Connected Trade Flows via Trade Cost: Spatial Autoregressive Framework

Hanbat Jeong* and Jieun Lee†

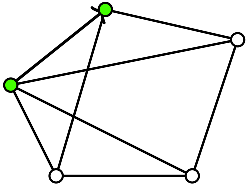
* Department of Economics, Macquarie Business School, Macquarie University
† Department of Economics, Emory University

Presentation Slides

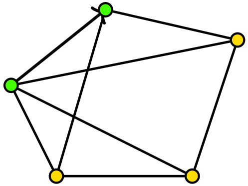
1. Introduction

Motivation - What we do?

- Our paper develops econometric model specificaton and estimation/inference for analyzing forces, signals, and flows within a *system*.
  - "system": subject to connectivity/networks

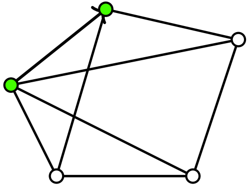


Without considering the system

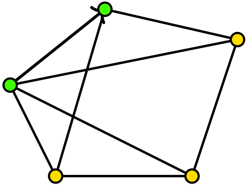


With considering the system

- Our paper develops econometric model specificaton and estimation/inference for analyzing forces, signals, and flows within a *system*.
    - "system": subject to connectivity/networks



Without considering the system



With considering the system

- Examples of economic flows: **Trade flows**, migration flows, commuting flows

- **Gravity equations** are the workhorse for explaining **trade flows**.

- **Gravity equations** are the workhorse for explaining **trade flows**.
- Specifying **trade costs** is essential.
  - Why? Shape the composition of partners and the quantities exchanged.
  - **Logistics** serve as a fundamental source of economic frictions.

- **Gravity equations** are the workhorse for explaining **trade flows**.
- Specifying **trade costs** is essential.
  - Why? Shape the composition of partners and the quantities exchanged.
  - **Logistics** serve as a fundamental source of economic frictions.
- The *iceberg-cost* specification (Samuelson 1952, 1954)—under which a fraction of the shipped good "melts" in transit—has become the default,
  - enters trade shares multiplicatively (and is therefore tractable).
  - This formulation treats trade costs as largely **exogenous** and unavoidable.

- Question: Do countries' network-leveraging behaviors significantly explain the residuals from iceberg trade-cost models?

- Question: Do countries' network-leveraging behaviors significantly explain the residuals from iceberg trade-cost models?
  - Yes. Model fit significantly improves: McFadden's R^2 28% (>20%).

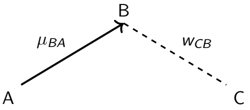
- Question: Do countries' network-leveraging behaviors significantly explain the residuals from iceberg trade-cost models?
    - Yes. Model fit significantly improves: McFadden's R^2 28% (>20%).
- Implication: How do the network spillovers reshape the global trade?
    - How do bilateral trade policy shocks (e.g., tariffs, sanctions, or supply-chain disruptions) propagate through the trade network?
    - How this affects *third-country* trade patterns?

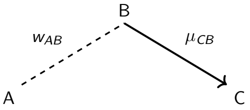
- Question: Do countries' network-leveraging behaviors significantly explain the residuals from iceberg trade-cost models?
  - Yes. Model fit significantly improves: McFadden's R^2 28% (>20%).
- Implication: How do the network spillovers reshape the global trade?
  - How do bilateral trade policy shocks (e.g., tariffs, sanctions, or supply-chain disruptions) propagate through the trade network?
  - How this affects *third-country* trade patterns?
- Some notes
  - Two types of networks
    1. Expected trade flows
    2. Countries' geographic/economic proximities
  - Cross-sectional analysis

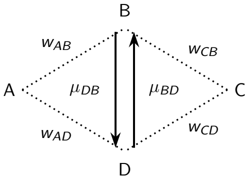
1. Introduction

Core Idea

- Case 1 (Cross-destination linkage): Suppose B and C are connected. Suppose A wants to export to C and A knows the expected trade flow from A to B.
- Case 2 (Cross-origin linkage): Suppose A and B are connected. Suppose A wants to export to C and A knows the expected trade flow from B to C.

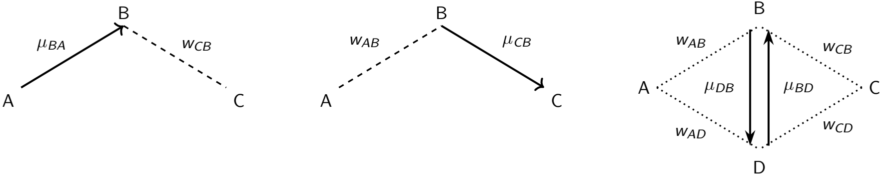






- Q. How would A account for trade cost exporting to C?

- Case 1 (Cross-destination linkage): Suppose B and C are connected. Suppose A wants to export to C and A knows the expected trade flow from A to B.
- Case 2 (Cross-origin linkage): Suppose A and B are connected. Suppose A wants to export to C and A knows the expected trade flow from B to C.

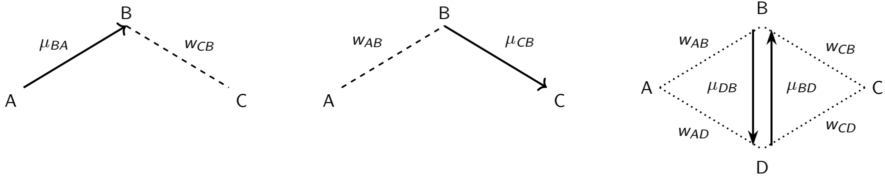


- Q. How would A account for trade cost exporting to C?
- A1. A would think if it's worth routing through B when exporting to C.

1. Introduction

Core Idea

- Case 1 (Cross-destination linkage): Suppose B and C are connected. Suppose A wants to export to C and A knows the expected trade flow from A to B.
- Case 2 (Cross-origin linkage): Suppose A and B are connected. Suppose A wants to export to C and A knows the expected trade flow from B to C.

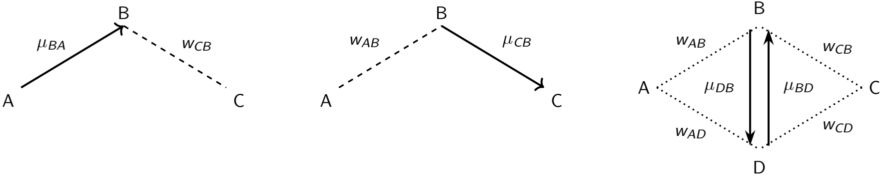


- Q. How would A account for trade cost exporting to C?
- A1. A would think if it's worth routing through B when exporting to C.
- A2. A would think if it's worth co-shipping with B when exporting to C.

1. Introduction

Core Idea

- Case 1 (Cross-destination linkage): Suppose B and C are connected. Suppose A wants to export to C and A knows the expected trade flow from A to B.
- Case 2 (Cross-origin linkage): Suppose A and B are connected. Suppose A wants to export to C and A knows the expected trade flow from B to C.



- Q. How would A account for trade cost exporting to C?
- A1. A would think if it's worth routing through B when exporting to C.
- A2. A would think if it's worth co-shipping with B when exporting to C.
  - The expected trade volume from B to C is important. (If not big, A will less likely leverage B when exporting to C.)
  - If leveraging B, how close with B is also important.

- Countries may leverage their **trade networks** as a *resource* to operate efficient trade costs scheme.
  - (i) expected trade flows and (ii) the geographic/economic ties

- Countries may leverage their **trade networks** as a *resource* to operate efficient trade costs scheme.
  - (i) expected trade flows and (ii) the geographic/economic ties

$\Rightarrow$ *Interdependent trade flows* via specifying a cost function.

- Countries may leverage their **trade networks** as a *resource* to operate efficient trade costs scheme.
  - (i) expected trade flows and (ii) the geographic/economic ties

$\Rightarrow$ *Interdependent trade flows* via specifying a cost function.

  - Trade costs emerge *endogenously* from trade networks, rather than remaining exogenous.
    - i.e., a country's trade costs *depend* on the expected trade flows of others with countries' connectivities.
  - This feature is captured by the **spatial autoregressive (SAR)** framework.

- Countries may leverage their **trade networks** as a *resource* to operate efficient trade costs scheme.
  - (i) expected trade flows and (ii) the geographic/economic ties

$\Rightarrow$ *Interdependent trade flows* via specifying a cost function.
  - Trade costs emerge *endogenously* from trade networks, rather than remaining exogenous.
    - ⋆ i.e., a country's trade costs *depend* on the expected trade flows of others with countries' connectivities.
  - This feature is captured by the **spatial autoregressive (SAR)** framework.

- Note. The iceberg costs are a special case of our framework.

1. Introduction

Role of **Spatial Autoregressive (SAR)** Model

- SAR represents interdependence: $y_i = w_{i1}y_1 + w_{i2}y_2 + \cdots + w_{in}y_n + \ldots$
  - y_i: i's outcome, $w_{i1}y_1 + w_{i2}y_2 + \cdots + w_{in}y_n$: neighbors' outcome
  - Can accommodate endogenous trade costs scheme

Role of **Spatial Autoregressive (SAR)** Model

- SAR represents interdependence: $y_i = w_{i1}y_1 + w_{i2}y_2 + \cdots + w_{in}y_n + \ldots$
  - y_i: i's outcome, $w_{i1}y_1 + w_{i2}y_2 + \cdots + w_{in}y_n$: neighbors' outcome
  - Can accommodate endogenous trade costs scheme
- SAR represents *network-induced* pair-specific heterogeneity
  - Multilateral resistance traditionally reflects average trade barriers across *all* partners, modeled by individual fixed effects.
  - Our framework *extends* this concept by allowing for interdependence within these terms.
  - The resulting *pair-specific heterogeneity* is thus inherently relational, manifesting at the *pair* level.

