}{\sqrt{N}}\partial_{\phi\phi}\ell_N\right)^{-1}\cdot\frac{1}{\sqrt{N}}\partial_{\phi}\ell_N\right]_g\cdot\mathbb{E}\left(-\frac{1}{\sqrt{N}}\partial_{\phi\phi}\ell_N\right)^{-1}\cdot\frac{1}{\sqrt{N}}\partial_{\phi}\ell_N,$$

$\mathcal{R}(\theta)$ denotes the remainder term satisfying $\|\mathcal{R}(\theta)\| = o_p(1) + o_p(n\cdot\|\theta-\theta_0\|)$ for $\theta\in\mathcal{B}(\theta_0, r_\theta)$, ϕ_g is the gth-element of $\boldsymbol{\phi}$ and $\dim(\boldsymbol{\phi}) = 2n$.

For a given $\theta\in\mathcal{B}(\theta_0, r_\theta)$, the Taylor expansion of $\hat{\boldsymbol{\phi}}(\theta)$ around $\boldsymbol{\phi}^0$ is

$$\widehat{\boldsymbol{\phi}}(\theta) - \boldsymbol{\phi}^0 = \left(-\frac{1}{\sqrt{N}}\partial_{\phi\phi}\ell_N\right)^{-1}\cdot\frac{1}{\sqrt{N}}\partial_{\phi}\ell_N + \left(-\frac{1}{\sqrt{N}}\partial_{\phi\phi}\ell_N\right)^{-1}\frac{1}{\sqrt{N}}\partial_{\phi\theta}\ell_N\cdot(\theta-\theta_0)$$
$$+\frac{1}{2}\left(-\frac{1}{\sqrt{N}}\partial_{\phi\phi}\ell_N\right)^{-1}\cdot\sum_{j=1}^{n}\left\{u_j^{\alpha}\cdot\frac{1}{\sqrt{N}}\partial_{\phi\phi\alpha_j}\ell_N\cdot\left(-\frac{1}{\sqrt{N}}\partial_{\phi\phi}\ell_N\right)^{-1}\cdot\frac{1}{\sqrt{N}}\partial_{\phi}\ell_N\right\}$$
$$+\frac{1}{2}\left(-\frac{1}{\sqrt{N}}\partial_{\phi\phi}\ell_N\right)^{-1}\cdot\sum_{i=1}^{n}\left\{u_i^{\eta}\cdot\frac{1}{\sqrt{N}}\partial_{\phi\phi\eta_i}\ell_N\cdot\left(-\frac{1}{\sqrt{N}}\partial_{\phi\phi}\ell_N\right)^{-1}\cdot\frac{1}{\sqrt{N}}\partial_{\phi}\ell_N\right\} + \mathcal{R}^{\phi}(\theta),$$

where $u_{N,j}^{\alpha}$ is the jth element of $\overline{\mathcal{H}}_N^{\alpha\alpha}\frac{1}{\sqrt{N}}\partial_{\boldsymbol{\alpha}_n}\ell_N + \overline{\mathcal{H}}_N^{\alpha\eta}\frac{1}{\sqrt{N}}\partial_{\boldsymbol{\eta}_n}\ell_N$, $u_{N,i}^{\eta}$ denotes the ith element of $\overline{\mathcal{H}}_N^{\eta\alpha}\frac{1}{\sqrt{N}}\partial_{\boldsymbol{\alpha}_n}\ell_N + \overline{\mathcal{H}}_N^{\eta\eta}\frac{1}{\sqrt{N}}\partial_{\boldsymbol{\eta}_n}\ell_N$, $\mathcal{R}^{\phi}(\theta)$ denotes the remainder term. Note that $\|\mathcal{R}^{\phi}(\theta)\|_q = o_p\left(n^{-1+\frac{1}{q}}\right) + o_p\left(n^{\frac{1}{q}}\cdot\|\theta-\theta_0\|\right)$ for $\theta\in\mathcal{B}(\theta_0, r_\theta)$.

Another main target is $\{\mu_{ij}(\theta)\}$ for each $\theta\in\Theta_\theta$, where $\mu_{ij}(\theta) = \mu_{ij}\left(\theta, \widehat{\boldsymbol{\phi}}(\theta)\right)$. Note that $\hat{\boldsymbol{\phi}}(\theta) = (\hat{\boldsymbol{\alpha}}(\theta)', \hat{\boldsymbol{\eta}}(\theta)')' = (\widehat{\alpha}_1(\theta), \cdots, \widehat{\alpha}_n(\theta), \widehat{\eta}_1(\theta), \cdots, \widehat{\eta}_n(\theta))' = \operatorname{argmax}_{\boldsymbol{\phi}\in\boldsymbol{\Theta}_\phi}\ell_N(\theta, \boldsymbol{\phi})$ and

$$\mu_{ij}(\theta) = \exp\left(\tilde{\mu}_{ij}\left(\theta, \widehat{\boldsymbol{\phi}}(\theta)\right)\right) = \exp\left(\sum_{k,l=1}^{n}s_{ij,kl}(\lambda)\left(x_{kl}'\beta + \widehat{\alpha}_l(\theta) + \widehat{\eta}_k(\theta)\right)\right).$$

For each $\theta\in\Theta_\theta$, let $\tilde{\mu}_{ij}(\theta) = \sum_{k,l=1}^{n}s_{ij,kl}(\lambda)\left(x_{kl}'\beta + \widehat{\alpha}_l(\theta) + \widehat{\eta}_k(\theta)\right)$ to have $\tilde{\mu}_{ij}(\theta) = \tilde{\mu}_{ij}\left(\theta, \widehat{\boldsymbol{\phi}}(\theta)\right)$.

Lemma 2.10. Assume Assumptions 2.1, 2.5, 2.6, and 2.7 hold. Throughout the lemma, $\theta\in\Theta_\theta$ is arbitrarily chosen and fixed.

(i) We have uniform L_p-boundedness of $\{\mu_{ij}(\theta)\}$. That is, $\sup_{n,i,j}\|\mu_{ij}(\theta)\|_{L_{2+c}} < \infty$.

(ii) Let $\mathcal{M} = \{\mu_{ij}(\theta) : ij \in \mathcal{D}_n\times\mathcal{D}_n, n\geq 1\}$. Assume Ξ is an α-mixing random field with spatial α-mixing coefficient $\alpha(u,v,r)\leq(u+v)^{\tau}\widehat{\alpha}(r)$ for some $\tau\geq 0$ and for some $0<\tilde{\eta}<2+\frac{\eta}{2}$, $\widehat{\alpha}(r)$ satisfies $\sum_{r=1}^{\infty}r^{2d(\tau_*+1)-1}\widehat{\alpha}(r)^{\frac{\tilde{\eta}}{4+2\tilde{\eta}}} < \infty$. In addition, we assume $0\leq w_{ij}\leq C\cdot d_{ij}^{-a}$ for some $C>0$ and $a>2d$.

Then, $\mathcal{M}$ is uniformly L_2-NED on Ξ. That is,

$$\|\mu_{ij}(\theta) - \mathbb{E}\left(\mu_{ij}(\theta)|\mathcal{F}_{ij}(s)\right)\|_{L_2} \le C \cdot s^{2d-a} \text{ for some } C > 0.$$

Here, $\mathcal{F}_{ij}(s) = \sigma\left(x_{kl}, \xi_{kl} : d^p_{ij,kl} \le s\right)$ for $s \ge 0$.

Proof of Lemma 2.10 To prove Lemma 2.10, it suffices to show that $\{\tilde{\mu}_{ij}(\theta)\}$ is NED on Ξ. The remaining part can be proven as the proof of Lemma 2.2.

$$\begin{aligned}\tilde{\mu}_{ij}(\theta) - \mathbb{E}\left(\tilde{\mu}_{ij}(\theta)|\mathcal{F}_{ij}(s)\right) = &\sum_{k,l=1}^{n} s_{ij,kl}(\lambda) \sum_{m=1}^{K} \beta_m \left(x_{kl,m} - \mathbb{E}\left(x_{kl,m}|\mathcal{F}_{ij}(s)\right)\right) \\ &+ \sum_{k,l=1}^{n} s_{ij,kl}(\lambda) \left(\widehat{f}_{kl}(\theta) - \mathbb{E}\left(\widehat{f}_{kl}(\theta)|\mathcal{F}_{ij}(s)\right)\right).\end{aligned}$$

* Uniform convergence of the sample average of the log-likelihood function

Need to show: $\sup_{\theta\in\Theta_\theta} \left|\frac{1}{N}\ell^c_N(\theta) - \frac{1}{N}\mathbb{E}\left(\ell^c_N(\theta)\right)\right| \xrightarrow{p} 0$ as $n \to \infty$.

Let $\mu_{ij}(\theta) = \mu_{ij}\left(\theta, \widehat{\boldsymbol{\phi}}(\theta)\right)$ for each $\theta \in \Theta_\theta$ and for any ij. Note that

$$\begin{aligned}\frac{1}{N}\ell^c_N(\theta) - \frac{1}{N}\mathbb{E}\left(\ell^c_N(\theta)\right) = &-\frac{1}{N}\sum_{i,j=1}^{n}\left(\mu_{ij}(\theta) - \mathbb{E}\left(\mu_{ij}(\theta)\right)\right) \\ &+ \frac{1}{N}\sum_{i,j=1}^{n}\left(y_{ij}\ln\left(\mu_{ij}(\theta)\right) - \mathbb{E}\left(y_{ij}\ln\left(\mu_{ij}(\theta)\right)\right)\right) \\ &- \frac{1}{N}\sum_{i,j=1}^{n}\left(\ln(\Gamma(y_{ij}+1)) - \mathbb{E}\left(\ln(\Gamma(y_{ij}+1))\right)\right).\end{aligned}$$

Consider the first term above.

* Uniform equicontinuity in $\theta \in \Theta_\theta$

Need to show: Uniform equicontinuity of $\frac{1}{N}\mathbb{E}\left(\ell^c_N(\theta)\right)$ in $\theta \in \Theta_\theta$

$$\frac{1}{N}\mathbb{E}\left(\ell^c_N(\theta)\right) = -\frac{1}{N}\sum_{i,j=1}^{n}\mathbb{E}\left(\mu_{ij}(\theta)\right) + \frac{1}{N}\sum_{i,j=1}^{n}\mathbb{E}\left(y_{ij}\ln\left(\mu_{ij}(\theta)\right)\right) - \frac{1}{N}\sum_{i,j=1}^{n}\mathbb{E}\left(\ln\left(\Gamma(y_{ij}+1)\right)\right).$$

* Consistency of $\widehat{\theta}$

2.4 Variance estimation

The assumptions below are regularity conditions.

Assumption 2.8. (i) For the structure of $\mathbf{u} = (u_{11}, \cdots, u_{n1}, \cdots, u_{1n}, \cdots, u_{nn})'$, we assume

$$\mathbf{u} = \mathbf{BH}\boldsymbol{\epsilon}, \tag{2.8}$$

where $\mathbf{B}$ denotes some $N \times N$ matrix, $\mathbf{H} = \operatorname{diag}(\sigma_{11}^*, \cdots, \sigma_{n1}^*, \cdots, \sigma_{1n}^*, \cdots, \sigma_{nn}^*)$, 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) $0 < \inf_{i,j,n} \sigma_{ij}^* \leq \sup_{i,j,n} \sigma_{ij}^* < \infty$.

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

Assumption 2.9. (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 $\mathbf{x}$, $\nu_{ij,kl} = o(d_N)$, 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 $d_{ij,kl}^* \leq d_N$. Define $\deg_{ij}^* = \sum_{k,l=1}^n \mathbb{I}\{d_{ij,kl}^* \leq d_N\}$ and $\deg^* = \frac{1}{N} \sum_{i,j=1}^n \deg_{ij}^*$. Based on these definitions, we define

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

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

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

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

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

Assumption 2.10. (i) The kernel $\mathsf{K} : \mathbb{R} \to [-1, 1]$ such that $\mathsf{K}(0) = 1$, $\mathsf{K}(x) = \mathsf{K}(-x)$, $\mathsf{K}(x) = 0$ for $|x| > 1$.

(ii) $\mathsf{K}(\cdot)$ is Lipschitz, i.e., $|\mathsf{K}(x_1) - \mathsf{K}(x_2)| \leq C \cdot |x_1 - x_2|$ for some $C > 0$ and for $x_1, x_2 \in \mathbb{R}$.

(iii) $q \leq q_d$, where $q = \max\left\{\widetilde{q} : K_{\widetilde{q}} \equiv \lim_{x\to 0} \frac{1-\mathsf{K}(x)}{|x|^{\widetilde{q}}} < \infty, \widetilde{q} \in [0, \infty)\right\}$ is the Parzen characteristic exponent of $\mathsf{K}(\cdot)$.