1. Introduction

Related literature - Contributions

Our paper is about **Econometric model specification.**

Our paper is about **Econometric model specification.**

1. [Why] We **endogenize trade costs** through the lens of trade networks as a resource in operating logistics-related trade costs.
   - Improve model fit by reducing residuals through econometric model specification.
   - Structural gravity equation for trade flows/iceberg trade cost specification
     - Krugman (1995), Eaton and Kortum (2002 *Ecta*), Anderson and van Wincoop (2003 *AER*), Arkolakis et al. (2012 *AER*), Tyazhelnikov and Romalis (2024 *JIE*)

1. Introduction

Related literature - Contributions

Our paper is about **Econometric model specification.**

1. [Why] We **endogenize trade costs** through the lens of trade networks as a resource in operating logistics-related trade costs.
    - Improve model fit by reducing residuals through econometric model specification.
    - Structural gravity equation for trade flows/iceberg trade cost specification
        - ★ Krugman (1995), Eaton and Kortum (2002 *Ecta*), Anderson and van Wincoop (2003 *AER*), Arkolakis et al. (2012 *AER*), Tyazhelnikov and Romalis (2024 *JIE*)

2-1. [How] From this motivation, we characterize **interdependent trade flows**.
- Interdependence from different motivations
    - ★ Allen et al. (2020 *JPE*), Lind and Ramondo (2023 *AER*)
- Spatial autoregressive framework
    - ★ Cliff and Ord (1995), Ord (1975 *JASA*), Lee (2004 *Ecta*, 2007 *JoE*)
    - ★ **LeSage and Pace (2008 *JRS*)**
    - ★ Behrens et al. (2012 *JAE*), Pesaran and Yang (2021 *JoE*), Jin et al. (2023 *ER*), Jeong et al. (2023, *EL*), Jeong and Lee (2024 *JoE*)

Our paper is about **Econometric model specification.**

1. [Why] We **endogenize trade costs** through the lens of trade networks as a resource in operating logistics-related trade costs.
   - Improve model fit by reducing residuals through econometric model specification.
   - Structural gravity equation for trade flows/iceberg trade cost specification
     - ★ Krugman (1995), Eaton and Kortum (2002 *Ecta*), Anderson and van Wincoop (2003 *AER*), Arkolakis et al. (2012 *AER*), Tyazhelnikov and Romalis (2024 *JIE*)

2-1. [How] From this motivation, we characterize **interdependent trade flows**.
   - Interdependence from different motivations
     - ★ Allen et al. (2020 *JPE*), Lind and Ramondo (2023 *AER*)
   - Spatial autoregressive framework
     - ★ Cliff and Ord (1995), Ord (1975 *JASA*), Lee (2004 *Ecta*, 2007 *JoE*)
     - ★ **LeSage and Pace (2008 *JRS*)**
     - ★ Behrens et al. (2012 *JAE*), Pesaran and Yang (2021 *JoE*), Jin et al. (2023 *ER*), Jeong et al. (2023, *EL*), Jeong and Lee (2024 *JoE*)
   - For estimation, we developed the Poisson pseudo-maximum likelihood estimation with network effects.
   - For inference, spatial HAC estimator.

2-2. [Innovation] Efficient computation method

- Network multiplier matrix: need to calculate an *inverse* of a huge matrix
  - ⋆ Makes computational cost heavy

2-2. [Innovation] Efficient computation method

- Network multiplier matrix: need to calculate an *inverse* of a huge matrix
  - ⋆ Makes computational cost heavy
- Based on the spectral decomposition of a network matrix, our algorithm
  - ⋆ yields a linear equation of the network multiplied quantities;
  - ⋆ and eventually leads to a closed form solution.
  - ⋆ i.e., We don't have to calculate an inverse of the matrix at all!

2-2. [Innovation] Efficient computation method

- Network multiplier matrix: need to calculate an *inverse* of a huge matrix
  - ★ Makes computational cost heavy
- Based on the spectral decomposition of a network matrix, our algorithm
  - ★ yields a linear equation of the network multiplied quantities;
  - ★ and eventually leads to a closed form solution.
  - ★ i.e., We don't have to calculate an inverse of the matrix at all!
- Computation time comparison

n	$N = n^2$	Default (A)	Ours (B)	A/B
9	81	0.5668	0.3518	1.6112
25	625	12.3862	0.3967	31.2265
49	2401	1252.0077	3.9725	315.1687
64	4096	9465.7145	9.2816	1019.8380

3. Empirical analysis
   - Four key phases of global trade
     1. Phase 1 (1986, trade liberalization)
     2. Phase 2 (1997, active NAFTA implementation)
     3. Phase 3 (2007, emergence of the China trade shock)
     4. Phase 4 (2016, expansion of global value chains)

3. Empirical analysis
    - Four key phases of global trade
        1. Phase 1 (1986, trade liberalization)
        2. Phase 2 (1997, active NAFTA implementation)
        3. Phase 3 (2007, emergence of the China trade shock)
        4. Phase 4 (2016, expansion of global value chains)
    - Network analysis
        - ★ Trade networks became more interconnected and denser, deviating from a bipartite structure

Related literature - Contributions

3. Empirical analysis
   - Four key phases of global trade
     1. Phase 1 (1986, trade liberalization)
     2. Phase 2 (1997, active NAFTA implementation)
     3. Phase 3 (2007, emergence of the China trade shock)
     4. Phase 4 (2016, expansion of global value chains)
   - Network analysis
     - ★ Trade networks became more interconnected and denser, deviating from a bipartite structure
   - Estimation results
     - ★ Phase 1: competition dominating in early liberalization
     - ★ Phase 2: complementarity rising under NAFTA
     - ★ Phase 3: renewed substitution emerging during the China trade shock
     - ★ Phase 4: shifted to complementarity under mature global value chains

3. Empirical analysis
   - Four key phases of global trade
     1. Phase 1 (1986, trade liberalization)
     2. Phase 2 (1997, active NAFTA implementation)
     3. Phase 3 (2007, emergence of the China trade shock)
     4. Phase 4 (2016, expansion of global value chains)
   - Network analysis
     - ★ Trade networks became more interconnected and denser, deviating from a bipartite structure
   - Estimation results
     - ★ Phase 1: competition dominating in early liberalization
     - ★ Phase 2: complementarity rising under NAFTA
     - ★ Phase 3: renewed substitution emerging during the China trade shock
     - ★ Phase 4: shifted to complementarity under mature global value chains
   - Counterfactual experiments and their policy implications
     - ★ How conflicts between major economies (e.g., US–China trade war) spill over to third countries and reshape global trade structures

1. Specification
   - Microfoundation
   - Econometric Model specification
2. Estimation/Inference
   - Poisson Pseudo Maximum Likelihood Estimation Under Network Influence
   - Asymptotic Distribution - Main and fixed-effect parameters
3. Monte Carlo Simulations
4. Empirical Application
   - Basic setup/Network statistics
   - Estimation results
   - Counterfactual experiments

Assumption (Spatial Setting)

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

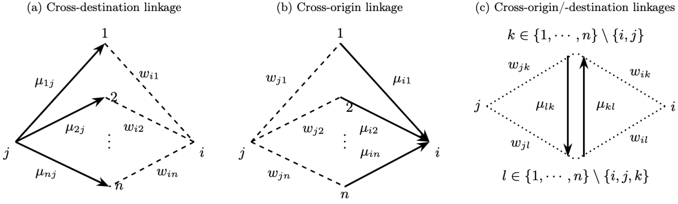
Assumption (Spatial Setting)

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

- Trade flow from j to i, y_{ij}, is generated by two locations, origin j and destination i.

$\Rightarrow$ There are n locations in a sample, so $N = n^2$ flow outcomes are observed.

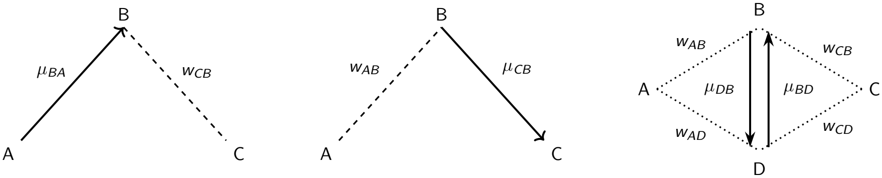
- Each w_{ij} denotes proximity between i and j.
  - Note. In practice, W can be constructed using historical trade flows.
    - ★ w_{ij} = long-run relationship between i and j
  - Free from the construction issue of W, thanks to the nature of network data.



- **Stage 1 (partner (hub) selection):** For each ν, i forms a probability distribution over potential partners $k \in \{1, \cdots, n\} \setminus \{i\}$, and j forms a probability distribution over $l \in \{1, \cdots, n\} \setminus \{j\}$:

$$\Pr(k \text{ is } i\text{'s partner at } \nu) = w_{ik}^d \text{ and } \Pr(l \text{ is } j\text{'s partner at } \nu) = w_{jl}^o.$$

  - For empirical application, we adopt a single set of proximity weights and impose $W^d = W^o = W$.



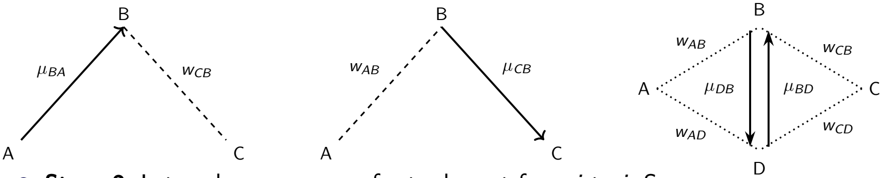
- **Stage 2:** Let π_{CA} be a measure of a trade cost from A to C. Suppose

$$\pi_{CA}(\boldsymbol{\mu}) = \left(\underbrace{\left(\mu_{BA}^{w_{CB}}\right)^{\tilde{\lambda}_d}}_{\substack{\text{outflows from } A \\ \text{cross-destination linkage}}} \times \underbrace{\left(\mu_{CB}^{w_{AB}}\right)^{\tilde{\lambda}_o}}_{\substack{\text{inflows to } C \\ \text{cross-origin linkage}}} \times \underbrace{\left(\cdots\left(\mu_{DB}^{w_{AB}w_{CD}}\right)\cdot\left(\mu_{BD}^{w_{AD}w_{CB}}\right)\cdots\right)^{\tilde{\lambda}_w}}_{\substack{\text{flows among third-party units} \\ \text{cross-origin and cross-destination linkage}}} \right)^{-1} \times \underbrace{D_{CA}^{\tilde{\beta}}}_{\substack{\text{bilateral} \\ \text{characteristics}}}$$

  - μ_{ij}: expected trade flow y_{ij}
  - $\tilde{\lambda}_d$, $\tilde{\lambda}_o$, $\tilde{\lambda}_w$, $\tilde{\beta}$: structural parameters