(iv) 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 2.9 (i) characterizes an admissible type of dependence. It excludes the infill asymptotic. An example is $\|R_{ij}R'_{kl}\|\leq\frac{C}{\left(1+d_{ij,kl}\right)^{c+\Delta}}$ 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 2.9 (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 in the main draft, we evaluate the simulation results for possible distance measures for pairs. Under Assumption 2.9 (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 2.9 (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 2.9 (v), note that $\frac{1}{\sqrt{\mathbb{E}(\deg^*)}}\sum_{kl:d^*_{ij,kl}\leq d_N}\left(\mathbf{G}'\mathbf{S}^{-1\prime}\mathbf{M}'_{\mathbf{D}}\mathbf{u}\right)_{\cdot,kl}$ is a local average around ij, while $\frac{1}{\sqrt{N}}\sum_{k,l=1}^{n}\left(\mathbf{G}'\mathbf{S}^{-1\prime}\mathbf{M}'_{\mathbf{D}}\mathbf{u}\right)_{\cdot,kl}$ is the global average. If $ij\in\mathcal{E}$ (interior), the local average and the global average have the same asymptotic variance. Assumption 2.10 is conventional in spatial HAC literature (Kelejian and Prucha, 2007; Kim and Sun, 2011).[11]

Assumption on kernel functions

Here, q shows the smoothness of $\mathsf{K}(x)$ at $x=0$. When $\mathsf{K}(u)=1-|u|$ for $|u|\leq 1$ (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 $K_q=6$. If $\mathsf{K}(u)$ is the Tukey-Hanning kernel, $q=2$ and $K_q=\frac{\pi^2}{4}$. This quantity characterizes the bias of $\widetilde{\mathbf{\Omega}}_{\theta,N}$. In detail, since $\mathsf{K}\left(\frac{d^*_{ij,kl}}{d_N}\right)-1\simeq$

[11] In particular, Assumption 2.10 (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 $|u|\leq 1$ (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 $\overline{\mathsf{K}}=\frac{1}{60}$.

$-K_p\left(\frac{d^*_{ij,kl}}{d_N}\right)^q = -\frac{K_q}{d_N^q}\cdot(d^*_{ij,kl})^q$ around 0, we have

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

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 2.1. Assume that Assumptions 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7 and 2.8 hold for Theorems 3.1 and 3.2. Also, we suppose that Assumptions 2.8, 2.9, and 2.10 hold. Then, we have the following results:

(i) (Variance) $\lim_{n\to\infty}\frac{N}{\mathbb{E}(\deg^*)}\mathsf{Var}\left(\mathrm{vec}\left(\widetilde{\mathbf{\Omega}}_{\theta,N}\right)\right) = \overline{\mathsf{K}}(1+C)(\mathbf{\Omega}_\theta\otimes\mathbf{\Omega}_\theta)$, where C denotes the $(3+K)^2\times(3+K)^2$ commutation matrix[12];

(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 $\mathbf{\Omega}_\theta^{(q)} = \lim_{n\to\infty}\frac{1}{N}\sum_{i,j,k,l=1}^n R_{ij}R'_{kl}\cdot\mathbb{E}\left((d^*_{ij,kl})^q\right)$ for each q; 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 $\sqrt{\frac{N}{\mathbb{E}(\deg^*)}}\left(\widehat{\mathbf{\Omega}}_{\theta,N}-\widetilde{\mathbf{\Omega}}_{\theta,N}\right) = O_p(1)$.

First, Theorem 2.1 states consistency of $\widehat{\mathbf{\Omega}}_{\theta,N}$. When $\frac{\mathbb{E}(\deg^*)}{N}\to 0$, $\mathsf{Var}\left(\mathrm{vec}\left(\widetilde{\mathbf{\Omega}}_{\theta,N}\right)\right)\to 0$ by Theorem 2.1 (i).

3 Additional simulation analysis

[To be added]

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

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

4 Empirical Application

This section provides information about the network statistics employed in the application section. The contents in this section are based on Horn and Johnson (1985); Wasserman and Faust (1994); Chung (1997); Bramoullé et al. (2014).

4.1 Network Construction

4.1.1 Network candidate 1: Historical trade flows

See the main draft (Section 4).

4.1.2 Network candidate 2: Text-based construction of country–country networks

A growing literature demonstrates how unstructured text can be systematically incorporated into economic analysis using modern machine-learning tools (Gentzkow et al., 2019; Ke et al., 2019; Dugoua et al., 2022). These approaches transform written documents into quantitative measures that capture economically meaningful relationships. For example, Hoberg and Phillips (2016) analyze firms' own written descriptions of the products they sell in their annual regulatory reports to construct text-based similarity measures, which quantify how closely firms compete in product markets and allow industry boundaries and competitive relationships to change over time.

In a similar spirit, we construct a country–country connection matrix using a fully automated, text-based procedure that does not impose any *ex ante* structure on bilateral relationships. The only inputs to the construction are (i) a set of countries, (ii) a time window, and (iii) a publicly available text corpus with a deterministic inclusion rule.

Our primary text corpus is constructed from country-level Wikipedia articles accessed via the MediaWiki Action API (`https://en.wikipedia.org/w/api.php`). All texts are processed deterministically. For each year t, we collect the set of sentences mentioning country i, denoted by $\mathcal{T}_{i,t}$, and the set of sentences jointly mentioning country pair (i, j), denoted by $\mathcal{T}_{ij,t}$. Two pretrained and fixed text models are then applied: a semantic embedding model that maps text into a latent vector space, and a sentiment model that assigns a signed polarity score to text.

Specifically, we represent each country i by the average semantic embedding of sentences in $\mathcal{T}_{i,t}$, and measure semantic similarity between countries i and j using the cosine similarity of these embeddings. Separately, we compute the average sentiment score of co-mention sentences in $\mathcal{T}_{ij,t}$ to capture the positive or negative tone of bilateral discourse. The signed affinity between countries i and j is defined as the product of nonnegative semantic similarity and average co-mention sentiment, yielding positive weights for positive discourse and negative weights for negative discourse.

Note that all model parameters are pretrained on external data and remain fixed throughout the analysis. Semantic embeddings are computed using the `SentenceTransformer` class from the `sentence-transformers` Python library (model: `sentence-transformers/all-MiniLM-L6-v2`), which provides a frozen mapping from text to a latent semantic vector space. Sentiment scores are computed using the VADER sentiment analyzer (the `SentimentIntensityAnalyzer` class from the `vaderSentiment` Python library), which assigns a signed polarity score (the `compound` score in $[-1, 1]$) to each sentence without any task-specific retraining. To ensure that the final connection matrix is nonnegative, sentiment is incorporated as a fixed attenuation factor by mapping the polarity score $s_{ij,t} \in [-1, 1]$ into the interval $[0, 1]$ via the transformation $(1 + s_{ij,t})/2$, which downweights semantic similarity under negative discourse and upweights it under positive discourse without introducing negative edge weights. All models are applied in inference mode only, and no parameters are estimated or tuned using the study data. Consequently, the resulting network is a deterministic, nonnegative function of the country set and the corpus definition, rather than researcher-imposed judgments about bilateral relationships.

4.2 Network statistics

Degree statistics. First, we consider three degree statistics:

- Degree $\deg_i = \sum_{j=1}^{n} \mathbb{I}(w_{ij} > 0)$: The degree is computed from the support of the network. It captures how many partners each country i is meaningfully connected to. A higher $\overline{\deg} = \frac{1}{n} \sum_{i=1}^{n} \deg_i$ represents a denser or more diversified connection structure across countries. A lower variance of $\{\deg_i\}$ implies that W is close to the uniform connectivity. On the other hand, if its variance is high, it implies W has a core-periphery or centralized structure.

- High-intensity degree $\deg_i^+ = \sum_{j=1}^{n} \mathbb{I}(w_{ij} > w_{0.95})$: This high-intensity degree accounts for where the strongest trade relationships concentrate. If only a few countries have many top-5% links, the network might be hub-dominated (highly centralized). On the other hand, if many countries share comparable top-link degrees, trade intensity is more evenly distributed. Since $\mathsf{Var}\left(\deg^+\right)$ captures the dispersion in strong-link intensity, a high variance of $\deg^+$ implies a super-hub structure (only a few countries dominate the strongest trade links). On the other hand, if $\mathsf{Var}\left(\deg^{+}\right)$ is low, strong trade relationships are more evenly distributed across countries (less centralized connectivity network).

- $c_j = \sum_{i=1}^{n} w_{ij}$ (Column sum): To understand this, recall that w_{ij} illustrates the choice probability of j for country i. Then, $c_j = \sum_{i=1}^{n} w_{ij}$ shows the summation of the choice probability of j when every country chooses a partner. Hence, c_j captures the j's popularity/centrality.

Variations in networks. The second-type network statistics capture the variations in W. Further, these statistics are generated since W is also a row-stochastic matrix.

- Herfindahl–Hirschman index (HHI): For each country i, $\mathrm{HHI}_i = \sum_{j=1}^{n} w_{ij}^2$. To understand this index, consider the two extreme cases. First, if $w_{ik} = 1$ for some $k \in \{1, \cdots, n\} \setminus \{i\}$ and $w_{ij} = 0$ if $j \neq k$, $\mathrm{HHI}_i = 1$. Second, if $w_{ij} = \frac{1}{n-1}$ for all $j \neq i$, $\mathrm{HHI}_i = (n-1) \cdot \left(\frac{1}{n-1}\right)^2 = \frac{1}{n-1}$. Hence, (uniform) $\frac{1}{n-1} \leq \mathrm{HHI}_i \leq 1$ (concentrated).

- Effective number of partners I $n_i^{\mathrm{HHI}} = \frac{1}{\mathrm{HHI}_i}$: This is the first measure of effective number of partners. If $w_{ij} = \frac{1}{n-1}$ for all $j \neq i$ (uniform), $n_i^{\mathrm{HHI}} = n - 1$ (n partners are evenly distributed). On the other hand, if $w_{ik} = 1$ for some $k \in \{1, \cdots, n\} \setminus \{i\}$ and $w_{ij} = 0$ if $j \neq k$, $n_i^{\mathrm{HHI}} = 1$ (Indeed, there is only one partner).

- (Shannon) partner diversification entropy (PDE) $H_i = -\sum_{j=1}^{n} w_{ij} \ln(w_{ij})$: For country i, H_i isthe Shannon entropy of its partner-selection distribution. This measure captures the dispersion of partners employed by country i. A larger H_i indicates that i's partner choice is more evenly spread across many countries. On the other hand, a smaller H_i means concentration on a few partners. First, if $w_{ik} = 1$ for some $k \in \{1, \cdots, n\} \setminus \{i\}$ and $w_{ij} = 0$ if $j \neq k$, $H_i = 0$ (all mass on a single partner). Second, if $w_{ij} = \frac{1}{n-1}$ for all $j \neq i$, $H_i = \ln(n-1)$ (perfectly even across all $n-1$ partners). Hence, $0 \leq H_i \leq \ln(n-1)$.

- Normalized partner-diversification entropy $\widetilde{H}_i = \frac{H_i}{\ln(n-1)} \in [0, 1]$: Based on the properties of H_i, $\widetilde{H}_i$ is constructed as the normalized entropy. If $\widetilde{H}_i = 1$, i's partners are perfectly evenly distributed. On the other hand, $\widetilde{H}_i = 0$ means complete concentration on a single partner.

- Effective number of partners II $n_i^{\mathrm{E}} = \exp(H_i)$: This is a second measure for the effective number of partners. Intuitively, this measure means how many partners would I need to generate the same level of diversification as the current distribution if partner choice were perfectly even. First, if $w_{ik} = 1$ for some $k \in \{1, \cdots, n\} \setminus \{i\}$ and $w_{ij} = 0$ if $j \neq k$, $n_i^{\mathrm{E}} = 1$. Second, if $w_{ij} = \frac{1}{n-1}$ for all $j \neq i$, $n_i^{\mathrm{E}} = n - 1$.

- Kullback-Leibler (KL) divergence (Relative entropy or I-divergence) $D_i^{\mathrm{KL}}(w_{i.} \| \mathcal{U}) = \ln(n-1) - H_i$: This measure captures the statistical distance between $(w_{i1}, \cdots, w_{in})$ and uniform distribution. Here, the uniform distribution is the benchmark for full diversification: $u_{ij} = \frac{1}{n-1}$ for all $j \neq i$. Then, the Kullback-Leibler (KL) divergence of $(w_{i1}, \cdots, w_{in})$ from $\mathcal{U}$ is

$$D_i^{\mathrm{KL}}(w_{i.} \| \mathcal{U}) = \sum_{j=1}^{n} w_{ij} \ln \left(\frac{w_{ij}}{u_{ij}} \right) = \ln(n-1) - H_i$$

  since $\sum_{j=1}^{n} w_{ij} = 1$.

- Discussion: Note that n^{HHI} and n^{E} are both representing the effective number of partners. If $w_{ik} = 1$ for some $k \in \{1, \cdots, n\} \setminus \{i\}$ and $w_{ij} = 0$ if $j \neq k$, $n_i^{\mathrm{HHI}} = n_i^{\mathrm{E}} = 1$. Second, if $w_{ij} = \frac{1}{n-1}$ for all $j \neq i$, $n_i^{\mathrm{HHI}} = n_i^{\mathrm{E}} = n - 1$. That is, first, n^{HHI} and n^{E} have the common range. Second, if n_i^{HHI} and n_i^{E} are both decreasing functions of the variance of $w_{i1}, \cdots, w_{in}$.

  However, there are several distinctions. First, $n_i^{\mathrm{E}} \geq n_i^{\mathrm{HHI}}$ and the equality holds only if $w_{i1}, \cdots, w_{in}$ are uniformly distributed. Second, the entropy-based measure n_i^{E} represents overall diversification, including small partners in the long tail. The HHI-based measure n_i^{HHI} is more conservative and reflects the number of partners that are effectively important in terms of hub dominance. Hence, the gap $n_i^{\mathrm{E}} - n_i^{\mathrm{HHI}}$ illustrates the role of small versus large partners.

Tables 1 - 4 report the detailed network statistics.

Common patterns across all four phases.

- Highly connected hubs
  - Countries such as the United States, Germany, France, the United Kingdom, China, Japan, India, Singapore, Korea, and Australia systematically appear as network hubs. Their degree and weighted degree are close to the maximum (in the 140s in later phases).

Table 1: Detailed network statistics (Phase 1)

Countries	deg	deg^+	c	HHI	n^{HHI}	H_i	$\widetilde{H}_i$	n^{E}	KL divergence
United States	134	122	24.3963	0.0973	10.2786	3.1214	0.6363	22.6779	1.7839
Japan	135	98	13.6553	0.1144	8.7380	3.1067	0.6333	22.3475	1.7986
South Africa	78	3	0.7206	0.1188	8.4203	2.5467	0.5192	12.7651	2.3586
Algeria	107	5	0.7924	0.1233	8.1109	2.6073	0.5315	13.5621	2.2980
Libya	78	5	0.6489	0.1237	8.0863	2.6110	0.5323	13.6122	2.2943
Morocco	119	0	0.2443	0.1003	9.9703	2.9428	0.5999	18.9680	1.9625
Sudan	62	0	0.0675	0.0638	15.6762	3.0905	0.6300	21.9888	1.8147
Tunisia	112	0	0.1719	0.1232	8.1158	2.7186	0.5542	15.1594	2.1867
Egypt	114	1	0.4852	0.0812	12.3212	2.9925	0.6101	19.9353	1.9128
Cameroon	58	0	0.0542	0.1845	5.4214	2.1973	0.4480	9.0009	2.7079
Central African Republic	41	0	0.0028	0.2865	3.4904	1.9159	0.3906	6.7932	2.9893
Chad	36	0	0.0059	0.1802	5.5504	2.0687	0.4217	7.9145	2.8366
Gabon	53	0	0.0576	0.2016	4.9612	2.1092	0.4300	8.2420	2.7960
Angola	63	0	0.0806	0.1884	5.3081	2.2448	0.4576	9.4389	2.6604
Burundi	41	0	0.0116	0.1862	5.3695	2.2161	0.4518	9.1714	2.6892
Comoros	36	0	0.0026	0.2771	3.6086	1.7361	0.3539	5.6753	3.1692
Democratic Republic of the Congo	56	0	0.0488	0.1325	7.5489	2.5326	0.5163	12.5859	2.3727
Benin	44	0	0.0140	0.0844	11.8483	2.7865	0.5681	16.2235	2.1188
Equatorial Guinea	32	0	0.0010	0.1848	5.4110	2.0140	0.4106	7.4929	2.8913
Ethiopia	90	0	0.0341	0.1003	9.9660	2.7979	0.5704	16.4096	2.1074
Gambia	42	0	0.0070	0.0937	10.6764	2.6781	0.5460	14.5569	2.2272
Ghana	61	0	0.0416	0.1086	9.2046	2.7318	0.5569	15.3605	2.1735
Guinea	44	0	0.0153	0.1327	7.5365	2.3656	0.4823	10.6509	2.5396
Côte d'Ivoire	119	6	0.9470	0.1057	9.4606	2.9062	0.5925	18.2873	1.9991
Kenya	101	4	0.6365	0.0641	15.5943	3.2897	0.6706	26.8338	1.6156
Liberia	82	0	0.0927	0.1174	8.5201	2.5969	0.5294	13.4227	2.3083
Madagascar	81	0	0.0655	0.1320	7.5779	2.7029	0.5510	14.9230	2.2024
Malawi	94	0	0.1126	0.1179	8.4841	2.7919	0.5692	16.3120	2.1134
Mali	48	0	0.0255	0.1658	6.0321	2.3231	0.4736	10.2077	2.5821
Mauritania	46	0	0.0093	0.1360	7.3509	2.3269	0.4744	10.2464	2.5784
Mauritius	60	1	0.1147	0.0890	11.2309	2.8415	0.5793	17.1422	2.0637
Mozambique	55	0	0.0377	0.0529	18.9200	3.1785	0.6480	24.0100	1.7268
Niger	40	0	0.0088	0.4581	2.1828	1.4906	0.3039	4.4399	3.4146
Nigeria	61	4	0.3725	0.1099	9.0982	2.5199	0.5137	12.4270	2.3854
Guinea-Bissau	29	0	0.0023	0.1762	5.6762	2.2379	0.4562	9.3735	2.6674
Rwanda	37	0	0.0177	0.1768	5.6569	2.2087	0.4503	9.1041	2.6965
Senegal	55	0	0.0423	0.1676	5.9680	2.4504	0.4995	11.5924	2.4549
Seychelles	60	0	0.0141	0.0699	14.3004	2.9857	0.6087	19.8009	1.9195
Sierra Leone	63	0	0.0105	0.1015	9.8497	2.6952	0.5495	14.8089	2.2100
Somalia	57	0	0.0105	0.1606	6.2262	2.3874	0.4867	10.8848	2.5179
Zimbabwe	113	2	0.2349	0.0851	11.7565	3.0971	0.6314	22.1337	1.8082
Togo	45	0	0.0170	0.1129	8.8538	2.7321	0.5570	15.3644	2.1732
Uganda	53	0	0.0400	0.1071	9.3383	2.5524	0.5203	12.8383	2.3528
Tanzania	62	0	0.0351	0.0694	14.4096	3.0846	0.6288	21.8596	1.8206
Burkina Faso	37	0	0.0213	0.2046	4.8877	2.0591	0.4198	7.8390	2.8462
Zambia	55	0	0.0837	0.0992	10.0805	2.7570	0.5621	15.7531	2.1482
Canada	122	12	2.6152	0.5805	1.7226	1.3317	0.2715	3.7874	3.5736
Bermuda	47	0	0.0111	0.1940	5.1555	2.2031	0.4491	9.0531	2.7022
Greenland	71	0	0.0145	0.4058	2.4645	1.5970	0.3256	4.9380	3.3083
Argentina	115	7	1.1980	0.0619	16.1476	3.2802	0.6687	26.5823	1.6250
Bolivia	79	1	0.0652	0.1714	5.8332	2.2875	0.4663	9.8506	2.6177
Brazil	130	17	2.4845	0.0919	10.8868	3.1941	0.6512	24.3885	1.7112
Chile	106	1	0.2659	0.0953	10.4908	2.9595	0.6033	19.2888	1.9458

Detailed network statistics (Phase 1, continued)

Countries	deg	deg^+	c	HHI	n^{HHI}	H_i	$\widetilde{H}_i$	n^{E}	KL divergence
Colombia	105	0	0.2562	0.1675	5.9714	2.6079	0.5317	13.5708	2.2974
Ecuador	85	0	0.1402	0.2759	3.6248	2.1664	0.4416	8.7267	2.7389
Mexico	73	2	0.3162	0.5080	1.9684	1.3770	0.2807	3.9628	3.5283
Paraguay	70	0	0.0609	0.1320	7.5740	2.6128	0.5327	13.6374	2.2925
Peru	106	0	0.1819	0.1775	5.6330	2.5693	0.5238	13.0573	2.3359
Uruguay	89	0	0.0881	0.1297	7.7112	2.7582	0.5623	15.7722	2.1470
Costa Rica	55	0	0.0194	0.3560	2.8093	1.8731	0.3818	6.5083	3.0322
El Salvador	55	0	0.0171	0.4206	2.3775	1.5809	0.3223	4.8592	3.3244
Guatemala	60	0	0.0230	0.2982	3.3538	2.0166	0.4111	7.5126	2.8887
Honduras	56	0	0.0238	0.3305	3.0260	1.9345	0.3944	6.9205	2.9708
Nicaragua	48	0	0.0142	0.0968	10.3267	2.7043	0.5513	14.9432	2.2010
Bahamas	52	0	0.0288	0.3755	2.6629	1.7683	0.3605	5.8606	3.1370
Barbados	97	1	0.1208	0.3456	2.8939	1.8594	0.3791	6.4196	3.0459
Cuba	60	0	0.0882	0.0840	11.9030	2.8361	0.5782	17.0494	2.0692
Dominican Republic	53	0	0.0240	0.5846	1.7106	1.2230	0.2493	3.3975	3.6822
Haiti	45	0	0.0078	0.5989	1.6698	1.1389	0.2322	3.1232	3.7664
Jamaica	94	0	0.1277	0.2868	3.4867	2.0911	0.4263	8.0937	2.8142
Saint Kitts and Nevis	17	0	0.0043	0.6802	1.4702	0.7472	0.1523	2.1112	4.1580
Trinidad and Tobago	106	5	0.6208	0.2408	4.1531	2.1490	0.4381	8.5766	2.7562
Belize	40	0	0.0044	0.3683	2.7151	1.5957	0.3253	4.9316	3.3096
Guyana	44	0	0.0582	0.1586	6.3045	2.2210	0.4528	9.2162	2.6843
Panama	61	0	0.1526	0.2246	4.4524	2.1999	0.4485	9.0240	2.7054
Suriname	45	0	0.0191	0.1605	6.2298	2.2382	0.4563	9.3763	2.6671
Israel	102	0	0.2471	0.1451	6.8930	2.6032	0.5307	13.5070	2.3021
Bahrain	60	2	0.3568	0.0984	10.1649	2.7635	0.5634	15.8550	2.1418
Cyprus	109	0	0.0668	0.0630	15.8772	3.2025	0.6529	24.5929	1.7028
Iran	65	6	0.8512	0.0761	13.1342	2.9549	0.6024	19.1995	1.9504
Iraq	59	5	0.5534	0.0838	11.9297	2.7877	0.5683	16.2441	2.1175
Jordan	104	0	0.1866	0.0682	14.6673	3.1878	0.6499	24.2354	1.7175
Kuwait	64	2	0.4199	0.0824	12.1317	2.9490	0.6012	19.0871	1.9563
Oman	85	0	0.1700	0.2226	4.4916	2.2224	0.4531	9.2295	2.6829
Qatar	75	0	0.1157	0.2715	3.6835	2.1873	0.4459	8.9112	2.7180
Saudi Arabia	111	17	2.0364	0.1184	8.4467	2.8721	0.5855	17.6747	2.0331
Syria	93	0	0.2531	0.0795	12.5746	3.0545	0.6227	21.2114	1.8507
United Arab Emirates	107	5	0.8347	0.1994	5.0143	2.5299	0.5158	12.5526	2.3753
Turkey	108	3	0.6365	0.0757	13.2156	3.0229	0.6163	20.5513	1.8824
Bangladesh	115	1	0.2791	0.0580	17.2269	3.3613	0.6852	28.8265	1.5440
Bhutan	25	0	0.0003	0.1683	5.9424	2.2313	0.4549	9.3120	2.6740
Brunei	41	0	0.0440	0.4023	2.4859	1.4072	0.2869	4.0844	3.4981
Myanmar	63	0	0.0679	0.0977	10.2390	2.9095	0.5931	18.3480	1.9958
Cambodia	35	0	0.0011	0.3823	2.6155	1.7559	0.3580	5.7885	3.1494
Sri Lanka	121	1	0.2165	0.0584	17.1265	3.3047	0.6737	27.2394	1.6006
Hong Kong	134	5	1.6278	0.1430	6.9936	2.5564	0.5212	12.8892	2.3489
India	132	8	1.6921	0.0677	14.7712	3.2769	0.6680	26.4938	1.6284
Indonesia	106	4	0.8277	0.2078	4.8128	2.2621	0.4612	9.6036	2.6431
South Korea	123	17	2.0511	0.1468	6.8099	2.7667	0.5640	15.9055	2.1386
Laos	35	0	0.0020	0.2228	4.4882	1.9913	0.4060	7.3252	2.9140
Malaysia	108	7	0.8852	0.1370	7.2978	2.5978	0.5296	13.4336	2.3075
Maldives	33	0	0.0044	0.1597	6.2617	2.2581	0.4603	9.5648	2.6472
Nepal	57	0	0.0251	0.3003	3.3295	1.9788	0.4034	7.2341	2.9265
Pakistan	131	2	0.6248	0.0606	16.5080	3.2717	0.6670	26.3563	1.6336
Philippines	113	0	0.2227	0.1542	6.4846	2.5590	0.5217	12.9234	2.3462
Singapore	101	28	3.8457	0.0844	11.8421	3.0293	0.6176	20.6820	1.8760
Thailand	130	15	1.6611	0.0903	11.0760	3.0537	0.6225	21.1940	1.8516
China	80	7	1.3193	0.1716	5.8288	2.4058	0.4905	11.0877	2.4994

Detailed network statistics (Phase 1, continued)

Countries	deg	deg^{+}	c	HHI	n^{HHI}	H_i	$\widetilde{H}_i$	n^{E}	KL divergence
Mongolia	29	0	0.0015	0.1510	6.6223	2.2008	0.4487	9.0318	2.7045
Vietnam	49	0	0.0222	0.1717	5.8258	2.2497	0.4586	9.4847	2.6556
Denmark	134	6	1.7737	0.0910	10.9860	2.9561	0.6026	19.2237	1.9491
France	134	80	11.3272	0.0826	12.1116	3.1899	0.6503	24.2860	1.7154
Germany	134	103	11.2417	0.0671	14.8969	3.2389	0.6603	25.5064	1.6663
Greece	129	3	0.6054	0.0787	12.7135	3.0704	0.6259	21.5515	1.8348
Ireland	129	2	0.4299	0.2066	4.8411	2.3630	0.4817	10.6228	2.5423
Italy	134	74	6.9649	0.0775	12.9072	3.2270	0.6579	25.2049	1.6782
Netherlands	134	45	4.3124	0.1201	8.3249	2.9113	0.5935	18.3804	1.9940
Portugal	129	5	1.1586	0.0667	14.9858	3.2263	0.6577	25.1875	1.6789
Spain	135	30	3.4745	0.0608	16.4461	3.3960	0.6923	29.8453	1.5092
United Kingdom	135	87	8.8630	0.0689	14.5209	3.2601	0.6646	26.0518	1.6452
Austria	133	6	1.0360	0.1923	5.1998	2.6016	0.5304	13.4857	2.3036
Finland	135	2	0.6910	0.0892	11.2075	2.9514	0.6017	19.1333	1.9538
Iceland	102	0	0.0317	0.0956	10.4597	2.7106	0.5526	15.0390	2.1946
Norway	133	9	1.0327	0.1170	8.5464	2.6637	0.5430	14.3497	2.2415
Sweden	134	8	1.8701	0.0804	12.4332	3.0099	0.6136	20.2852	1.8954
Switzerland	134	9	1.7432	0.1054	9.4905	2.9185	0.5950	18.5143	1.9867
Malta	61	0	0.0176	0.1470	6.8043	2.4511	0.4997	11.6010	2.4542
Albania	43	0	0.0070	0.0956	10.4586	2.6361	0.5374	13.9589	2.2692
Bulgaria	67	0	0.1591	0.0694	14.3991	3.1197	0.6360	22.6391	1.7856
Hungary	121	5	0.7400	0.0804	12.4378	3.1899	0.6503	24.2857	1.7154
Australia	135	7	2.0280	0.1204	8.3029	2.8765	0.5864	17.7526	2.0287
New Zealand	123	4	0.5803	0.1083	9.2367	2.8264	0.5762	16.8848	2.0789
Solomon Islands	44	0	0.0092	0.1401	7.1401	2.4833	0.5063	11.9808	2.4220
Fiji	76	1	0.1164	0.1526	6.5512	2.3038	0.4697	10.0123	2.6015
Kiribati	36	0	0.0048	0.1129	8.8575	2.5446	0.5188	12.7387	2.3606
Papua New Guinea	76	0	0.0684	0.1417	7.0579	2.3647	0.4821	10.6407	2.5406

  - They exhibit low concentration (low HHI, high n^{HHI} and high entropy-based effective number of partners n^{E}), indicating that trade links are spread relatively evenly over a large set of partners.