2. Specification

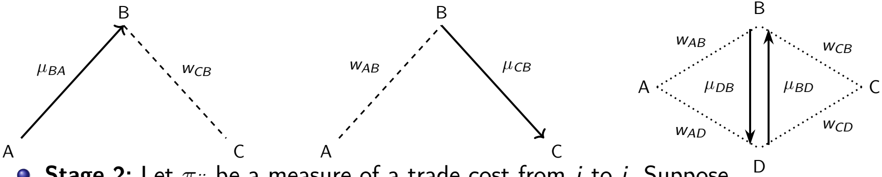
Microfoundation



- **Stage 2:** Let π_{ij} be a measure of a trade cost from j to i. Suppose

$$\pi_{ij}(\boldsymbol{\mu}) = \left(\underbrace{\left(\prod_{k=1}^{n} \mu_{kj}^{w_{ik}}\right)^{\tilde{\lambda}_d}}_{\substack{\text{outflows from } j \\ \text{cross-destination linkage}}} \times \underbrace{\left(\prod_{l=1}^{n} \mu_{il}^{w_{jl}}\right)^{\tilde{\lambda}_o}}_{\substack{\text{inflows to } i \\ \text{cross-origin linkage}}} \times \underbrace{\left(\prod_{k,l=1}^{n} \mu_{kl}^{w_{ik} w_{jl}}\right)^{\tilde{\lambda}_w}}_{\substack{\text{flows among third-party units} \\ \text{cross-origin and cross-destination linkage}}} \right)^{-1} \times \underbrace{D_{ij}^{\tilde{\beta}}}_{\substack{\text{bilateral} \\ \text{characteristics}}}$$

  - μ_{ij}: expected trade flow y_{ij}
  - $\tilde{\lambda}_d$, $\tilde{\lambda}_o$, $\tilde{\lambda}_w$, $\tilde{\beta}$: structural parameters



- **Stage 2:** Let π_{ij} be a measure of a trade cost from j to i. Suppose

$$\pi_{ij}(\boldsymbol{\mu}) = \underbrace{\left(\underbrace{\left(\prod_{k=1}^{n} \mu_{kj}^{w_{ik}}\right)^{\tilde{\lambda}_d}}_{\substack{\text{outflows from } j \\ \text{cross-destination linkage}}} \times \underbrace{\left(\prod_{l=1}^{n} \mu_{il}^{w_{jl}}\right)^{\tilde{\lambda}_o}}_{\substack{\text{inflows to } i \\ \text{cross-origin linkage}}} \times \underbrace{\left(\prod_{k,l=1}^{n} \mu_{kl}^{w_{ik}w_{jl}}\right)^{\tilde{\lambda}_w}}_{\substack{\text{flows among third-party units} \\ \text{cross-origin and cross-destination linkage}}} \right)^{-1}}_{=:\pi_{ij}^{e}(\boldsymbol{\mu}) \text{ (endogenous part)}} \times \underbrace{\underbrace{D_{ij}^{\tilde{\beta}}}_{\substack{\text{bilateral} \\ \text{characteristics}}}}_{=:\pi_{ij}^{+} \text{ (exogenous part)}}$$

Stage 3: Given $\pi_{ij}(\boldsymbol{\mu})$ from **Stage 2**, the optimal trade flows are determined as in Anderson and van Wincoop (2003).

- At equilibrium, the total trade flow from exporter j to importer i is

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

  - G_i: i's budget (exogenously given)
  - G^W: World budget
  - $\Pi_j(\boldsymbol{\mu})$ and $P_i(\boldsymbol{\mu})$: multilateral resistance terms at $\boldsymbol{\mu}$.
  - ϱ: CES elasticity

Stage 3: Given $\pi_{ij}(\boldsymbol{\mu})$ from **Stage 2**, the optimal trade flows are determined as in Anderson and van Wincoop (2003).

- At equilibrium, the total trade flow from exporter j to importer i is

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

  - G_i: i's budget (exogenously given)
  - G^W: World budget
  - $\Pi_j(\boldsymbol{\mu})$ and $P_i(\boldsymbol{\mu})$: multilateral resistance terms at $\boldsymbol{\mu}$.
  - ϱ: CES elasticity

- Plugging $\pi_{ij}(\boldsymbol{\mu})$ from **Stage 2**,

$$\mu_{ij} = \underbrace{\left(\pi_{ij}^e(\boldsymbol{\mu})\right)^{1-\varrho}}_{\substack{\text{explicitly}\\ \text{endogenous term}}} \cdot \underbrace{P_i^{\varrho-1}(\boldsymbol{\mu})\Pi_j^{\varrho-1}(\boldsymbol{\mu})}_{\substack{\text{implicitly}\\ \text{endogenous term}}} \cdot \underbrace{G_i G_j (G^W)^{-1}(\pi_{ij}^+)^{1-\varrho}}_{\substack{\text{purely}\\ \text{exogenous term}}},$$

  - $(\pi_{ij}^e(\boldsymbol{\mu}))^{1-\varrho} = \left(\prod_{k=1}^n \mu_{kj}^{w_{ik}}\right)^{\lambda_d}\left(\prod_{l=1}^n \mu_{il}^{w_{jl}}\right)^{\lambda_o}\left(\prod_{k,l=1}^n \mu_{kl}^{w_{ik}w_{jl}}\right)^{\lambda_w}$,
    where $\lambda_d = (\varrho-1)\tilde{\lambda}_d$, $\lambda_o = (\varrho-1)\tilde{\lambda}_o$, $\lambda_w = (\varrho-1)\tilde{\lambda}_w$.

2. Specification

Microfoundation

Assumption (Equilibrium uniqueness (i))

$\rho_{\text{spec}}(\mathbf{A}) < 1$, where $\mathbf{A}$ is an aggregate network matrix defined by

$$\mathbf{A} := \underbrace{\lambda_d (I_n \otimes W)}_{\substack{\text{cross-destination} \\ \text{linkage}}} + \underbrace{\lambda_o (W \otimes I_n)}_{\substack{\text{cross-origin} \\ \text{linkage}}} + \underbrace{\lambda_w (W \otimes W)}_{\substack{\text{cross-origin} \\ \text{and -destination linkages}}},$$

with $\lambda_d = (\varrho - 1)\tilde{\lambda}_d$, $\lambda_o = (\varrho - 1)\tilde{\lambda}_o$, $\lambda_w = (\varrho - 1)\tilde{\lambda}_w$.

Assumption (Equilibrium uniqueness (i))

$\rho_{\text{spec}}(\mathbf{A}) < 1$, where $\mathbf{A}$ is an aggregate network matrix defined by

$$\mathbf{A} := \underbrace{\lambda_d (I_n \otimes W)}_{\substack{\text{cross-destination} \\ \text{linkage}}} + \underbrace{\lambda_o (W \otimes I_n)}_{\substack{\text{cross-origin} \\ \text{linkage}}} + \underbrace{\lambda_w (W \otimes W)}_{\substack{\text{cross-origin} \\ \text{and -destination linkages}}},$$

with $\lambda_d = (\varrho - 1)\tilde{\lambda}_d$, $\lambda_o = (\varrho - 1)\tilde{\lambda}_o$, $\lambda_w = (\varrho - 1)\tilde{\lambda}_w$.

- Semi-reduced form: Under (i) and (ii), there is a unique $\boldsymbol{\mu}^*$ satisfying

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

where $x_{kl} = (\ln(D_{kl,1}), \cdots, \ln(D_{kl,K}))'$ and $\beta = (\beta_1, \cdots, \beta_K)'$ with $\beta_k = (1 - \varrho)\tilde{\beta}_k$.

Assumption (Equilibrium uniqueness (i))

$\rho_{\text{spec}}(\mathbf{A}) < 1$, where $\mathbf{A}$ is an aggregate network matrix defined by

$$\mathbf{A} := \underbrace{\lambda_d(I_n \otimes W)}_{\substack{\text{cross-destination}\\ \text{linkage}}} + \underbrace{\lambda_o(W \otimes I_n)}_{\substack{\text{cross-origin}\\ \text{linkage}}} + \underbrace{\lambda_w(W \otimes W)}_{\substack{\text{cross-origin}\\ \text{and -destination linkages}}},$$

with $\lambda_d = (\varrho - 1)\tilde{\lambda}_d$, $\lambda_o = (\varrho - 1)\tilde{\lambda}_o$, $\lambda_w = (\varrho - 1)\tilde{\lambda}_w$.

- Semi-reduced form: Under (i) and (ii), there is a unique $\boldsymbol{\mu}^*$ satisfying

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

where $x_{kl} = (\ln(D_{kl,1}),\cdots,\ln(D_{kl,K}))'$ and $\beta = (\beta_1,\cdots,\beta_K)'$ with $\beta_k = (1-\varrho)\tilde{\beta}_k$.

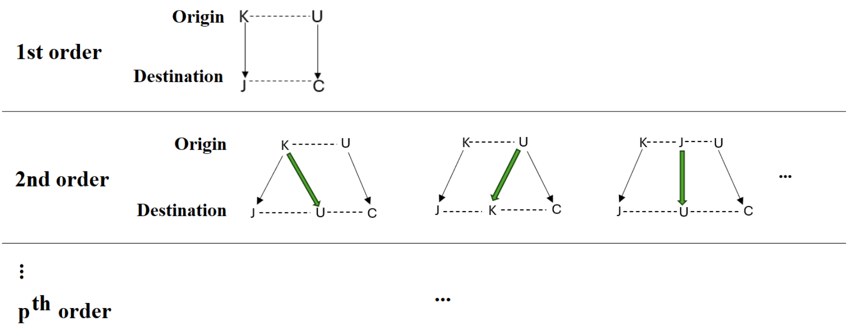
  - $\mathbf{S}^{-1}$: Network multiplier matrix
    - $s_{ij,kl} \equiv ((j-1)n + i, (l-1)n + k)$-element of $\mathbf{S}^{-1}$.
    - $\mathbf{S} = I_N - \mathbf{A}$: Network SAR operator (LeSage and Pace (2008) *JRS*)
  - *Network Multilateral Resistance* fixed effects: $\alpha_l(\boldsymbol{\mu}^*)$ and $\eta_k(\boldsymbol{\mu}^*)$

2. Specification

Microfoundation - Structure of Network Multiplier Matrix

Remark. This specification encompasses all intermediate flows that carry the signal $s_{ij,kl}$ between the flows (k, l) and (i, j) by p-th order, $p \in \mathbb{N}$.

Eg. The signal $s_{KJ,UC}$ between $(Korea, Japan)$ and $(US, China)$ is carried by inifinitely many intermediate flows as....



Assumption (Equilibrium uniqueness (ii))

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

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

In words, a small change of μ_{kl} does *not* yield dramatic changes in the network fixed effects, which include the multilateral resistance terms.

- $\alpha_q(\boldsymbol{\mu}) = \text{const.} + \ln\left(G_q\right) + \ln\left(\Pi_q^{\varrho-1}(\boldsymbol{\mu})\right)$ for $q = 1, \cdots, n$,
- $\eta_p(\boldsymbol{\mu}) = \text{const.} + \ln\left(G_p\right) + \ln\left(P_p^{\varrho-1}(\boldsymbol{\mu})\right)$ for $p = 1, \cdots, n$,
- $\Pi_j(\boldsymbol{\mu}) = \left(\sum_{i=1}^{n} \frac{G_i}{G^W} \left(\frac{\pi_{ij}(\boldsymbol{\mu})}{P_i(\boldsymbol{\mu})} \right)^{1-\varrho} \right)^{\frac{1}{1-\varrho}}$: Price index for origin j
- $P_i(\boldsymbol{\mu}) = \left(\sum_{j=1}^{n} \frac{G_j}{G^W} \left(\frac{\pi_{ij}(\boldsymbol{\mu})}{\Pi_j(\boldsymbol{\mu})} \right)^{1-\varrho} \right)^{\frac{1}{1-\varrho}}$: Price index for destination i

- From the semi-reduced form, the true data-generating process can be specified by

$$y_{ij} = \mu_{ij}^0 \times \xi_{ij}, \text{ where } \mu_{ij}^0 = \exp\left(\sum_{k,l=1}^{n} s_{ij,kl}\left(x_{kl}'\beta^0 + \alpha_l^0 + \eta_k^0\right)\right),$$

  - $\mu_{ij}^0 = \mathbb{E}(y_{ij}|\mathbf{z})$: Economic Model; ξ_{ij}: Error satisfying $\mathbb{E}(\xi_{ij}|\mathbf{z}) = 1$
  - $\mathbf{z}$ stands for a vector of exogenous characteristics
  - $u_{ij} = \mu_{ij}^0(\xi_{ij} - 1)$: additive error, $\mathbb{E}(u_{ij}|\mathbf{z}) = 0$
  - $\lambda^0 = \left(\lambda_d^0, \lambda_o^0, \lambda_w^0\right)'$ denotes a vector of the true network influence parameters
  - $\beta^0 = \left(\beta_1^0, \cdots, \beta_K^0\right)'$ is the true parameter for x_{kl}
  - α_j^0 and η_i^0 are the true origin- and destination- fixed effects, respectively.

- From the semi-reduced form, the true data-generating process can be specified by

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

  - $\mu^0_{ij} = \mathbb{E}(y_{ij}|\mathbf{z})$: Economic Model; ξ_{ij}: Error satisfying $\mathbb{E}(\xi_{ij}|\mathbf{z}) = 1$
  - $\mathbf{z}$ stands for a vector of exogenous characteristics
  - $u_{ij} = \mu^0_{ij}(\xi_{ij} - 1)$: additive error, $\mathbb{E}(u_{ij}|\mathbf{z}) = 0$
  - $\lambda^0 = \left(\lambda^0_d, \lambda^0_o, \lambda^0_w\right)'$ denotes a vector of the true network influence parameters
  - $\beta^0 = \left(\beta^0_1, \cdots, \beta^0_K\right)'$ is the true parameter for x_{kl}
  - α^0_j and η^0_i are the true origin- and destination- fixed effects, respectively.

- Remarks
  - The purpose of this model is to identify λ^0, β^0, $\alpha^0_1, \cdots, \alpha^0_n$, $\eta^0_1, \cdots,$ and η^0_n.
  - μ^0_{ij} $(i, j = 1, \cdots, n)$ can be identified by the semi-reduced form.

Estimation

- **Poisson pseudo maximum likelihood (PPML) estimator** (Gourieroux et al. 1984 *Ecta*) has become the *standard* since Santos Silva and Tenreyro (2006 *REStat*).
- Pseudo log-likelihood

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

where $\phi := (\boldsymbol{\alpha}', \boldsymbol{\eta}')'$.

- **Poisson pseudo maximum likelihood (PPML) estimator** (Gourieroux et al. 1984 *Ecta*) has become the *standard* since Santos Silva and Tenreyro (2006 *REStat*).
- Pseudo log-likelihood

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

where $\boldsymbol{\phi} := (\boldsymbol{\alpha}', \boldsymbol{\eta}')'$.

  - PPMLE: $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ell_N(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta', \boldsymbol{\phi}')'$
- First-order conditions: $\partial_\theta \ell_N(\boldsymbol{\theta}) = \sum_{i,j=1}^{n} \partial_\theta \tilde{\mu}_{ij}(\boldsymbol{\theta}) u_{ij}(\boldsymbol{\theta})$
  - Requirement: Correctly specifying $\mathbb{E}(y_{ij}|\mathbf{z})$ to be exponential conditional mean
- Can accommodate many zero ys $\Rightarrow$ No need to have $\ln(y+1)$

- Main issue here
  - Need to compute $\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta})$ for each $\boldsymbol{\theta}$.
  - Do we need to invert $\mathbf{S}(\lambda)$?
  - When there are $n = 150$ countries, $\mathbf{S}(\lambda)$ is an $150^2 \times 150^2$ matrix $(506, 250, 000$ elements!)

PPMLE - Efficient computation algorithm for $\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta})$

- Main issue here
  - Need to compute $\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta})$ for each $\boldsymbol{\theta}$.
  - Do we need to invert $\mathbf{S}(\lambda)$?
  - When there are $n = 150$ countries, $\mathbf{S}(\lambda)$ is an $150^2 \times 150^2$ matrix $(506, 250, 000$ elements!)
- Core idea
  - Utilize the structure of $\mathbf{A} = \lambda_d(I_n \otimes W) + \lambda_o(W \otimes I_n) + \lambda_w(W \otimes W)$ $(\mathbf{S}(\lambda) = I_N - \mathbf{A}(\lambda))$
  - Utilize $W = QDQ^{-1}$ (obtain this decomposition before estimation).
    - ★ Q is the eigenvector basis of W
    - ★ D is the diagonal matrix consisting of eigenvalues of W

Recall $\mathbf{S}^{-1} = (I_N - \mathbf{A})^{-1}$, where $\mathbf{A} = \lambda_d(I_n \otimes W) + \lambda_o(W \otimes I_n) + \lambda_w(W \otimes W)$.

- For the invertibility of **S**, we assume $\rho_{\mathsf{spec}}(\mathbf{A}) < 1$.

Recall $\mathbf{S}^{-1} = (I_N - \mathbf{A})^{-1}$, where $\mathbf{A} = \lambda_d(I_n \otimes W) + \lambda_o(W \otimes I_n) + \lambda_w(W \otimes W)$.

- For the invertibility of **S**, we assume $\rho_{\text{spec}}(\mathbf{A}) < 1$.
- We have $\rho_{\text{spec}}(\mathbf{A}) = \max\{b(1,1), b(1,\varphi_{\min}), b(\varphi_{\min},1), b(\varphi_{\min},\varphi_{\min})\}$, where
  - $b(1,1) = \lambda_d + \lambda_o + \lambda_w,\ b(1,\varphi_{\min}) = \lambda_d\varphi_{\min} + \lambda_o + \lambda_w\varphi_{\min}$,
    $b(\varphi_{\min},1) = \lambda_d + \lambda_o\varphi_{\min} + \lambda_w\varphi_{\min}$,
    $b(\varphi_{\min},\varphi_{\min}) = \lambda_d\varphi_{\min} + \lambda_o\varphi_{\min} + \lambda_w\varphi_{\min}^2$.

Recall $\mathbf{S}^{-1} = (I_N - \mathbf{A})^{-1}$, where $\mathbf{A} = \lambda_d(I_n \otimes W) + \lambda_o(W \otimes I_n) + \lambda_w(W \otimes W)$.

- For the invertibility of **S**, we assume $\rho_{\text{spec}}(\mathbf{A}) < 1$.
- We have $\rho_{\text{spec}}(\mathbf{A}) = \max\{b(1,1), b(1,\varphi_{\min}), b(\varphi_{\min},1), b(\varphi_{\min},\varphi_{\min})\}$, where
  - $b(1,1) = \lambda_d + \lambda_o + \lambda_w$, $b(1,\varphi_{\min}) = \lambda_d\varphi_{\min} + \lambda_o + \lambda_w\varphi_{\min}$, $b(\varphi_{\min},1) = \lambda_d + \lambda_o\varphi_{\min} + \lambda_w\varphi_{\min}$, $b(\varphi_{\min},\varphi_{\min}) = \lambda_d\varphi_{\min} + \lambda_o\varphi_{\min} + \lambda_w\varphi_{\min}^2$.
- Why? $I_n \otimes W$, $W \otimes I_n$, and $W \otimes W$ share the same eigenvector basis:

$$(I_n \otimes W)(q_i \otimes q_j) = q_i \otimes Wq_j = q_i \otimes \varphi_j q_j = \varphi_j(q_i \otimes q_j)$$
$$(W \otimes I_n)(q_i \otimes q_j) = Wq_i \otimes q_j = \varphi_i q_i \otimes q_j = \varphi_i(q_i \otimes q_j)$$
$$(W \otimes W)(q_i \otimes q_j) = Wq_i \otimes Wq_j = (\varphi_i q_i) \otimes (\varphi_j q_j) = (\varphi_j \varphi_j)(q_i \otimes q_j),$$

where Q is some eigenvector basis of W and q_i is the i[th] column vector of Q.

Recall $\mathbf{S}^{-1} = (I_N - \mathbf{A})^{-1}$, where $\mathbf{A} = \lambda_d(I_n \otimes W) + \lambda_o(W \otimes I_n) + \lambda_w(W \otimes W)$.

- For the invertibility of **S**, we assume $\rho_{\textbf{spec}}(\mathbf{A}) < 1$.
- We have $\rho_{\textbf{spec}}(\mathbf{A}) = \max\{b(1,1), b(1,\varphi_{\textbf{min}}), b(\varphi_{\textbf{min}},1), b(\varphi_{\textbf{min}},\varphi_{\textbf{min}})\}$, where
  - $b(1,1) = \lambda_d + \lambda_o + \lambda_w,\ b(1,\varphi_{\textbf{min}}) = \lambda_d\varphi_{\textbf{min}} + \lambda_o + \lambda_w\varphi_{\textbf{min}}$,
    $b(\varphi_{\textbf{min}},1) = \lambda_d + \lambda_o\varphi_{\textbf{min}} + \lambda_w\varphi_{\textbf{min}}$,
    $b(\varphi_{\textbf{min}},\varphi_{\textbf{min}}) = \lambda_d\varphi_{\textbf{min}} + \lambda_o\varphi_{\textbf{min}} + \lambda_w\varphi_{\textbf{min}}^2$.
- Why? $I_n \otimes W,\ W \otimes I_n,$ and $W \otimes W$ share the same eigenvector basis:

$$(I_n \otimes W)(q_i \otimes q_j) = q_i \otimes Wq_j = q_i \otimes \varphi_j q_j = \varphi_j(q_i \otimes q_j)$$
$$(W \otimes I_n)(q_i \otimes q_j) = Wq_i \otimes q_j = \varphi_i q_i \otimes q_j = \varphi_i(q_i \otimes q_j)$$
$$(W \otimes W)(q_i \otimes q_j) = Wq_i \otimes Wq_j = (\varphi_i q_i) \otimes (\varphi_j q_j) = (\varphi_j\varphi_j)(q_i \otimes q_j),$$

  where Q is some eigenvector basis of W and q_i is the i^{th} column vector of Q.