- Peripheral/small countries
  - Countries such as Saint Kitts and Nevis, Comoros, Niger, Greenland, Haiti, and the Turks and Caicos Islands have substantially lower degrees (often in the 20–80 range).
  - Their HHI values are high (around 0.3-0.6) and n^{HHI} values are small (around 2-4), implying heavy dependence on a small number of partners.
  - Their high KL divergence values indicate that their partner distributions differ markedly from a uniform distribution (benchmark), i.e., their networks are highly skewed.

- Many African, Caribbean, and Oceanian countries
  - These countries have moderate degrees with high concentration. For example, Niger, Haiti, and Greenland have a non-trivial number of concentrations but remain heavily concentrated on a few key partners. These countries are therefore

Table 2: Detailed network statistics (Phase 2)

Countries	deg	deg^{+}	c	HHI	n^{HHI}	H_i	$\widetilde{H}_i$	n^{E}	KL divergence
United States	141	119	21.5263	0.0848	11.7871	3.1445	0.6354	23.2085	1.8042
Japan	141	94	10.7785	0.0970	10.3082	3.1167	0.6298	22.5721	1.8320
South Africa	141	12	2.2133	0.0619	16.1626	3.2959	0.6660	27.0022	1.6528
Algeria	133	1	0.3451	0.1006	9.9409	2.8219	0.5702	16.8084	2.1269
Libya	92	0	0.1969	0.1678	5.9589	2.4902	0.5032	12.0636	2.4586
Morocco	134	0	0.3055	0.1227	8.1494	2.9314	0.5923	18.7535	2.0174
Sudan	128	0	0.0842	0.0401	24.9529	3.5708	0.7216	35.5444	1.3780
Tunisia	135	0	0.1930	0.1427	7.0086	2.6579	0.5371	14.2668	2.2908
Egypt	139	0	0.3768	0.0674	14.8402	3.3071	0.6683	27.3046	1.6417
Cameroon	122	3	0.4656	0.1118	8.9426	2.8964	0.5853	18.1097	2.0523
Central African Republic	103	0	0.0203	0.2114	4.7296	2.5218	0.5096	12.4514	2.4269
Chad	87	0	0.0247	0.1332	7.5085	2.6597	0.5375	14.2926	2.2890
Gabon	119	0	0.0813	0.2333	4.2870	2.2762	0.4600	9.7398	2.6725
Angola	93	0	0.0393	0.2900	3.4483	2.0903	0.4224	8.0870	2.8585
Burundi	106	0	0.0360	0.0582	17.1884	3.2865	0.6641	26.7504	1.6622
Comoros	87	0	0.0063	0.2349	4.2575	2.3155	0.4679	10.1300	2.6333
Democratic Republic of the Congo	94	0	0.0979	0.0806	12.4108	3.0670	0.6198	21.4778	1.8817
Benin	124	1	0.1907	0.0639	15.6585	3.3067	0.6682	27.2940	1.6421
Equatorial Guinea	67	0	0.0167	0.1253	7.9799	2.4827	0.5017	11.9740	2.4660
Ethiopia	117	0	0.0362	0.0757	13.2074	3.0853	0.6235	21.8748	1.8634
Gambia	102	0	0.0370	0.0705	14.1933	3.0844	0.6233	21.8543	1.8644
Ghana	128	1	0.1705	0.0693	14.4281	3.1629	0.6391	23.6381	1.7859
Guinea	124	0	0.0850	0.0675	14.8204	3.1954	0.6457	24.4206	1.7533
Côte d'Ivoire	128	7	1.1217	0.0818	12.2225	3.1884	0.6443	24.2496	1.7604
Kenya	124	5	0.6734	0.0464	21.5687	3.5067	0.7086	33.3379	1.4421
Liberia	94	0	0.0511	0.1755	5.6991	2.3777	0.4805	10.7796	2.5711
Madagascar	129	0	0.0485	0.1468	6.8100	2.7517	0.5560	15.6690	2.1971
Malawi	123	0	0.0727	0.0981	10.1915	2.9936	0.6049	19.9579	1.9551
Mali	109	2	0.1093	0.1103	9.0697	2.9190	0.5898	18.5225	2.0298
Mauritania	103	0	0.0662	0.0949	10.5350	2.9255	0.5911	18.6428	2.0233
Mauritius	133	1	0.1536	0.0915	10.9302	2.9752	0.6012	19.5928	1.9736
Mozambique	106	0	0.0697	0.1648	6.0694	2.7512	0.5559	15.6614	2.1976
Niger	105	1	0.1181	0.1339	7.4682	2.6813	0.5418	14.6036	2.2675
Nigeria	128	7	0.8116	0.1239	8.0698	2.8480	0.5755	17.2533	2.1008
Guinea-Bissau	71	0	0.0162	0.1036	9.6509	2.7621	0.5581	15.8325	2.1867
Rwanda	97	0	0.0528	0.0630	15.8684	3.2194	0.6506	25.0136	1.7293
Senegal	131	1	0.2426	0.1100	9.0933	3.1225	0.6310	22.7035	1.8262
Seychelles	92	0	0.0187	0.0918	10.8982	2.7865	0.5631	16.2237	2.1623
Sierra Leone	91	0	0.0113	0.0768	13.0186	3.0984	0.6261	22.1619	1.8504
Somalia	79	0	0.0225	0.1040	9.6112	2.7653	0.5588	15.8832	2.1835
Zimbabwe	137	3	0.2780	0.1247	8.0212	2.9492	0.5959	19.0900	1.9996
Togo	122	0	0.1050	0.0472	21.2048	3.4695	0.7011	32.1216	1.4792
Uganda	133	2	0.2304	0.0661	15.1389	3.2122	0.6491	24.8332	1.7366
Tanzania	130	1	0.1799	0.0496	20.1683	3.4157	0.6902	30.4391	1.5330
Burkina Faso	114	1	0.0808	0.1472	6.7933	2.7450	0.5547	15.5650	2.2037
Zambia	118	0	0.1240	0.0780	12.8189	3.0536	0.6170	21.1907	1.8952
Canada	141	8	1.9126	0.5675	1.7621	1.3778	0.2784	3.9662	3.5709
Bermuda	103	0	0.0191	0.0946	10.5697	2.8816	0.5823	17.8429	2.0672
Greenland	96	0	0.0123	0.4620	2.1646	1.4336	0.2897	4.1938	3.5152
Argentina	137	5	1.0267	0.0953	10.4914	3.0778	0.6219	21.7103	1.8710
Bolivia	124	0	0.0895	0.1083	9.2304	2.6517	0.5358	14.1786	2.2970
Brazil	141	12	2.4140	0.0772	12.9590	3.2884	0.6645	26.8003	1.6603
Chile	136	3	0.4811	0.0784	12.7591	3.1329	0.6331	22.9405	1.8159
Colombia	140	2	0.3996	0.1807	5.5332	2.6679	0.5391	14.4102	2.2808

Detailed network statistics (Phase 2, continued)

Countries	deg	deg^+	c	HHI	n^{HHI}	H_i	$\widetilde{H}_i$	n^{E}	KL divergence
Ecuador	128	0	0.1585	0.1597	6.2621	2.7172	0.5491	15.1382	2.2315
Mexico	139	7	0.9549	0.6004	1.6655	1.2647	0.2556	3.5422	3.6840
Paraguay	98	0	0.0809	0.1574	6.3544	2.4391	0.4929	11.4630	2.5096
Peru	134	1	0.2252	0.0897	11.1459	3.0759	0.6215	21.6684	1.8729
Uruguay	121	0	0.1117	0.1222	8.1841	2.7819	0.5621	16.1497	2.1669
Costa Rica	130	3	0.2395	0.2805	3.5647	2.3159	0.4680	10.1341	2.6329
El Salvador	114	2	0.2154	0.2327	4.2978	2.3161	0.4680	10.1362	2.6326
Guatemala	119	3	0.3306	0.2545	3.9288	2.4132	0.4876	11.1694	2.5356
Honduras	122	0	0.1204	0.3756	2.6626	1.9602	0.3961	7.1006	2.9886
Nicaragua	110	0	0.0775	0.1514	6.6044	2.7102	0.5477	15.0322	2.2386
Bahamas	102	0	0.0429	0.1274	7.8505	2.6754	0.5406	14.5188	2.2733
Barbados	111	0	0.0877	0.1943	5.1475	2.3724	0.4794	10.7234	2.5763
Cuba	100	0	0.0806	0.1009	9.9147	2.8397	0.5738	17.1111	2.1090
Dominican Republic	103	0	0.0796	0.5756	1.7372	1.3516	0.2731	3.8638	3.5971
Haiti	88	0	0.0113	0.4048	2.4703	1.8298	0.3697	6.2325	3.1190
Jamaica	129	2	0.1774	0.2740	3.6503	2.2378	0.4522	9.3725	2.7110
Saint Kitts and Nevis	83	0	0.0110	0.3412	2.9306	1.8033	0.3644	6.0697	3.1454
Trinidad and Tobago	123	5	0.4808	0.2499	4.0021	2.4348	0.4920	11.4133	2.5140
Belize	98	0	0.0228	0.1841	5.4331	2.4895	0.5031	12.0553	2.4593
Guyana	86	0	0.0524	0.1643	6.0865	2.3272	0.4703	10.2491	2.6216
Panama	115	1	0.2685	0.1644	6.0845	2.5656	0.5184	13.0078	2.3832
Suriname	98	0	0.0374	0.1479	6.7597	2.4739	0.4999	11.8682	2.4749
Israel	124	0	0.3270	0.1167	8.5686	2.8440	0.5747	17.1838	2.1048
Bahrain	134	0	0.1029	0.0778	12.8469	3.1154	0.6295	22.5428	1.8333
Cyprus	139	0	0.1154	0.0544	18.3730	3.3491	0.6768	28.4779	1.5996
Iran	110	3	0.5109	0.0552	18.1313	3.3285	0.6726	27.8959	1.6203
Iraq	80	1	0.1373	0.4277	2.3382	1.4881	0.3007	4.4287	3.4606
Jordan	124	1	0.7635	0.0462	21.6528	3.5316	0.7136	34.1790	1.4171
Kuwait	123	0	0.2362	0.0848	11.7956	2.9940	0.6050	19.9654	1.9548
Lebanon	110	1	0.1417	0.0631	15.8571	3.2565	0.6580	25.9573	1.6923
Oman	120	2	0.1985	0.1102	9.0747	2.7622	0.5582	15.8354	2.1865
Qatar	124	0	0.0958	0.1645	6.0802	2.6912	0.5438	14.7491	2.2576
Saudi Arabia	136	10	1.5408	0.0807	12.3869	3.1054	0.6275	22.3174	1.8434
Syria	126	1	0.2510	0.0645	15.5086	3.2724	0.6613	26.3744	1.6764
United Arab Emirates	130	11	1.0687	0.1015	9.8494	3.0305	0.6124	20.7074	1.9183
Turkey	139	8	1.0822	0.0758	13.1912	3.2433	0.6554	25.6185	1.7054
Yemen	110	3	0.3840	0.0497	20.1228	3.3806	0.6831	29.3876	1.5682
Bangladesh	136	2	0.2513	0.0674	14.8433	3.1782	0.6422	24.0031	1.7706
Bhutan	67	0	0.0038	0.4565	2.1906	1.6167	0.3267	5.0364	3.3321
Brunei	109	0	0.0291	0.1557	6.4224	2.2437	0.4534	9.4284	2.7050
Myanmar	96	0	0.0614	0.1183	8.4507	2.6103	0.5275	13.6032	2.3385
Cambodia	83	0	0.0099	0.2115	4.7282	2.1227	0.4289	8.3534	2.8261
Sri Lanka	135	1	0.1873	0.0716	13.9662	3.2287	0.6524	25.2466	1.7201
Hong Kong	140	26	2.7984	0.1828	5.4690	2.4706	0.4992	11.8300	2.4781
India	141	21	2.7996	0.0522	19.1410	3.5076	0.7088	33.3697	1.4411
Indonesia	137	1	0.8909	0.1098	9.1049	2.9126	0.5886	18.4052	2.0361
South Korea	139	34	3.4298	0.0971	10.3013	3.0894	0.6243	21.9648	1.8593
Laos	72	0	0.0129	0.2414	4.1425	2.1769	0.4399	8.8186	2.7719
Malaysia	140	10	1.4587	0.1374	7.2768	2.6144	0.5283	13.6596	2.3343
Maldives	83	0	0.0125	0.1169	8.5522	2.8102	0.5679	16.6139	2.1385
Nepal	106	0	0.0323	0.1218	8.2096	2.6747	0.5405	14.5081	2.2741
Pakistan	137	0	0.5887	0.0493	20.2652	3.4661	0.7004	32.0126	1.4826
Philippines	138	0	0.3579	0.1318	7.5870	2.7000	0.5456	14.8801	2.2487
Singapore	136	32	3.7530	0.0925	10.8164	2.9614	0.5984	19.3248	1.9874
Thailand	140	20	2.8987	0.1100	9.0907	2.9459	0.5953	19.0274	2.0029