  - Thus,

$$\begin{aligned}\mathbf{A}(q_i \otimes q_j) &= (\lambda_d(I_n \otimes W) + \lambda_o(W \otimes I_n) + \lambda_w(W \otimes W))(q_i \otimes q_j) \\ &= (\lambda_d\varphi_j + \lambda_o\varphi_i + \lambda_w\varphi_i\varphi_j)(q_i \otimes q_j), \quad i,j = 1,\ldots,n.\end{aligned}$$

  - Note that the eigenvalue of **A** is a bilinear map:
    $b(\varphi_i,\varphi_j) = \lambda_d\varphi_j + \lambda_o\varphi_i + \lambda_w\varphi_i\varphi_j$ for $(\varphi_i,\varphi_j) \in [\varphi_{\textbf{min}}, 1]^2$.

- Recall $\mathrm{A}(\lambda)(q_i \otimes q_j) = (\lambda_d \varphi_j + \lambda_o \varphi_i + \lambda_w \varphi_i \varphi_j)(q_i \otimes q_j)$ for $i, j = 1, \cdots, n$.
- We want to obtain the matrix fixed-point $T(\boldsymbol{\theta})$ satisfying $\mathrm{vec}\,(T(\boldsymbol{\theta})) = \mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta})$ for each $\boldsymbol{\theta}$.
  - $Z^{\mathrm{mat}}(\boldsymbol{\theta})$ is such that $\mathbf{Z}(\boldsymbol{\theta}) = \mathrm{vec}\,(Z^{\mathrm{mat}}(\boldsymbol{\theta}))$.

2. Specification

Microfoundation - Faster computation algorithm for $\mathbf{S}^{-1}\mathbf{Z}(\theta)$

- Recall $\mathrm{A}(\lambda)(q_i \otimes q_j) = (\lambda_d \varphi_j + \lambda_o \varphi_i + \lambda_w \varphi_i \varphi_j)(q_i \otimes q_j)$ for $i, j = 1, \cdots, n$.
- We want to obtain the matrix fixed-point $T(\boldsymbol{\theta})$ satisfying $\mathrm{vec}(T(\boldsymbol{\theta})) = \mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta})$ for each $\boldsymbol{\theta}$.
  - $Z^{\mathrm{mat}}(\boldsymbol{\theta})$ is such that $\mathbf{Z}(\boldsymbol{\theta}) = \mathrm{vec}(Z^{\mathrm{mat}}(\boldsymbol{\theta}))$.
- Note that Q and D are invariant in the estimation procedure. Hence, we have

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

  - $\widetilde{Z}^{\mathrm{mat}}(\boldsymbol{\theta}) = Q^{-1} Z^{\mathrm{mat}}(\boldsymbol{\theta}) Q^{-1\prime}$
  - $\widetilde{T}(\boldsymbol{\theta}) = Q^{-1} T(\boldsymbol{\theta}) Q^{-1\prime}$

2. Specification

Microfoundation - Faster computation algorithm for $\mathbf{S}^{-1}\mathbf{Z}(\boldsymbol{\theta})$

- Recall $\mathrm{A}(\lambda)(q_i \otimes q_j) = (\lambda_d \varphi_j + \lambda_o \varphi_i + \lambda_w \varphi_i \varphi_j)(q_i \otimes q_j)$ for $i, j = 1, \cdots, n$.
- We want to obtain the matrix fixed-point $T(\boldsymbol{\theta})$ satisfying $\mathrm{vec}\,(T(\boldsymbol{\theta})) = \mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta})$ for each $\boldsymbol{\theta}$.
  - $Z^{\mathrm{mat}}(\boldsymbol{\theta})$ is such that $\mathbf{Z}(\boldsymbol{\theta}) = \mathrm{vec}\,(Z^{\mathrm{mat}}(\boldsymbol{\theta}))$.
- Note that Q and D are invariant in the estimation procedure. Hence, we have

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

  - $\widetilde{Z}^{\mathrm{mat}}(\boldsymbol{\theta}) = Q^{-1} Z^{\mathrm{mat}}(\boldsymbol{\theta}) Q^{-1\prime}$
  - $\widetilde{T}(\boldsymbol{\theta}) = Q^{-1} T(\boldsymbol{\theta}) Q^{-1\prime}$
- This implies

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

  by $\mathrm{A}(\lambda)(q_i \otimes q_j) = (\lambda_d \varphi_j + \lambda_o \varphi_i + \lambda_w \varphi_i \varphi_j)(q_i \otimes q_j)$ for $i, j = 1, \cdots, n$.
  - Then, we can easily recover $T(\boldsymbol{\theta}) = Q \widetilde{T}(\boldsymbol{\theta}) Q'$.

3. Estimation

PPMLE - Efficient computation algorithm for $\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta})$

n	$N = n^2$	Default (A)	Ours (B)	A/B
9	81	0.5668	0.3518	1.6112
25	625	12.3862	0.3967	31.2265
49	2401	1252.0077	3.9725	315.1687
64	4096	9465.7145	9.2816	1019.8380

- Default (A) maximizes the log-likelihood based on the direct computation of $\mathbf{S}^{-1}(\lambda)$.
- Ours (B) maximizes the log-likelihood based on our suggested method.
- When $n \simeq 150$, our proposed method is **more than four orders of magnitude faster**, rendering maximum likelihood estimation computationally feasible even for moderately large n.

3. Estimation

PPMLE - Identification

Assumption (Identification)

Let Θ_θ denote a compact parameter space of θ and $\mathbf{\Phi}$ represent a parameter space of ϕ.
(i) Assume $\liminf_{n\to\infty} \inf_{\phi\in\mathbf{\Phi}} \varphi_{\min}\left(\frac{1}{n}\mathbf{D}'\mathbf{S}^{-1\prime}(\lambda)\text{Diag}\left(\boldsymbol{\mu}(\boldsymbol{\theta})\right)\mathbf{S}^{-1}(\lambda)\mathbf{D} + \text{other terms}\right) > 0$ for each $\theta \in \Theta_\theta$. Then, $\hat{\phi}(\theta) = \text{argmax}_{\phi\in\mathbf{\Phi}} \ell_N(\theta, \phi)$ is unique for each $\theta \in \Theta_\theta$ and for a large n.
(ii) For each $\theta \in \Theta_\theta$, let

$$\widehat{\mathbf{H}}(\theta) = \frac{1}{N}\widehat{\mathbf{G}}'(\theta)\mathbf{S}^{-1\prime}(\lambda)\text{Diag}\left(\widehat{\boldsymbol{\mu}}(\theta)\right)\mathbf{S}^{-1}(\lambda)\widehat{\mathbf{G}}(\theta) + \text{other terms},$$

where $\widehat{\mathbf{G}}(\theta) = \mathbf{G}\left(\theta, \hat{\phi}(\theta)\right)$ and $\widehat{\boldsymbol{\mu}}(\theta) = \boldsymbol{\mu}(\theta, \hat{\phi}(\theta))$ for each $\theta \in \Theta_\theta$.
Assume $\liminf_{n\to\infty} \inf_{\theta\in\Theta_\theta} \varphi_{\min}\left(\widehat{\mathbf{H}}(\theta)\right) > 0$.
$\Rightarrow$ Uniqueness of $\theta^0 = \text{argmax}_{\theta\in\Theta_\theta} \ell_\infty(\theta, \phi(\theta))$, and consequently, $\phi^0 = \phi(\theta^0)$.

- Notations
  - $\mathbf{D} = [\mathrm{I}_n \otimes l_n, l_n \otimes \mathrm{I}_n]$ is an $N \times 2n$ matrix for dummy variables
  - $\mathbf{G}(\boldsymbol{\theta}) = \left[(I_n \otimes W)\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta}), (W \otimes I_n)\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta}), (W \otimes W)\mathbf{S}^{-1}(\lambda)\mathbf{Z}(\boldsymbol{\theta}), \mathbf{X}\right]$ and $\mathbf{Z}(\boldsymbol{\theta}) = \mathbf{X}\beta + \boldsymbol{\alpha} \otimes l_n + l_n \otimes \boldsymbol{\eta}$
  - $\boldsymbol{\mu}(\boldsymbol{\theta}) = \left(\exp\left(\tilde{\mu}_{11}(\boldsymbol{\theta})\right), \cdots, \exp\left(\tilde{\mu}_{n1}(\boldsymbol{\theta})\right), \cdots, \exp\left(\tilde{\mu}_{1n}(\boldsymbol{\theta})\right), \cdots, \exp\left(\tilde{\mu}_{nn}(\boldsymbol{\theta})\right)\right)$ with $\tilde{\mu}_{ij}(\boldsymbol{\theta}) = \sum_{k,l=1}^{n} s_{ij,kl}(\lambda)\left(x_{kl}'\beta + \alpha_l + \eta_k\right)$

Theorem

Under some regularity conditions,

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

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

- Here,

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

  and

$$\boldsymbol{\Omega}_{\theta,N}=\frac{1}{N}\mathbf{G}'\mathbf{S}^{-1\prime}\mathbf{M}_{\mathbf{D}}'\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)\mathbf{M}_{\mathbf{D}}\mathbf{S}^{-1}\mathbf{G}.$$

  - $\mathbf{M}_{\mathbf{D}}=I_N-\mathbf{P}_{\mathbf{D}}\mathrm{Diag}(\boldsymbol{\mu})$ and $\mathbf{P}_{\mathbf{D}}=\mathbf{S}^{-1}\mathbf{D}(\widetilde{\mathbf{D}'\mathbf{D}})^{-1}\mathbf{D}'\mathbf{S}^{-1\prime}$ with $\widetilde{\mathbf{D}'\mathbf{D}}=\mathbf{D}'\mathbf{S}^{-1\prime}\mathrm{Diag}(\boldsymbol{\mu})\mathbf{S}^{-1}\mathbf{D}-\mathbf{H}^{\phi\phi}$
- No asymptotic bias

Theorem (Fixed-effect estimators)

Under some regularity conditions,

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

$$\sqrt{n}\left(\widehat{\eta}_i - \eta_i^0\right) \xrightarrow{d} N(0, \lim_{n\to\infty} e'_{2n,n+i}\boldsymbol{V}_{\phi,N}e_{2n,n+i})$$

as $n \to \infty$.

- Using $\{\widehat{\alpha}_j\}$ and $\{\widehat{\eta}_i\}$, we can identify/estimate γ_o^0 and γ_d^0.
- Note that

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

  where

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

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

- $\mathbf{\Omega}_\theta$ contains $\mathbb{E}\left(\mathbf{u}\mathbf{u}'\right)$, where $\mathbf{u} = \left(u_{11}, \cdots, u_{n1}, \cdots, u_{1n}, \cdots, u_{nn}\right)'$.

- $\mathbf{\Omega}_\theta$ contains $\mathbb{E}(\mathbf{u}\mathbf{u}')$, where $\mathbf{u} = (u_{11}, \cdots, u_{n1}, \cdots, u_{1n}, \cdots, u_{nn})'$.

Theorem (Variance estimators)

Let d_N denote the pre-specified threshold distance and $d^*_{ij,kl}$ the distance between ij and kl. Let the spatial HAC estimator and the infeasible spatial HAC estiamator, respectively,

$$\widehat{\mathbf{\Omega}}_{\theta,N} = \frac{1}{N}\sum_{i,j,k,l=1}^{n} b_{ij}c'_{kl}\hat{u}_{ij}\hat{u}_{kl} \cdot \mathrm{K}\left(\frac{d^*_{ij,kl}}{d_N}\right) \text{ for some } b_{ij} \text{ and } c_{kl},$$

$$\widetilde{\mathbf{\Omega}}_{\theta,N} = \frac{1}{N}\sum_{i,j,k,l=1}^{n} b_{ij}c'_{kl}u_{ij}u_{kl} \cdot \mathrm{K}\left(\frac{d^*_{ij,kl}}{d_N}\right) \text{ for some } b_{ij} \text{ and } c_{kl}.$$

Under some regularity conditions,

(i) (Variance) $\lim_{n\to\infty} \frac{N}{\mathbb{E}(deg^*)}\mathrm{Var}\left(vec\left(\widetilde{\mathbf{\Omega}}_{\theta,N}\right)\right) = \overline{\mathrm{K}}(1+C)(\mathbf{\Omega}_\theta \otimes \mathbf{\Omega}_\theta)$;

(ii) (Bias) $\lim_{n\to\infty} d_N^q\left(\mathbb{E}\left(\widetilde{\mathbf{\Omega}}_{\theta,N}\right) - \mathbf{\Omega}_{\theta,N}\right) = -K_q\mathbf{\Omega}_\theta^{(q)}$;

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

$\sqrt{\frac{N}{\mathbb{E}(deg^*)}}\left(\widehat{\mathbf{\Omega}}_{\theta,N} - \widetilde{\mathbf{\Omega}}_{\theta,N}\right) = O_p(1)$.

Monte Carlo Simulations

4. Monte Carlo Simulations

- We generate the data based on the structural gravity model.
  - Basic setting
  - $n = 49$, $\lambda_d = 0.2$, $\lambda_o = 0.2$, $\lambda_w = 0.1$, $\beta_1 = 0.6$, and $\beta_2 = 0.2$
  - W is row-normalized, constructed by supposing two influential units.

4. Monte Carlo Simulations

- We generate the data based on the structural gravity model.
  - Basic setting
  - $n = 49$, $\lambda_d = 0.2$, $\lambda_o = 0.2$, $\lambda_w = 0.1$, $\beta_1 = 0.6$, and $\beta_2 = 0.2$
  - W is row-normalized, constructed by supposing two influential units.

- Results

	λ_d	λ_o	λ_w	β_1	β_2	α_{49}	η_{49}
Empirical bias	0.0123	-0.0004	-0.0163	-0.0011	-0.0001	0.0131	0.0044
Empirical STD	0.0266	0.0260	0.0355	0.0137	0.0130	0.0845	0.0850
`s.e.`	0.0229	0.0232	0.0331	0.0120	0.0118	0.0297	0.0228
CP	0.9330	0.9340	0.9280	0.8970	0.9180	0.8630	0.8910

  - Among the four possible kernel choices (Bartlett, Parzen, Tukey–Hanning, QS), Parzen performs best.
  - Among the three distance measures, $d^{*,2}_{ij,kl}$ performs best.
    - ★ $d^{*,1}_{ij,kl} = D_{ik} + D_{jl}$
    - ★ $d^{*,2}_{ij,kl} = \left(D^2_{ik} + D^2_{jl}\right)^{\frac{1}{2}}$
    - ★ $d^{*,\infty}_{ij,kl} = \max\left(D_{ik}, D_{jl}\right)$

Empirical Application

5. Empirical Application

Setup and Data

- We suspect different network structures across the following phases:
  - Phase 1: 1986, trade liberalization
  - Phase 2: 1997, active NAFTA implementation
  - Phase 3: 2007, emergence of the China trade shock
  - Phase 4: 2016, expansion of global supply chains.

5. Empirical Application

Setup and Data

- We suspect different network structures across the following phases:
  - Phase 1: 1986, trade liberalization
  - Phase 2: 1997, active NAFTA implementation
  - Phase 3: 2007, emergence of the China trade shock
  - Phase 4: 2016, expansion of global supply chains.

- Data
  - World trade flows: the Center for International Data at UC Davis
  - Distance, border, legal, language, colony, currency, islands, landlock from Helpman et al. (2008), Chen et al. (2021)
  - FTA from WTO data
  - Due to the data availability, the 136, 142, 146, and 147 countries by phase, respectively, are included.

5. Empirical Application

Setup and Data

- We suspect different network structures across the following phases:
  - Phase 1: 1986, trade liberalization
  - Phase 2: 1997, active NAFTA implementation
  - Phase 3: 2007, emergence of the China trade shock
  - Phase 4: 2016, expansion of global supply chains.
- Data
  - World trade flows: the Center for International Data at UC Davis
  - Distance, border, legal, language, colony, currency, islands, landlock from Helpman et al. (2008), Chen et al. (2021)
  - FTA from WTO data
  - Due to the data availability, the 136, 142, 146, and 147 countries by phase, respectively, are included.
- Countries' connectivity matrix: Average total trade flows over the recent past years
  - $w_{ij}^{\text{Phase}} = \frac{\widetilde{w}_{ij}^{\textbf{Phase}}}{\sum_{k=1}^{n} \widetilde{w}_{ik}^{\textbf{Phase}}}$, where $\widetilde{w}_{ij}^{\text{Phase}} = \frac{1}{\#(\mathcal{T}^{\textbf{Phase}})} \sum_{t \in \mathcal{T}^{\textbf{Phase}}} \left(y_{ij,t} + y_{ji,t}\right)$
  - e.g., $\mathcal{T}^{\text{Phase}=2} = \{1987, 1988, \cdots, 1996\}$
  - W is undirected and row-normalized, with its diagonal elements being zero.

5. Empirical Application

Network Statistics

	linear-in-means	Bipartite	Phase 1	Phase 2	Phase 3	Phase 4
degree (deg)	149.0000	75.0000	81.3088	118.9859	136.3973	138.4354
Std(deg)	0.0000	0.0000	34.9280	21.9047	12.5476	12.1543
Herfindahl–Hirschman Index (HHI)	0.0067	0.0133	0.1673	0.1419	0.1249	0.1182
Std(HHI)	0.0000	0.0000	0.1206	0.1040	0.0932	0.0881
n^{HHI}	149.0000	75.0000	8.3230	9.7423	11.0296	11.5994
$\varphi_{(2)}$	-0.0067	$\simeq 0$	0.5413	0.5653	0.5752	0.5553
$\varphi_{\min}$	-0.0067	-1	-0.5296	-0.5151	-0.5088	-0.4419
Density	1	0.5034	0.4948	0.7560	0.8910	0.9105

- Summary network statistics
  - Networks become more *connected and diversified* as the phases progress.
  - Networks deviate from a bipartite structure but continue to display heterogeneity that differs from the linear-in-means structure.
  - A persistent core-periphery structure across all phases is observed.

5. Empirical Application
Estimation Results

Phase	1	2	3	4
λ_d (same exporter)	-0.1261***	0.3002***	-0.6440***	0.1370*
	(0.0480)	(0.0362)	(0.1106)	(0.0712)
λ_o (same importer)	-0.1900**	0.3510***	-0.6246***	0.2830***
	(0.0968)	(0.0392)	(0.0913)	(0.0649)
λ_w (third-party)	-0.1533***	0.3354***	1.3110***	0.5734***
	(0.0561)	(0.0336)	(0.0742)	(0.0418)
Distance	-1.5011**	-1.5184***	-1.5808***	-1.7511***
	(0.7203)	(0.2449)	(0.3730)	(0.2849)
Border	1.1213***	0.9973***	-0.1417	0.9402***
	(0.3638)	(0.2134)	(0.2256)	(0.1575)
FTA	0.8220***	0.4123***	0.2586**	0.4968***
	(0.2338)	(0.1717)	(0.1303)	(0.0851)
McFadden's R^2	0.1295	0.2848	0.1037	0.1739

Note: Significance levels: *** (1%), ** (5%), * (10%). Standard errors are in parentheses. McFadden's $R^2 = 1 - \frac{\widehat{\ell}_N(\text{=our model})}{\widehat{\ell}_N^{trad.}(\text{traditional gravity model})}$

- λ_w changes from negative to positive and remains highly significant.
  - persistent and significant *third-party* network effects.

5. Empirical Application

Estimation Results

- λ_w changes from negative to positive and remains highly significant.
  - persistent and significant *third-party* network effects.

- **Phase 1 (1986, early trade liberalization):** All interaction parameters are negative and statistically significant,
  - indicating trade flows sharing the same exporter or the same importer behaved as *substitutes*.
  - In the early stage of market liberalization
    - exporters: exporters faced limited capacity or market-access constraints
    - importers: reallocated their demand across competing sources of supply.
  - At this stage, network competition dominated network complementarity.

- **Phase 2 (1997, active NAFTA implementation):** All interaction parameters turn positive,
  - reflecting a structural shift toward *complementarity*.
  - Regional integration under NAFTA
    - reduced trade frictions
    - deepened production linkages across member countries
  - capture the emergence of regional value chains and increasing returns to network connectivity.