Detailed network statistics (Phase 2, continued)

Countries	deg	deg^{+}	c	HHI	n^{HHI}	H_i	$\widetilde{H}_i$	n^{E}	KL divergence
China	141	27	3.7752	0.1702	5.8761	2.4913	0.5034	12.0768	2.4575
Mongolia	72	0	0.0086	0.2732	3.6606	1.9077	0.3855	6.7378	3.0410
Vietnam	112	1	0.2611	0.0903	11.0799	2.9150	0.5890	18.4481	2.0338
Denmark	141	7	1.7182	0.0915	10.9278	3.0239	0.6111	20.5722	1.9248
France	141	85	9.7643	0.0870	11.4913	3.1314	0.6328	22.9053	1.8174
Germany	141	101	10.5596	0.0579	17.2852	3.3180	0.6705	27.6042	1.6308
Greece	141	3	0.9295	0.0791	12.6371	3.1834	0.6433	24.1278	1.7654
Ireland	141	2	0.5698	0.1498	6.6764	2.6178	0.5290	13.7057	2.3310
Italy	141	69	6.7925	0.0791	12.6404	3.2558	0.6579	25.9407	1.6929
Netherlands	141	38	3.7817	0.1072	9.3261	3.0109	0.6084	20.3053	1.9379
Portugal	141	5	1.0843	0.1052	9.5046	2.8251	0.5709	16.8634	2.1236
Spain	141	30	3.4322	0.0891	11.2230	3.0958	0.6256	22.1058	1.8529
United Kingdom	141	72	7.5917	0.0660	15.1503	3.2836	0.6635	26.6716	1.6652
Austria	141	4	0.9691	0.2037	4.9082	2.5420	0.5137	12.7056	2.4067
Finland	141	2	0.6714	0.0690	14.4867	3.1520	0.6369	23.3837	1.7967
Iceland	124	0	0.0411	0.0834	11.9883	2.8653	0.5790	17.5543	2.0835
Norway	141	8	1.0374	0.0867	11.5396	2.9053	0.5871	18.2700	2.0435
Sweden	141	4	1.3254	0.0723	13.8385	3.1065	0.6277	22.3417	1.8423
Switzerland	141	11	1.7490	0.1058	9.4513	2.9551	0.5971	19.2038	1.9937
Malta	136	0	0.0570	0.1316	7.6009	2.6785	0.5412	14.5633	2.2703
Albania	94	0	0.0205	0.2269	4.4078	2.1095	0.4263	8.2442	2.8392
Bulgaria	132	1	0.2530	0.1003	9.9683	3.0269	0.6116	20.6332	1.9219
Czechia	140	0	0.3616	0.1876	5.3303	2.5760	0.5205	13.1444	2.3728
Hungary	141	0	0.3247	0.1378	7.2583	2.6682	0.5392	14.4136	2.2806
Poland	140	2	0.5717	0.1376	7.2676	2.8398	0.5738	17.1129	2.1089
Romania	138	0	0.3666	0.0749	13.3433	3.2379	0.6543	25.4790	1.7109
Russia	137	14	2.4346	0.0529	18.8963	3.3959	0.6862	29.8429	1.5528
Australia	141	6	2.0905	0.0900	11.1117	2.9970	0.6056	20.0248	1.9518
New Zealand	138	4	0.5362	0.1109	9.0167	2.8671	0.5794	17.5864	2.0816
Solomon Islands	55	0	0.0055	0.1833	5.4542	2.2544	0.4555	9.5293	2.6944
Fiji	117	1	0.1104	0.1681	5.9492	2.3139	0.4676	10.1141	2.6348
Kiribati	56	0	0.0060	0.1355	7.3807	2.4549	0.4961	11.6453	2.4939
Papua New Guinea	80	0	0.0598	0.1962	5.0980	2.1939	0.4433	8.9702	2.7549

Table 3: Detailed network statistics (Phase 3)

Countries	deg	deg^+	c	HHI	n^{HHI}	H_i	$\widetilde{H}_i$	n^{E}	KL divergence
United States	145	122	21.2007	0.0788	12.6827	3.2184	0.6467	24.9871	1.7584
Germany	145	92	8.9951	0.0499	20.0467	3.3844	0.6800	29.5010	1.5923
South Africa	145	14	2.4110	0.0508	19.6769	3.4878	0.7008	32.7131	1.4890
Algeria	142	0	0.2899	0.0904	11.0610	2.8894	0.5806	17.9824	2.0873
Libya	119	1	0.1631	0.1539	6.4983	2.5577	0.5139	12.9063	2.4190
Morocco	142	0	0.2747	0.0886	11.2904	3.1414	0.6312	23.1368	1.8353
Sudan	140	0	0.1084	0.1371	7.2913	2.8436	0.5714	17.1783	2.1331
Tunisia	144	0	0.1640	0.1406	7.1147	2.6857	0.5397	14.6687	2.2910
Egypt	144	0	0.4459	0.0512	19.5237	3.5178	0.7068	33.7095	1.4590
Cameroon	139	1	0.2448	0.0734	13.6268	3.1875	0.6405	24.2290	1.7892
Central African Republic	121	0	0.0103	0.1380	7.2444	2.7627	0.5551	15.8419	2.2141
Chad	108	0	0.0162	0.3392	2.9485	1.9460	0.3910	7.0005	3.0307
Gabon	135	0	0.0932	0.2342	4.2700	2.3978	0.4818	10.9992	2.5789
Angola	119	0	0.0724	0.1845	5.4186	2.3442	0.4710	10.4252	2.6325
Burundi	130	0	0.0339	0.0457	21.8777	3.4641	0.6961	31.9472	1.5126
Comoros	117	0	0.0051	0.0782	12.7956	3.1971	0.6424	24.4605	1.7797
Democratic Republic of the Congo	116	1	0.1301	0.1423	7.0293	2.7022	0.5430	14.9131	2.2745
Benin	140	0	0.1423	0.1106	9.0379	3.0771	0.6183	21.6955	1.8996
Equatorial Guinea	98	0	0.0370	0.1657	6.0364	2.2960	0.4614	9.9348	2.6807
Ethiopia	145	1	0.0964	0.0546	18.3251	3.3441	0.6719	28.3346	1.6326
Gambia	128	0	0.0271	0.0671	14.8970	3.2557	0.6542	25.9365	1.7211
Ghana	144	2	0.3017	0.0446	22.4416	3.4943	0.7021	32.9258	1.4825
Guinea	136	0	0.0619	0.0506	19.7475	3.3979	0.6828	29.9011	1.5788
Côte d'Ivoire	144	8	0.7961	0.0751	13.3172	3.3046	0.6640	27.2382	1.6721
Kenya	145	5	0.8122	0.0419	23.8840	3.5783	0.7190	35.8118	1.3985
Liberia	122	0	0.0421	0.1306	7.6589	2.6094	0.5243	13.5903	2.3674
Madagascar	144	0	0.0900	0.1009	9.9100	2.9680	0.5964	19.4524	2.0088
Malawi	143	0	0.0840	0.0924	10.8215	3.1634	0.6356	23.6500	1.8134
Mali	140	2	0.1834	0.0692	14.4601	3.1981	0.6426	24.4858	1.7786
Mauritania	137	0	0.0497	0.0558	17.9074	3.3371	0.6705	28.1378	1.6396
Mauritius	142	1	0.1686	0.0650	15.3821	3.2710	0.6573	26.3390	1.7057
Mozambique	141	2	0.1389	0.1060	9.4302	2.9345	0.5896	18.8120	2.0422
Niger	137	0	0.0582	0.0646	15.4797	3.3071	0.6645	27.3056	1.6696
Nigeria	144	7	0.8759	0.1322	7.5625	2.9075	0.5842	18.3115	2.0692
Guinea-Bissau	101	0	0.0211	0.1092	9.1588	2.7662	0.5558	15.8987	2.2105
Rwanda	136	0	0.0515	0.0689	14.5201	3.2267	0.6484	25.1966	1.7500
Senegal	143	3	0.4136	0.0619	16.1662	3.4559	0.6944	31.6880	1.5208
Seychelles	129	0	0.0209	0.0813	12.3059	3.0034	0.6035	20.1543	1.9733
Sierra Leone	123	0	0.0177	0.0653	15.3167	3.2807	0.6592	26.5947	1.6960
Somalia	110	0	0.0282	0.1729	5.7846	2.5342	0.5092	12.6064	2.4425
Zimbabwe	141	4	0.3359	0.1264	7.9130	2.9142	0.5856	18.4341	2.0625
Togo	143	2	0.2330	0.0533	18.7454	3.4339	0.6900	30.9987	1.5428
Uganda	143	3	0.1887	0.0647	15.4611	3.3376	0.6706	28.1509	1.6392
Tanzania	144	1	0.2465	0.0444	22.5344	3.5035	0.7040	33.2310	1.4733
Burkina Faso	134	1	0.1565	0.0786	12.7299	3.1330	0.6295	22.9419	1.8438
Zambia	138	1	0.2452	0.1127	8.8700	2.8541	0.5735	17.3597	2.1226
Canada	145	12	1.9110	0.5475	1.8263	1.4602	0.2934	4.3069	3.5165
Bermuda	123	0	0.0206	0.0935	10.7006	2.7857	0.5597	16.2109	2.1911
Greenland	121	0	0.0111	0.4449	2.2477	1.5931	0.3201	4.9189	3.3837
Argentina	143	5	1.0159	0.0887	11.2785	3.2106	0.6451	24.7951	1.7661
Bolivia	139	0	0.0559	0.1238	8.0752	2.6355	0.5296	13.9503	2.3412
Brazil	145	15	2.3951	0.0749	13.3565	3.3792	0.6790	29.3472	1.5975
Chile	143	4	0.5158	0.0714	13.9975	3.1475	0.6324	23.2782	1.8292
Colombia	145	3	0.4849	0.1707	5.8592	2.8096	0.5645	16.6030	2.1672

Detailed network statistics (Phase 3, continued)