- **Phase 3 (2007, the emergence of the China trade shock):** λ_d and λ_o again become negative.
    - The sharp rise of China as a global exporter may have introduced strong competitive pressures in world markets.
    - This leads to substitution effects across both exporters and importers.
        - exporters increasingly competed for global market share
        - importers rebalanced sourcing patterns in response to China's dominance.
- **Phase 4 (2016, expansion of global value chains):** λ_d and λ_o turn to positive,
    - as the value of participating in the network increases when more countries become interconnected under global value chains,
    - reinforcing complementarities across the system.

- Counterfactual scenario: Focusing on Phase 4, we consider a threefold increase in $\pi^{+}_{\mathsf{US,CN}}$ and $\pi^{+}_{\mathsf{CN,US}}$, thereby illustrating the recent US-China trade war.

- Counterfactual scenario: Focusing on Phase 4, we consider a threefold increase in $\pi^{+}_{\mathrm{US,CN}}$ and $\pi^{+}_{\mathrm{CN,US}}$, thereby illustrating the recent US-China trade war.
- We compute $\widehat{s}_{ij}$ and $\widetilde{s}_{ij}$: for each $i = 1, \cdots, n$,

$$\widehat{s}_{ij} = \frac{\widehat{\mu}_{ij}}{\sum_{k=1,k\neq i} \widehat{\mu}_{ik}} \text{ (estimated) and } \widetilde{s}_{ij} = \frac{\widetilde{\mu}_{ij}}{\sum_{k=1,k\neq i}^{n} \widetilde{\mu}_{ik}} \text{ (counterfactual) for } j \neq i.$$

  - For comparison, we also compute $\widehat{s}_{ij}^{\mathrm{con}}$ and $\widetilde{s}_{ij}^{\mathrm{con}}$ from the conventional model.

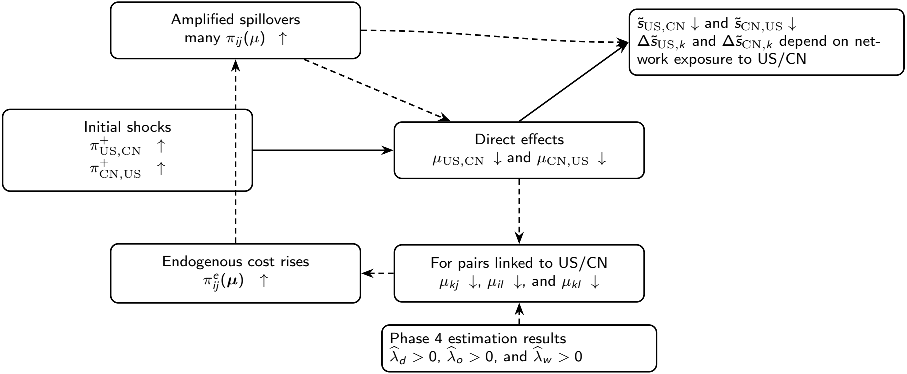
- Shock: $\pi^{+}_{\mathrm{US,CN}} \uparrow$, $\pi^{+}_{\mathrm{CN,US}} \uparrow$

- Shock: $\pi^{+}_{\mathrm{US,CN}} \uparrow$, $\pi^{+}_{\mathrm{CN,US}} \uparrow$
- Conventional gravity model ($\lambda = 0$)
  - Direct: $\mu_{\mathrm{US,CN}},\ \mu_{\mathrm{CN,US}} \downarrow$
  - Indirect
    - MR only: $\{P_i, \Pi_j\}$ adjust $\Rightarrow$ *limited* third-country effects

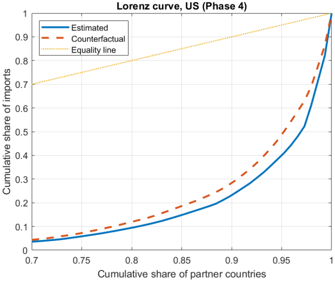
- Shock: $\pi^{+}_{\mathrm{US,CN}} \uparrow$, $\pi^{+}_{\mathrm{CN,US}} \uparrow$
- Conventional gravity model ($\lambda = 0$)
  - Direct: $\mu_{\mathrm{US,CN}},\ \mu_{\mathrm{CN,US}} \downarrow$
  - Indirect
    - ⋆ MR only: $\{P_i, \Pi_j\}$ adjust $\Rightarrow$ *limited* third-country effects
- Our model (Phase 4: $\hat{\lambda}_d, \hat{\lambda}_o, \hat{\lambda}_w > 0$)
  - Direct: $\mu_{\mathrm{US,CN}},\ \mu_{\mathrm{CN,US}} \downarrow$
  - Indirect
    - ⋆ Endogenous network cost rises for pairs linked to US/CN: $\widetilde{\pi}^{e}_{ij}(\boldsymbol{\mu}; k, l) \uparrow$
    - ⋆ Amplified propagation: $\pi(\boldsymbol{\mu}) \Rightarrow \{P_i(\boldsymbol{\mu}), \Pi_j(\boldsymbol{\mu})\}$

    $\Rightarrow$ Large, system-wide reallocation of import shares

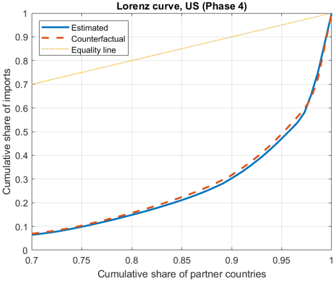
- Model's mechanism
  - Solid arrows: channels that are present in the conventional model
  - Dashed arrows: additional propagation mechanisms implied by $\widehat{\lambda} > 0$



(a) U.S. (Our model)

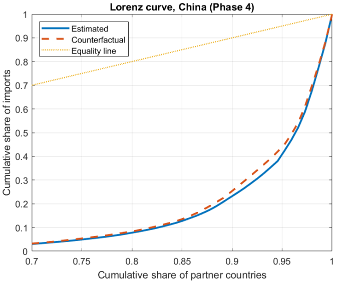


(b) U.S. (Conventional)

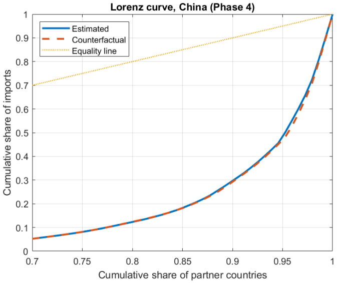


- U.S.-China shock $\Rightarrow$ U.S. import shares strongly reallocated toward many alternative suppliers.
  - Our model: Loss absorbed by many partners, Import concentration $\downarrow$
  - Conventional gravity: Adjustment concentrated on few large suppliers.

5. Empirical Application
Counterfactual analysis



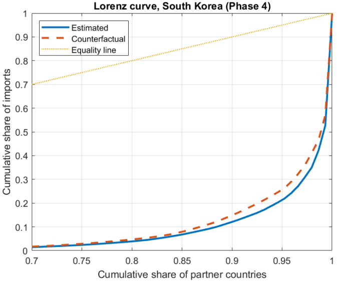
(a) China (Our model)



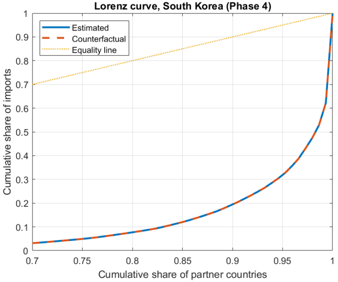
(b) China (Conventional)

- U.S.-China shock
  - Our model: China experiences diversification under spillovers.
  - Conventional gravity: Minimal third-country adjustment

(a) South Korea (Our model)



(b) South Korea (Conventional)



- U.S.-China shock ⇒ Non-targeted countries respond through network exposure, not bilateral costs.

5. Empirical Application

Counterfactual analysis: Overall results and policy implications

- Takeaway
  - Trade policy shocks reshape the *global trade structure* through networks.
    - Network-based endogenous trade costs generate large third-country adjustments and distributional responses.
    - These are largely absent in the conventional gravity models.

5. Empirical Application

Counterfactual analysis: Overall results and policy implications

- Takeaway
  - Trade policy shocks reshape the *global trade structure* through networks.
    - ★ Network-based endogenous trade costs generate large third-country adjustments and distributional responses.
    - ★ These are largely absent in the conventional gravity models.
  - Targeted hub economies (U.S., China):
    - ★ Import shares reallocated toward many partners $\Rightarrow$ Diversification (lower concentration)

- Takeaway
  - Trade policy shocks reshape the *global trade structure* through networks.
    - ★ Network-based endogenous trade costs generate large third-country adjustments and distributional responses.
    - ★ These are largely absent in the conventional gravity models.
  - Targeted hub economies (U.S., China):
    - ★ Import shares reallocated toward many partners $\Rightarrow$ Diversification (lower concentration)
  - Third countries (e.g. Korea):
    - ★ Large reallocation despite no direct bilateral shock $\Rightarrow$ Network exposure matters
    - ★ Spillovers occur even when π_{ij}^{+} is unchanged

5. Empirical Application

Counterfactual analysis: Overall results and policy implications

- Takeaway
  - Trade policy shocks reshape the *global trade structure* through networks.
    - ★ Network-based endogenous trade costs generate large third-country adjustments and distributional responses.
    - ★ These are largely absent in the conventional gravity models.
  - Targeted hub economies (U.S., China):
    - ★ Import shares reallocated toward many partners $\Rightarrow$ Diversification (lower concentration)
  - Third countries (e.g. Korea):
    - ★ Large reallocation despite no direct bilateral shock $\Rightarrow$ Network exposure matters
    - ★ Spillovers occur even when π_{ij}^{+} is unchanged
- **Policy implications**: Ignoring network spillovers leads to a systematic understatement of trade policy effects.

5. Empirical Application

Counterfactual analysis: Overall results and policy implications

- Takeaway
  - Trade policy shocks reshape the *global trade structure* through networks.
    - ★ Network-based endogenous trade costs generate large third-country adjustments and distributional responses.
    - ★ These are largely absent in the conventional gravity models.
  - Targeted hub economies (U.S., China):
    - ★ Import shares reallocated toward many partners $\Rightarrow$ Diversification (lower concentration)
  - Third countries (e.g. Korea):
    - ★ Large reallocation despite no direct bilateral shock $\Rightarrow$ Network exposure matters
    - ★ Spillovers occur even when π_{ij}^{+} is unchanged
- **Policy implications**: Ignoring network spillovers leads to a systematic understatement of trade policy effects.
  - Network-aware policy evaluation
    - ★ Bilateral tariffs generate global distributional effects
  - Hub countries matter
    - ★ Policies targeting hubs propagate disproportionately

Conclusion

- Research question: How do countries operate trade costs leveraging their trade networks?
- To this end, we characterized a network-interaction-based gravity equation.
- For estimation, we developed a Poisson pseudo maximum likelihood estimator with network effects and relevant tools for statistical inference.
- In empirical application, we expect to find evidence of endogenous trade costs, followed by implications.
- Counterfactual experiments illustrate how spillovers reshape global trade structures.