Countries	deg	deg^+	c	HHI	n^{HHI}	H_i	$\widetilde{H}_i$	n^{E}	KL divergence
Ecuador	141	2	0.2418	0.1451	6.8897	2.8449	0.5716	17.1998	2.1318
Mexico	145	11	1.2606	0.5294	1.8890	1.4883	0.2991	4.4295	3.4884
Paraguay	136	2	0.1844	0.1073	9.3193	2.7862	0.5598	16.2187	2.1906
Peru	145	2	0.2897	0.0898	11.1390	3.0822	0.6193	21.8073	1.8945
Uruguay	141	2	0.3521	0.0768	13.0268	3.2130	0.6456	24.8527	1.7638
Costa Rica	144	1	0.2556	0.1794	5.5743	2.7167	0.5459	15.1308	2.2600
El Salvador	135	3	0.2567	0.2300	4.3475	2.4037	0.4830	11.0641	2.5730
Guatemala	140	5	0.3978	0.2133	4.6887	2.5602	0.5144	12.9384	2.4165
Honduras	140	1	0.1868	0.4223	2.3679	1.8607	0.3739	6.4282	3.1160
Nicaragua	141	0	0.0842	0.1939	5.1565	2.5413	0.5106	12.6961	2.4354
Bahamas	125	0	0.0498	0.1296	7.7150	2.7021	0.5429	14.9110	2.2746
Barbados	144	0	0.1015	0.1547	6.4641	2.6719	0.5369	14.4669	2.3049
Cayman Islands	99	1	0.0547	0.1329	7.5257	2.5366	0.5097	12.6366	2.4401
Cuba	141	0	0.0997	0.0692	14.4529	3.1586	0.6347	23.5380	1.8181
Dominican Republic	143	1	0.1956	0.4334	2.3072	1.8329	0.3683	6.2521	3.1438
Haiti	114	0	0.0288	0.3614	2.7670	2.0250	0.4069	7.5762	2.9517
Jamaica	134	1	0.1503	0.1755	5.6964	2.6560	0.5337	14.2386	2.3208
Saint Kitts and Nevis	125	0	0.0087	0.3053	3.2756	2.0347	0.4088	7.6499	2.9420
Trinidad and Tobago	142	5	0.8026	0.2872	3.4818	2.3713	0.4765	10.7118	2.6054
Belize	128	0	0.0243	0.1582	6.3204	2.7921	0.5610	16.3158	2.1846
Guyana	138	0	0.0706	0.1268	7.8852	2.7181	0.5462	15.1513	2.2587
Panama	129	1	0.2834	0.1179	8.4799	2.8907	0.5808	18.0050	2.0861
Suriname	128	0	0.0699	0.0971	10.3030	2.8618	0.5750	17.4926	2.1150
Israel	143	0	0.4028	0.1237	8.0856	2.9111	0.5849	18.3763	2.0657
Japan	145	65	7.5501	0.0853	11.7256	3.1756	0.6381	23.9403	1.8012
Bahrain	139	2	0.2724	0.0913	10.9577	3.1977	0.6425	24.4764	1.7790
Cyprus	145	0	0.1186	0.1067	9.3689	2.9707	0.5969	19.5058	2.0060
Iran	142	2	0.5828	0.0636	15.7329	3.2198	0.6470	25.0223	1.7570
Iraq	121	1	0.2074	0.1552	6.4418	2.7440	0.5514	15.5484	2.2328
Jordan	138	1	0.1672	0.0531	18.8439	3.4429	0.6918	31.2761	1.5339
Kuwait	140	1	0.3112	0.0801	12.4871	3.0735	0.6176	21.6169	1.9033
Lebanon	145	0	0.1613	0.0419	23.8742	3.5645	0.7162	35.3225	1.4122
Oman	140	1	0.2099	0.1017	9.8339	2.7859	0.5598	16.2145	2.1908
Qatar	134	0	0.1506	0.1515	6.5996	2.6411	0.5307	14.0288	2.3356
Saudi Arabia	142	15	1.9016	0.0718	13.9306	3.2462	0.6523	25.6919	1.7306
Syria	137	1	0.2111	0.0637	15.7038	3.3435	0.6718	28.3177	1.6332
United Arab Emirates	144	19	2.3537	0.0623	16.0580	3.3813	0.6794	29.4084	1.5955
Turkey	145	10	1.3131	0.0538	18.6004	3.4578	0.6948	31.7484	1.5189
Yemen	143	1	0.1674	0.0732	13.6589	3.1397	0.6309	23.0975	1.8370
Afghanistan	124	0	0.0338	0.1141	8.7613	2.7312	0.5488	15.3506	2.2456
Bangladesh	145	0	0.2538	0.0591	16.9063	3.3079	0.6647	27.3277	1.6688
Bhutan	98	0	0.0032	0.3768	2.6539	1.9163	0.3851	6.7960	3.0604
Brunei	129	0	0.0237	0.1433	6.9760	2.3445	0.4711	10.4277	2.6323
Myanmar	119	0	0.0449	0.1339	7.4694	2.5427	0.5109	12.7142	2.4340
Cambodia	140	0	0.0289	0.1245	8.0300	2.5423	0.5108	12.7090	2.4344
Sri Lanka	143	1	0.1911	0.0668	14.9685	3.2546	0.6540	25.9089	1.7221
Hong Kong	144	14	1.8431	0.2238	4.4675	2.3468	0.4716	10.4523	2.6299
India	145	26	3.8904	0.0427	23.4287	3.6703	0.7375	39.2620	1.3065
Indonesia	145	3	1.2206	0.0862	11.6070	3.0441	0.6117	20.9912	1.9326
South Korea	145	37	3.9528	0.0836	11.9602	3.1998	0.6429	24.5264	1.7770
Laos	112	0	0.0136	0.3160	3.1642	1.9532	0.3925	7.0511	3.0236
Malaysia	145	7	1.4554	0.1031	9.7034	2.8689	0.5765	17.6185	2.1078
Maldives	109	0	0.0141	0.0858	11.6483	2.8983	0.5824	18.1440	2.0784
Nepal	124	0	0.0440	0.2639	3.7886	2.2014	0.4423	9.0381	2.7753

Detailed network statistics (Phase 3, continued)

Countries	deg	deg^{+}	c	HHI	n^{HHI}	H_i	$\widetilde{H}_i$	n^{E}	KL divergence
Pakistan	145	2	0.7450	0.0521	19.1888	3.4853	0.7003	32.6307	1.4915
Philippines	145	1	0.4010	0.0992	10.0833	2.8183	0.5663	16.7482	2.1584
Singapore	145	26	3.1214	0.0725	13.7848	3.1237	0.6277	22.7302	1.8530
Thailand	145	17	2.7197	0.0758	13.1979	3.2186	0.6467	24.9922	1.7582
China	145	83	8.9125	0.1002	9.9801	3.0178	0.6064	20.4455	1.9590
Mongolia	116	0	0.0074	0.1987	5.0324	2.2143	0.4449	9.1546	2.7625
Vietnam	145	2	0.5281	0.0691	14.4650	3.1377	0.6305	23.0504	1.8390
Belgium	145	31	3.8489	0.0955	10.4726	2.9633	0.5954	19.3609	2.0135
Denmark	145	4	1.4764	0.0812	12.3226	3.0905	0.6210	21.9876	1.8863
France	145	69	7.5737	0.0713	14.0319	3.2389	0.6508	25.5064	1.7378
Greece	145	3	0.8192	0.0575	17.3891	3.3743	0.6780	29.2030	1.6025
Ireland	145	2	0.6351	0.1176	8.5048	2.7323	0.5490	15.3686	2.2444
Italy	145	52	6.0066	0.0613	16.3193	3.4157	0.6863	30.4381	1.5610
Netherlands	145	38	3.7473	0.0788	12.6932	3.1753	0.6380	23.9344	1.8014
Portugal	144	6	0.7721	0.1156	8.6497	2.8606	0.5748	17.4714	2.1162
Spain	145	28	3.8669	0.0755	13.2473	3.2453	0.6521	25.6685	1.7315
United Kingdom	145	68	5.8639	0.0607	16.4823	3.3246	0.6680	27.7890	1.6521
Austria	145	6	0.8569	0.1719	5.8171	2.6970	0.5419	14.8351	2.2797
Finland	145	3	0.6606	0.0621	16.0997	3.2630	0.6556	26.1270	1.7138
Iceland	139	0	0.0935	0.0685	14.6032	3.1017	0.6232	22.2350	1.8751
Norway	145	7	0.9439	0.0809	12.3546	2.9608	0.5949	19.3135	2.0159
Sweden	145	5	1.2810	0.0622	16.0875	3.2137	0.6458	24.8717	1.7630
Switzerland	145	12	1.8459	0.0873	11.4536	3.1511	0.6332	23.3614	1.8256
Malta	142	0	0.0714	0.0637	15.7071	3.2228	0.6476	25.0994	1.7539
Albania	137	0	0.0269	0.2071	4.8276	2.3789	0.4780	10.7925	2.5979
Bulgaria	145	1	0.2240	0.0642	15.5705	3.2080	0.6446	24.7301	1.7687
Czechia	145	1	0.4694	0.1547	6.4624	2.7310	0.5487	15.3476	2.2458
Hungary	145	3	0.4674	0.1185	8.4389	2.8958	0.5819	18.0982	2.0809
Poland	145	2	0.7275	0.1154	8.6674	2.9090	0.5845	18.3391	2.0677
Romania	145	2	0.3601	0.0842	11.8834	3.0620	0.6153	21.3695	1.9148
Serbia	144	0	0.1253	0.0801	12.4830	3.0406	0.6110	20.9186	1.9361
Russia	145	14	2.2772	0.0522	19.1409	3.3804	0.6792	29.3826	1.5963
Australia	145	8	2.2544	0.0727	13.7522	3.1495	0.6329	23.3252	1.8272
New Zealand	145	3	0.5155	0.0959	10.4233	3.0195	0.6067	20.4809	1.9572
Solomon Islands	98	0	0.0078	0.0953	10.4954	2.7878	0.5602	16.2451	2.1889
Fiji	137	1	0.1249	0.1462	6.8383	2.4115	0.4846	11.1506	2.5652
Kiribati	83	0	0.0052	0.1447	6.9104	2.5176	0.5059	12.3987	2.4591
Papua New Guinea	136	0	0.0784	0.2269	4.4082	2.2331	0.4487	9.3284	2.7437

Table 4: Detailed network statistics (Phase 4)

Countries	deg	deg^+	c	HHI	n^{HHI}	H_i	$\widetilde{H}_i$	n^{E}	KL divergence
United States	146	108	16.4964	0.0773	12.9335	3.2681	0.6558	26.2602	1.7156
China	147	137	17.4830	0.0584	17.1273	3.4661	0.6955	32.0101	1.5176
Germany	146	62	7.5205	0.0466	21.4564	3.4421	0.6907	31.2522	1.5415
South Africa	147	15	2.3228	0.0722	13.8414	3.3989	0.6820	29.9299	1.5847
Algeria	142	3	0.6466	0.0633	15.7941	3.2051	0.6431	24.6590	1.7785
Libya	118	0	0.1704	0.0955	10.4675	2.9147	0.5849	18.4431	2.0689
Morocco	143	0	0.3654	0.0693	14.4249	3.3234	0.6669	27.7540	1.6602
Sudan	143	1	0.1286	0.1683	5.9426	2.6846	0.5387	14.6523	2.2990
Tunisia	145	0	0.1958	0.1005	9.9469	3.0217	0.6063	20.5252	1.9620
Egypt	146	4	0.6953	0.0390	25.6605	3.6769	0.7378	39.5229	1.3067
Cameroon	143	2	0.1906	0.0599	16.6858	3.3576	0.6737	28.7202	1.6260
Central African Republic	125	0	0.0213	0.0802	12.4671	3.1552	0.6331	23.4570	1.8284
Chad	109	0	0.0314	0.3374	2.9636	1.9403	0.3893	6.9607	3.0433
Gabon	120	1	0.0833	0.0883	11.3223	2.8727	0.5764	17.6841	2.1109
Angola	137	1	0.1552	0.1897	5.2717	2.4007	0.4817	11.0307	2.5829
Burundi	132	0	0.0264	0.0492	20.3265	3.3277	0.6677	27.8742	1.6559
Comoros	105	0	0.0070	0.0780	12.8241	3.0274	0.6075	20.6443	1.9562
Democratic Republic of the Congo	113	2	0.2575	0.1263	7.9162	2.7457	0.5510	15.5761	2.2379
Benin	140	1	0.1131	0.1250	7.9972	2.9332	0.5886	18.7873	2.0504
Equatorial Guinea	100	0	0.0589	0.0719	13.9060	2.9142	0.5847	18.4332	2.0695
Ethiopia	142	1	0.2284	0.0766	13.0594	3.2318	0.6485	25.3240	1.7519
Djibouti	112	0	0.0192	0.1104	9.0589	2.9098	0.5839	18.3538	2.0738
Gambia	134	0	0.0270	0.1129	8.8557	2.9755	0.5971	19.5994	2.0081
Ghana	142	2	0.4314	0.0535	18.6818	3.3772	0.6777	29.2897	1.6064
Guinea	132	0	0.0586	0.0592	16.9003	3.3307	0.6683	27.9575	1.6529
Côte d'Ivoire	146	4	0.4743	0.0490	20.4276	3.5307	0.7085	34.1464	1.4530
Kenya	141	5	0.5874	0.0591	16.9172	3.4084	0.6839	30.2183	1.5752
Liberia	125	0	0.0444	0.1769	5.6544	2.2623	0.4539	9.6052	2.7213
Madagascar	145	0	0.0876	0.0672	14.8846	3.2416	0.6504	25.5736	1.7420
Malawi	143	0	0.0525	0.0577	17.3292	3.3640	0.6750	28.9049	1.6196
Mali	140	1	0.1626	0.0761	13.1443	3.1426	0.6306	23.1638	1.8410
Mauritania	142	0	0.0444	0.1011	9.8954	3.0617	0.6143	21.3628	1.9220
Mauritius	145	2	0.1450	0.0603	16.5816	3.3411	0.6704	28.2503	1.6425
Mozambique	146	2	0.1537	0.0903	11.0768	3.0672	0.6154	21.4807	1.9165
Niger	143	0	0.0515	0.1091	9.1675	2.9502	0.5920	19.1102	2.0334
Nigeria	143	9	2.1013	0.0630	15.8770	3.3117	0.6645	27.4310	1.6719
Guinea-Bissau	101	0	0.0158	0.1347	7.4254	2.7383	0.5495	15.4602	2.2453
Rwanda	140	0	0.0919	0.0572	17.4934	3.3079	0.6638	27.3281	1.6757
Senegal	144	4	0.4577	0.0485	20.6104	3.5277	0.7079	34.0465	1.4559
Seychelles	133	0	0.0214	0.0725	13.7900	3.1958	0.6413	24.4304	1.7878
Sierra Leone	129	0	0.0318	0.0902	11.0906	3.1725	0.6366	23.8670	1.8111
Somalia	108	0	0.0477	0.1179	8.4827	2.5500	0.5117	12.8077	2.4336
Zimbabwe	144	0	0.1499	0.2552	3.9179	2.3245	0.4664	10.2211	2.6592
Togo	136	2	0.2141	0.0872	11.4716	3.1571	0.6335	23.5027	1.8265
Uganda	146	3	0.2801	0.0571	17.5062	3.3790	0.6780	29.3427	1.6046
Tanzania	146	4	0.4661	0.0700	14.2892	3.2036	0.6428	24.6222	1.7800
Burkina Faso	137	0	0.1017	0.0598	16.7127	3.3649	0.6752	28.9316	1.6187
Zambia	145	3	0.3837	0.1162	8.6038	2.7276	0.5473	15.2968	2.2560
Canada	147	9	1.8347	0.4216	2.3717	1.8557	0.3724	6.3965	3.1279
Bermuda	137	0	0.0102	0.1802	5.5485	2.2316	0.4478	9.3151	2.7520
Greenland	125	0	0.0117	0.3548	2.8183	1.8207	0.3653	6.1759	3.1630
Argentina	147	6	0.9031	0.0869	11.5092	3.2713	0.6564	26.3448	1.7123
Bolivia	144	0	0.0757	0.1128	8.8653	2.7079	0.5434	14.9974	2.2757
Brazil	147	15	2.5999	0.0687	14.5507	3.4078	0.6838	30.1973	1.5759
Chile	147	7	0.5965	0.0971	10.2940	2.9929	0.6006	19.9444	1.9907

Detailed network statistics (Phase 4, continued)

Countries	deg	deg^{+}	c	HHI	n^{HHI}	H_i	$\widetilde{H}_i$	n^{E}	KL divergence
Colombia	147	4	0.5854	0.1274	7.8466	2.9577	0.5935	19.2542	2.0259
Ecuador	147	0	0.2476	0.1357	7.3717	2.8556	0.5730	17.3845	2.1280
Mexico	146	10	1.2428	0.4260	2.3474	1.7802	0.3572	5.9312	3.2034
Paraguay	141	0	0.1198	0.1073	9.3167	2.8715	0.5762	17.6631	2.1121
Peru	146	2	0.3447	0.0857	11.6637	3.0680	0.6156	21.4981	1.9156
Uruguay	145	0	0.1355	0.0869	11.5085	3.1620	0.6345	23.6170	1.8216
Costa Rica	146	3	0.2779	0.1768	5.6560	2.6623	0.5342	14.3293	2.3213
El Salvador	143	4	0.2818	0.1950	5.1275	2.4829	0.4982	11.9759	2.5007
Guatemala	140	3	0.3777	0.1727	5.7902	2.7162	0.5450	15.1227	2.2674
Honduras	132	2	0.1939	0.2998	3.3351	2.1583	0.4331	8.6567	2.8253
Nicaragua	143	1	0.1018	0.1820	5.4949	2.5683	0.5154	13.0442	2.4153
Bahamas	129	1	0.1535	0.1257	7.9560	2.7137	0.5445	15.0845	2.2699
Barbados	145	0	0.0909	0.2510	3.9835	2.1665	0.4347	8.7280	2.8171
Cayman Islands	106	0	0.0303	0.0908	11.0145	2.8088	0.5636	16.5907	2.1748
Cuba	128	0	0.0418	0.0839	11.9147	3.0841	0.6189	21.8486	1.8995
Dominican Republic	146	1	0.3767	0.2278	4.3900	2.5476	0.5112	12.7770	2.4360
Haiti	125	1	0.0522	0.2399	4.1681	2.2511	0.4517	9.4978	2.7325
Jamaica	134	0	0.1034	0.1908	5.2417	2.6388	0.5295	13.9964	2.3448
Saint Kitts and Nevis	127	0	0.0061	0.2015	4.9638	2.3907	0.4797	10.9213	2.5929
Trinidad and Tobago	141	6	0.7085	0.1510	6.6247	2.8643	0.5748	17.5374	2.1193
Turks and Caicos Islands	110	0	0.0090	0.4866	2.0550	1.6074	0.3225	4.9900	3.3762
Belize	130	0	0.0356	0.4089	2.4458	1.7976	0.3607	6.0352	3.1860
Guyana	143	0	0.0714	0.1190	8.4054	2.8034	0.5625	16.5012	2.1802
Panama	139	2	0.4344	0.1050	9.5259	2.7828	0.5584	16.1648	2.2008
Suriname	127	0	0.0842	0.0845	11.8394	2.9638	0.5947	19.3714	2.0198
Israel	146	1	0.4029	0.0978	10.2261	3.0796	0.6179	21.7501	1.9040
Japan	146	46	5.4337	0.0818	12.2187	3.2299	0.6481	25.2784	1.7537
Bahrain	147	0	0.1822	0.0580	17.2411	3.3862	0.6795	29.5549	1.5974
Cyprus	144	0	0.0925	0.0558	17.9175	3.2947	0.6611	26.9693	1.6889
Iran	137	4	0.5720	0.1050	9.5196	2.8213	0.5661	16.7986	2.1623
Iraq	113	2	0.3928	0.0844	11.8489	2.8836	0.5786	17.8793	2.1000
Jordan	141	0	0.1632	0.0606	16.4911	3.3807	0.6784	29.3914	1.6029
Kuwait	146	2	0.4483	0.0855	11.6899	2.9759	0.5971	19.6067	2.0077
Lebanon	147	1	0.2184	0.0380	26.3322	3.6583	0.7341	38.7940	1.3253
Oman	143	1	0.3135	0.1097	9.1134	2.8532	0.5725	17.3430	2.1304
Qatar	142	0	0.2968	0.1039	9.6237	2.8597	0.5738	17.4557	2.1239
Saudi Arabia	146	17	1.9545	0.0723	13.8336	3.1863	0.6394	24.1983	1.7973
Syria	141	0	0.1355	0.0482	20.7552	3.4294	0.6881	30.8573	1.5542
United Arab Emirates	146	37	4.0318	0.0558	17.9194	3.4598	0.6942	31.8091	1.5239
Turkey	146	17	1.9245	0.0446	22.3968	3.6033	0.7230	36.7205	1.3803
Yemen	140	1	0.1098	0.0766	13.0476	3.1449	0.6310	23.2174	1.8387
Afghanistan	126	0	0.0618	0.1077	9.2893	2.6534	0.5324	14.2020	2.3302
Bangladesh	146	1	0.2646	0.0567	17.6335	3.3737	0.6770	29.1866	1.6099
Bhutan	97	0	0.0045	0.4484	2.2303	1.6658	0.3343	5.2901	3.3178
Brunei	136	0	0.0267	0.1371	7.2934	2.3890	0.4794	10.9030	2.5946
Myanmar	139	0	0.0655	0.1884	5.3083	2.2287	0.4472	9.2875	2.7549
Cambodia	144	0	0.0601	0.0906	11.0331	2.8040	0.5626	16.5102	2.1796
Sri Lanka	146	1	0.1518	0.0673	14.8635	3.2589	0.6539	26.0222	1.7247
Hong Kong	146	10	1.4000	0.2965	3.3731	2.1725	0.4359	8.7800	2.8111
India	146	53	6.5293	0.0398	25.1201	3.7334	0.7491	41.8211	1.2502
Indonesia	147	5	1.3589	0.0781	12.8111	3.0997	0.6220	22.1905	1.8839
South Korea	146	39	4.7301	0.0838	11.9340	3.2764	0.6574	26.4797	1.7072
Laos	116	0	0.0185	0.3280	3.0487	1.6555	0.3322	5.2359	3.3281
Malaysia	147	11	1.6446	0.0921	10.8614	3.0001	0.6020	20.0873	1.9835
Maldives	113	0	0.0122	0.0848	11.7964	2.9649	0.5949	19.3928	2.0187

Detailed network statistics (Phase 4, continued)

Countries	deg	deg^+	c	HHI	n^{HHI}	H_i	$\widetilde{H}_i$	n^{E}	KL divergence
Nepal	143	0	0.0265	0.3417	2.9268	1.8367	0.3685	6.2758	3.1469
Pakistan	146	2	0.7250	0.0644	15.5177	3.3909	0.6804	29.6922	1.5927
Philippines	146	0	0.2916	0.0999	10.0144	2.8383	0.5695	17.0867	2.1453
Singapore	146	31	3.2812	0.0613	16.3242	3.2637	0.6549	26.1449	1.7200
Thailand	147	17	2.6661	0.0639	15.6411	3.3429	0.6708	28.3015	1.6407
Vietnam	143	5	1.0785	0.0849	11.7741	3.1123	0.6245	22.4734	1.8713
Belgium	147	26	2.8285	0.0857	11.6640	3.1054	0.6231	22.3182	1.8782
Denmark	146	4	1.1927	0.0758	13.1899	3.1795	0.6380	24.0348	1.8041
France	147	53	5.3800	0.0651	15.3640	3.3504	0.6723	28.5150	1.6332
Greece	146	5	0.6806	0.0444	22.5219	3.5430	0.7109	34.5705	1.4406
Ireland	146	1	0.4470	0.1060	9.4358	2.8569	0.5733	17.4076	2.1267
Italy	146	42	4.4342	0.0516	19.3653	3.5388	0.7101	34.4250	1.4448
Netherlands	146	41	3.8671	0.0736	13.5799	3.2561	0.6534	25.9483	1.7275
Portugal	147	4	0.7117	0.1102	9.0736	3.0092	0.6038	20.2707	1.9744
Spain	147	23	3.1078	0.0570	17.5509	3.4727	0.6968	32.2226	1.5109
United Kingdom	147	41	3.9471	0.0528	18.9374	3.4466	0.6916	31.3949	1.5370
Austria	146	5	0.7010	0.1680	5.9539	2.7449	0.5508	15.5627	2.2387
Finland	147	1	0.4434	0.0678	14.7407	3.2034	0.6428	24.6160	1.7802
Iceland	141	0	0.0752	0.0706	14.1654	3.1389	0.6298	23.0777	1.8447
Norway	146	4	0.7178	0.0794	12.5920	3.0165	0.6053	20.4197	1.9671
Sweden	146	4	1.1698	0.0604	16.5647	3.2647	0.6551	26.1724	1.7189
Switzerland	146	15	2.0226	0.0718	13.9226	3.2414	0.6504	25.5682	1.7423
Malta	144	0	0.1198	0.0625	15.9963	3.2434	0.6508	25.6212	1.7402
Albania	143	0	0.0432	0.1590	6.2904	2.6622	0.5342	14.3278	2.3214
Bulgaria	147	3	0.2935	0.0598	16.7185	3.2675	0.6557	26.2464	1.7161
Czechia	147	4	0.5972	0.1278	7.8242	2.8511	0.5721	17.3073	2.1325
Hungary	146	3	0.4800	0.0982	10.1792	2.9953	0.6010	19.9905	1.9883
Poland	146	6	1.0471	0.1025	9.7514	2.9961	0.6012	20.0072	1.9875
Romania	146	3	0.5229	0.0756	13.2314	3.1475	0.6316	23.2769	1.8361
Serbia	145	1	0.1595	0.0690	14.4917	3.1251	0.6271	22.7618	1.8585
Russia	147	19	2.3579	0.0575	17.3828	3.3424	0.6707	28.2876	1.6412
Australia	147	9	1.8989	0.1096	9.1218	2.9358	0.5891	18.8368	2.0478
New Zealand	147	2	0.4490	0.0857	11.6673	3.1459	0.6313	23.2412	1.8377
Solomon Islands	120	0	0.0109	0.1921	5.2057	2.3413	0.4698	10.3948	2.6423
Fiji	140	1	0.1138	0.1083	9.2347	2.6612	0.5340	14.3139	2.3224
Kiribati	100	0	0.0102	0.0946	10.5673	2.7383	0.5495	15.4602	2.2453
Papua New Guinea	134	0	0.0598	0.1744	5.7347	2.3991	0.4814	11.0132	2.5845

"connected but dependent": they are not isolated, but their trade is still dominated by a small subset of partners.

- Interpretation of the KL divergence: The KL divergence offers a useful summary of how far each country's partner distribution deviates from a uniform allocation over its neighbours.
  - Core countries (United States, Germany, France, China, etc.) typically have KL divergence in the range 1.5–2.1.
  - Small, highly concentrated economies often exhibit KL values between 3 and 4.

Phase-by-phase qualitative changes.

- Degrees deg and $\deg^+$ move closer to the upper bound over phases. As time passes, the network becomes closer to a complete graph, and China, in particular, transitions from a relatively less connected node to a global hub very rapidly.
  - For example:
    * United States: deg = 134 (Phase 1) $\rightarrow$ 141 (Phase 2) $\rightarrow$ 145 (Phase 3) $\rightarrow$ 146 (Phase 4).
    * Germany: deg = 134 (Phase 1) $\rightarrow$ 141 (Phase 2) $\rightarrow$ 145 (Phase 3) $\rightarrow$ 146 (Phase 4).
    * China: deg = 80 (Phase 1, relatively low) $\rightarrow$ 141 (Phase 2, dramatic increase) $\rightarrow$ 145 (Phase 3) $\rightarrow$ 147 (Phase 4).
- Concentration (HHI, KL) over phases
  - For core countries such as the United States, Germany, France, and China, HHI is generally low, and both n_i^{HHI} and n_i^{E} are high across phases.
  - China stands out in Phases 2-3 with higher HHI (Phase 2: HHI $\simeq$ 0.17; Phase 3: HHI $\simeq$ 0.10), consistent with a hub that is still relatively tilted towards a subset of key partners. Among small economies, HHI and KL divergence do not fall dramatically over time, so the problem of partner concentration in peripheral countries remains unresolved, even as the global network thickens.
- Entropy-based measures
  - Core countries typically have $H_i \simeq 3.1$ - 3.6 ($\widetilde{H}_i = 0.62$ - 0.73), and n_i^{E} in the high-20s to high-30s.

– Peripheral countries have much lower entropy $H_i \simeq 1.4$ - 2.7 ($\widetilde{H}_i = 0.30$ - 0.55), and n_i^{E} around 4-15. Over the phases, some African and Latin American countries show modest increases in H_i and n_i^{E}, meaning they gradually diversify their partners; however, a sizeable core–periphery gap persists.

4.2.1 Relationships between networks

The main purpose of this subsection is to examine some measures comparing two connectivity matrices W^{Phase} and $W^{\text{Phase}\prime}$. This comparison can be summarized by the difference:

$$\Delta W = W^{\text{Phase}} - W^{\text{Phase}\prime}.$$

Let Δw_{ij} denote each element of ΔW.

First, different matrix norms can capture different aspects of how two connectivity networks differ.

- Frobenius norm $\|\Delta W\|_F = \left(\sum_{i=1}^n \sum_{j=1}^n (\Delta w_{ij})^2\right)^{\frac{1}{2}}$: This measure captures the overall/global difference between two networks by treating all entries symmetrically. Since this measure is based on squared differences, it places heavy weight on large discrepancies.

- Column sum norm $\|\Delta W\|_1 = \max_{j=1,\cdots,n} \sum_{i=1}^n |\Delta w_{ij}|$: In a row-normalized connectivity matrix, the column j $(w_{1j}, w_{2j}, \cdots, w_{nj})$ represents how influential origin j is to all destinations. Hence, $\|\Delta W\|_1$ captures which origin country experiences the largest change in how much other countries rely on it.

- Row sum norm $\|\Delta W\|_\infty = \max_{i=1,\cdots,n} \sum_{j=1}^n |\Delta w_{ij}|$: In a row-normalized connectivity matrix, each row i indicates how destination i distribute relevance across origins. Hence, $\|\Delta W\|_\infty$ highlights which destination countries experience the largest change in their inbound structure.

For example, if $\|\Delta W\|_F$ is minor but $\|\Delta W\|_\infty$ is large, the networks are similar overall, but some destinations have dramatic changes. On the other hand, if $\|\Delta W\|_1$ is large but $\|\Delta W\|_F$ is modest, most of the connectivity networks remain stable, but some origins experience significant changes.

- Jaccard coefficient: This measure is defined by

$$J_{\text{Phase},\text{Phase}\prime} = \frac{\#\left(\mathcal{E}_{\text{Phase}} \cap \mathcal{E}_{\text{Phase}\prime}\right)}{\#\left(\mathcal{E}_{\text{Phase}} \cup \mathcal{E}_{\text{Phase}\prime}\right)},$$

where $\mathcal{E}_{\text{Phase}}$ denotes a set of edges of W_{Phase}. Hence, the Jaccard coefficient represents the topological similarity, while the similarity measures based on the matrix norms capture the weight similarity.

Table 5: Relationships among the connectivity networks across phases

Panel A. Relationships among the connectivity networks across phases via three norms

		$W^{\text{Phase}=1}$	$W^{\text{Phase}=2}$	$W^{\text{Phase}=3}$	$W^{\text{Phase}=4}$
$\|\Delta W\|_F$	$W^{\text{Phase}=2}$	2.3728 (0.0159)	0	*	*
$\|\Delta W\|_1$	$W^{\text{Phase}=2}$	7.5036 (0.0502)	0	*	*
$\|\Delta W\|_\infty$	$W^{\text{Phase}=2}$	1.5714 (0.0105)	0	*	*
$\|\Delta W\|_F$	$W^{\text{Phase}=3}$	2.8813 (0.0193)	1.9212 (0.0129)	0	*
$\|\Delta W\|_1$	$W^{\text{Phase}=3}$	9.9249 (0.0664)	5.1008 (0.0341)	0	*
$\|\Delta W\|_\infty$	$W^{\text{Phase}=3}$	1.5342 (0.0103)	1.6170 (0.0108)	0	*
$\|\Delta W\|_F$	$W^{\text{Phase}=4}$	3.6703 (0.0246)	2.8833 (0.0193)	1.9474 (0.0130)	0
$\|\Delta W\|_1$	$W^{\text{Phase}=4}$	15.1642 (0.1014)	13.1163 (0.0877)	8.6279 (0.0577)	0
$\|\Delta W\|_\infty$	$W^{\text{Phase}=4}$	1.7240 (0.0115)	1.6314 (0.0109)	1.2872 (0.0086)	0

Panel B. Jaccard coefficients

	$W^{\text{Phase}=1}$	$W^{\text{Phase}=2}$	$W^{\text{Phase}=3}$	$W^{\text{Phase}=4}$
$W^{\text{Phase}=2}$	0.6949	1.0000	0.8769	0.8659
$W^{\text{Phase}=3}$	0.6412	0.8769	1.0000	0.9465
$W^{\text{Phase}=4}$	0.6341	0.8659	0.9465	1.0000

Panel A of Table 5 summarizes the distance between the connectivity matrices across phases using the Frobenius, 1-, and ∞-norms of ΔW. The Frobenius norm, which captures the overall Euclidean distance between two networks, increases monotonically as phases become further apart (e.g., from 2.37 between Phases 1 and 2 to 3.67 between Phases 1 and 4; corresponding per-country averages are 0.016 and 0.025), indicating a gradual but non-trivial drift in the global structure of trade connectivity. The 1-norm, which is sensitive to changes in the columns of W and therefore to the outbound influence of origin countries, grows more sharply—especially in comparisons involving Phase 4—suggesting that a subset of origins substantially reallocated their relative importance in the network over time. By contrast, the ∞-norm, which reflects changes in the rows of W and thus in the inbound exposure of destination countries, remains in a narrower range (around 1.3–1.7), implying that destination-side sourcing patterns adjusted more moderately and in a more diffuse manner. Overall, the network appears far from static, but its evolution is gradual and driven primarily by changes on the origin side rather than by abrupt shifts in the import portfolios of destination countries.

From Panel B of Table 5, we observe that the transition from Phase 1 to Phase 2 already features a substantial reconfiguration of the connectivity network. The Jaccard coefficient of about 0.69 indicates that roughly 70% of the links present in Phase 1 are preserved in Phase

2, while the remaining links are either created or severed. At the same time, the Frobenius distance of 2.37 suggests a sizeable reweighting of the surviving links. Hence, Phase 1 to Phase 2 can be interpreted as an initial adjustment period in which both the topology and the intensities of connections are noticeably restructured, before the network becomes more stable in later phases. For other transitions, the Jaccard indices remain relatively high for adjacent later phases (e.g. 0.88 between Phases 2 and 3 and 0.95 between Phases 3 and 4), implying that the set of trading relationships stabilises after the early period and that most of the subsequent adjustment occurs through re-weighting existing links rather than creating or severing connections. The comparison between Phases 1 and 4, with a lower Jaccard coefficient of about 0.63 and the largest Frobenius distance, shows that both the topology and the associated trade intensities have substantially evolved relative to the initial network.

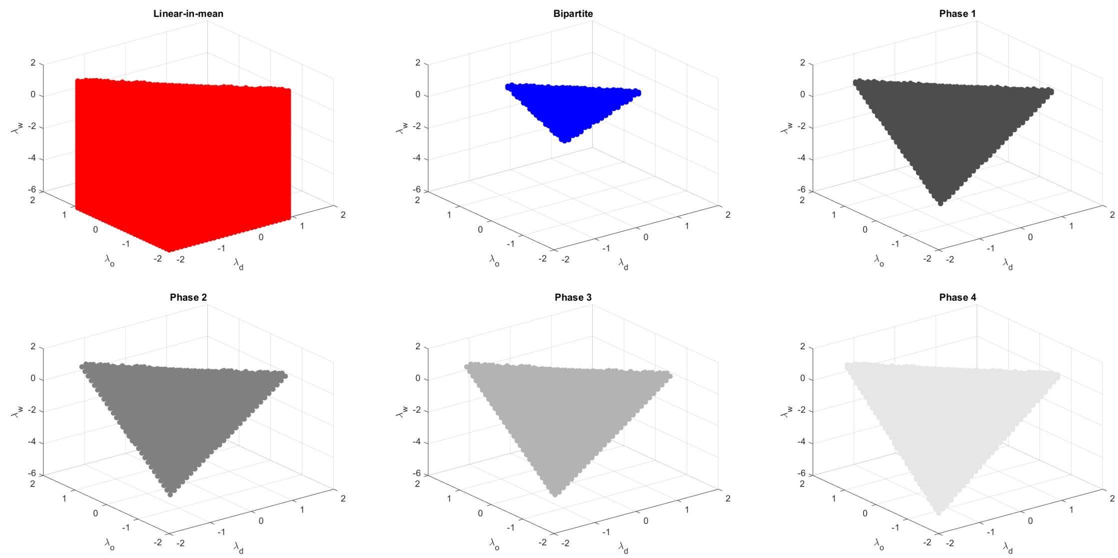
Figure 4 visualizes the admissible parameter spaces for λ across different network structures and phases. In each panel, the shaded region collects the values of λ for which the stability condition ($\rho_{\text{spec}}(\mathbf{A}) < 1$) holds. Our empirical estimates for λ lie well inside these regions in all phases, indicating that the stability constraint is not binding in practice. 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.

4.3 Additional coefficient interpretations

Beyond the network parameters, the coefficients on the standard gravity covariates have largely expected signs and are reasonably stable across phases.

The mean *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

Figure 4: Admissible parameter spaces for λ



Note: The figures above show the admissible parameter space for λ stated in (2.3) across different network structures: Panel (a): *linear-in-means* network, Panel (b): *bipartite* network, Panel (c): Phase 1 (1986, trade liberalization), Panel (d): Phase 2 (1997, active NAFTA implementation), Panel (e): Phase 3 (2007, emergence of the China trade shock), and Panel (f): Phase 4 (2016, expansion of global supply chains).

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. The incidence of regional trade agreements (*FTA*) rises gradually across phases, from about 0.04% in 1986 to about 1.1% in 2016, capturing the growing prevalence of formal trade cooperation.

The coefficient on *Distance* is negative and statistically significant in all four phases, with particularly precise and sizeable effects from Phase 2 onward. This confirms that geographic separation continues to impose substantial trade costs even in an increasingly interconnected world economy. The magnitude of the distance elasticity becomes larger in absolute value over time, especially in Phase 4, suggesting that despite technological advances and reduced communication costs, spatial frictions remain a first–order determinant of international trade patterns.

The *Border* variable is positive and significant in Phases 1, 2, and 4, indicating that countries sharing a common border trade more intensively than others, likely due to reduced transportation costs and various forms of institutional proximity. In Phase 3, however, the border effect becomes small and statistically insignificant, suggesting that the competitive pressures associated with the China trade shock may have partially offset the traditional advantages of geographic contiguity in that period. Sharing the same *Legal* system increases bilateral trade volumes in all phases, reflecting the role of institutional similarity in lowering transaction costs and facilitating contract enforcement.

The impact of *Language* is modest and statistically insignificant in the earlier and final phases, but becomes positive and statistically significant in Phase 3. This pattern implies that cultural and informational frictions gained particular importance around the period of intensified global competition associated with the China trade shock, when firms expanded more aggressively into diverse and distant markets and relied more on shared language to mitigate informational barriers.

Other structural and historical factors show heterogeneous patterns. The effect of *Colony* is small in magnitude and not robustly significant across phases, suggesting that historical colonial ties have weakened as a determinant of trade once more recent forms of integration and institutional similarity are taken into account. By contrast, the coefficient on *FTA* is positive and statistically significant in all four phases, confirming the trade–creating effect of regional trade agreements (including NAFTA and other arrangements) in our sample. The positive and persistent influence of *FTA* underscores the continued relevance of policy–driven integration alongside endogenous network formation captured by the λ parameters.

Geographic constraints, captured by the *Islands* and *Landlock* variables, exhibit unstable and sometimes extreme coefficient estimates across phases. In particular, their magnitudes and, in some phases, extremely small estimated standard errors are suggestive of quasi-complete separation or very limited within–group variation. As a result, the phase–specific coefficients on these indicators should be interpreted with caution. Rather than emphasizing these estimates, we view island and landlocked status as characteristics that are largely absorbed by the origin and destination fixed effects and by the distance measure.

Finally, the *Currency* variable displays mixed signs and is not statistically significant in most phases, consistent with the limited prevalence of common–currency arrangements in the sample and with the possibility that much of the effect of monetary integration is captured by other institutional or regional controls already included in the specification.

4.4 Counterfactual simulations

In this subsection, we provide details on the counterfactual simulations. The purpose of counterfactual analyses is to compare the trade flows from two scenarios: (i) the parameter estimates with the specified connectivity network ($\widehat{\boldsymbol{\mu}}$) (ii) counterfactual parameters or hypothetical connectivity network ($\widetilde{\boldsymbol{\mu}}$). Let $\boldsymbol{\mu}\left(\lambda, \boldsymbol{\phi}, W\right)$ denote the vector of the predicted trade flows evaluated at $(\lambda, \boldsymbol{\phi}, W)$. Mathematically, we studey the gap between $\widehat{\boldsymbol{\mu}} = \boldsymbol{\mu}\left(\widehat{\lambda}, \widehat{\boldsymbol{\phi}}, W\right)$ and $\widetilde{\boldsymbol{\mu}} = \boldsymbol{\mu}\left(\widetilde{\lambda}, \widehat{\boldsymbol{\phi}}, \widetilde{W}\right)$, where $\widetilde{\lambda}$ denotes the counterfactual network interaction parameter and $\widetilde{W}$ denotes the counterfactual connectivity network.

4.4.1 Designs

1. Network utilization. The first counterfactual scenario describes the trade flows when countries do not utilize information in the connectivity matrix. In our model framework, this scenario can be represented by $\lambda_d = \lambda_o = \lambda_w = 0$. In other words, we recompute the equilibrium trade flows under a scenario where countries do not exploit network-based spillover channels in managing trade costs, holding the underlying gravity structure fixed.

2. Changes in the network structures. Roughly, our model's main premitives can be categorized into two components: (i) behavioral parameters λ and (ii) connectivity structure W. In the second counterfactual analysis, we examine the trade patterns under a different connectivity structure with keeping the estimated behavioral parameters λ. For example, we can examine the trade patterns after the China trade shock if the Phase 2 behaviors (NAFTA) are maintained.

3. Changes in the behavioral parameters. On the other hand, we can consider different behavioral parameters under the fixed connectivity structure.

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.

Dugoua, E., Dumas, M., and Noailly, J. (2022). Text as data in environmental economics and policy. *Review of Environmental Economics and Policy*, 16(2):346–356.

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.

Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3):535–574.

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.

Hoberg, G. and Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of political economy*, 124(5):1423–1465.

Horn, R. A. and Johnson, C. R. (1985). *Matrix Analysis*. Cambridge University Press.

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.

Ke, Z. T., Kelly, B. T., and Xiu, D. (2019). Predicting returns with text data. *National Bureau of Economic Research*.

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.

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.

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Structural Analysis in the Social Sciences. Cambridge University Press.