Appendix

- A common way to organize movements is through **Origin-Destination (OD) flows,** which represent *non-negative* intensities between origins and destinations.

Origin-destination flow matrix

Destination /Origin	Origin 1	Origin 2	...	Origin n
Destination 1	$o_1 \to d_1$	$o_2 \to d_1$	...	$o_n \to d_1$
Destination 2	$o_1 \to d_2$	$o_2 \to d_2$	...	$o_n \to d_2$
$\vdots$		$\vdots$		
Destination n	$o_1 \to d_n$	$o_2 \to d_n$	...	$o_n \to d_n$

Source: LeSage and Pace (2009)

- OD flows capture *directional* interactions between two distinct *locations* (i.e., an origin and a destination).
    - A structured framework for analyzing heterogeneous policy effects (e.g., directional heterogeneities in economic activities).

- OD flows are *spatial* in nature.
    - Why? Because OD flows inherently involve two locations.
    - Reflect regional dynamics shaped by geographic coordinates (e.g., latitude and longitude) or socioeconomic factors (e.g., income levels, population density).

- Thus, *distance* naturally emerges as a key determinant.
    - As distance increases, flows typically weaken and may even cease when costs become prohibitive.
    - Highlights the importance of incorporating distance into OD flows models.

Background

Constant elasticity models for origin-destination (OD) flows

- OD flows are often analyzed using the **constant elasticity models.**
  - E.g. $Y = AX^{\alpha}Z^{\beta}$
  - Examines how proportional changes in explanatory variables influence the intensity of flows.
  - E.g., Cobb-Douglas function, migration, activities in networks, household production functions, consumer preference, firm production functions, etc.
- **The gravity equation** stands out for its profound and enduring influence on trade literature. (Isard 1954; Tinbergen 1962)
  - A robust framework for predicting trade flows between countries.
  - Accounting for factors such as economic size and geographic distance. (Anderson 1979; Helpman and Krugman 1985; Anderson and van Wincoop 2003; Chen et al. 2021)
  - Traditionally employed in a bilateral context.

Issues with the traditional gravity equation

However, traditional gravity models overlook the role of dominant units, encountering several limitations.

1. Fail to capture the interdependencies among individual units under their influence. (i.e., assume all pairwise interactions are independent of dominant units.)
   - **Multilateral interactions**
   - **Spatial correlations in errors**
2. Lack individual units' inherent heterogeneities
   - Heterogeneity in trade costs shaped by dominant units, represented as **multilateral resistance**
   - Extent of the influence of dominant units varies across countries depending on their unique characteristics, leading to **heteroskedasticity in errors**
3. Use the **log-transformed specification**, leading to invalid inference.

Issues with the traditional gravity equation

However, traditional gravity models overlook the role of dominant units, encountering several limitations.

1. Fail to capture the interdependencies among individual units under their influence. (i.e., assume all pairwise interactions are independent of dominant units.)
   - **Multilateral interactions** (Addressed by LeSage and Pace 2008)
   - **Spatial correlations in errors**
2. Lack individual units' inherent heterogeneities
   - Heterogeneity in trade costs shaped by dominant units, represented as **multilateral resistance**
   - Extent of the influence of dominant units varies across countries depending on their unique characteristics, leading to **heteroskedasticity in errors**
3. Use the **log-transformed specification**, leading to invalid inference.

Background

Traditional Gravity Equation (Isard 1954; Tinbergen 1962)

- The *gravity equation* in trade literature:

$$\mathrm{E}(Y_{ij}|\mathrm{X}) = G\frac{X_i^{\beta_1} X_j^{\beta_2}}{D_{ij}^{\beta_3}},$$

  - Y_{ij} represents volume of trade from country j to country i;
  - X_i and X_j typically represent the GDPs for countries i and j; and
  - D_{ij} denotes the distance between the two countries.

Background

Traditional Gravity Equation (Isard 1954; Tinbergen 1962)

- The *gravity equation* in trade literature:

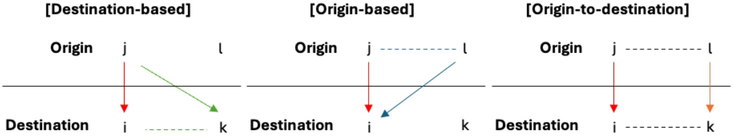
$$E(Y_{ij}|X) = G\frac{X_i^{\beta_1} X_j^{\beta_2}}{D_{ij}^{\beta_3}},$$

  - Y_{ij} represents volume of trade from country j to country i;
  - X_i and X_j typically represent the GDPs for countries i and j; and
  - D_{ij} denotes the distance between the two countries.

- Often use Poisson pseudo-maximum likelihood estimator (PPMLE) (Gourieroux et al. 1984, *Ecta*)
  - Became a standard (Silva and Tenreyro 2006, *REStat*)
  - Works for nonnegative outcomes (Note: Trade data often contain many zeros, with right-skewed and outliers)
  - PPMLE works well with various error structures and no asymptotic bias from the incidental parameters

Extension: Spatial Gravity Equation (LeSage and Pace 2008)

- Traditional gravity equation only addresses *bilateral* interactions.
  (Isard 1954; Tinbergen 1962; Anderson 1979; Helpman and Krugman 1985; Helpman 1987; Feenstra 2002; Anderson and van Wincoop 2003; Chen et al. 2021)
  - i.e., All pairwise interactions are independent, even of dominant units.
- Why should we care about *multilateral* interactions?
  - Dominant units exert significant influence on other units, creating interdependencies, i.e., multilateral. (Acemoglu et al. 2012, 2016)
  - Literature also finds third-party effects through networks.
    (Porojon 2001; Lee and Pace 2005; Tiefelsdorf 2003; LeSage and Pace 2008; Behrens et al. 2012)
- *Spatial* gravity equation (LeSage and Pace 2008) addresses multilateral interactions among cross-sectional units.
  - Idea: Modeling *three* types of dependencies



Background

Extension: Spatial Gravity Equation (LeSage and Pace 2008)

- Augmenting spatial terms,

$$\ln(y_{ij}+1) = \underbrace{\lambda_d \sum_{k=1}^{n} w_{ik} \ln(y_{kj}+1)}_{\text{Destination-based dependence}} + \underbrace{\lambda_o \sum_{l=1}^{n} m_{lj} \ln(y_{il}+1)}_{\text{Origin-based dependence}}$$

$$+ \underbrace{\lambda_w \sum_{k,l=1}^{n} w_{ik} m_{lj} \ln(y_{kl}+1)}_{\text{origin-to-destination dependence}} + x_{ij}'\beta + error_{ij}$$

- This reduces to

Derivation

$$\ln(y_{ij}+1) = \sum_{k,l=1}^{n} s_{ij\leftarrow kl}(x_{kl}'\beta + error_{kl}),$$

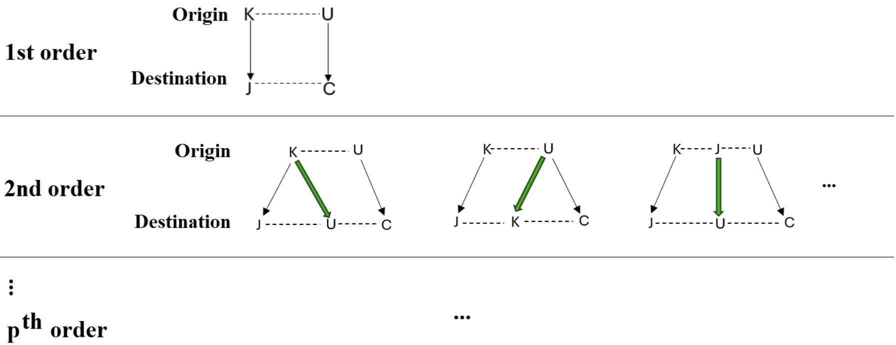
where $s_{ij\leftarrow kl}$ represents the signal from flow (k,l) to flow (i,j) as an element in $\mathrm{S}_N^{-1} := (\mathrm{I}_N - \lambda_d \mathrm{W}_N - \lambda_o \mathrm{M}_N - \lambda_w \mathrm{R}_N)^{-1}$.

Background

How Signals Capture Rich Structure of Dependencies

Remark. This specification encompasses all intermediate flows that carry the signal $s_{ij,kl}$ between the flows (k, l) and (i, j) by p-th order, $p \in \mathbb{N}$.

Eg. The signal $s_{KJ,UC}$ between (K, J) and (U, C) is carried by inifinitely many intermediate flows as....



Issue 1: Absence of Multilateral *Resistance* (MR)

Motivation: Unresolved issues in the traditional/spatial gravity models

- MR represents the general equilibrium effect of trade frictions within the entire global trade system.
  - Eg. If a country has high MR, it faces significant trade frictions relative to other countries.
  - MR captures *heterogeneity* in trade frictions by accounting for a country's relative access to global markets.
    Why? Countries with different geographic, institutional, and economic characteristics face varying degrees of trade resistance.
- Consider the gravity equation with MR (Anderson and van Wincoop 2003)

$$y_{ij} = \frac{x_i x_j}{x_w}\left(\frac{c_{ij}}{R_i R_j}\right)^{1-\sigma},$$

  - y_{ij} is the trade volume from country j to country i,
  - x_i, x_j are the GDP of country i and j and x_w is the world GDP,
  - $\sigma(> 1)$ is the elasticity of substitution between goods,
  - c_{ij} represents the trade costs between countries i and j,
  - R_i, R_j denote the multilateral resistance terms for countries i and j.

- The MRs are represented as

$$\begin{cases} R_i^{1-\sigma} & = \sum_j \left(\frac{c_{ij}}{R_j}\right)^{1-\sigma} \frac{x_j}{x_w} \\ R_j^{1-\sigma} & = \sum_i \left(\frac{c_{ij}}{R_i}\right)^{1-\sigma} \frac{x_i}{x_w} \end{cases}$$

- Thus, MR can be absorbed into *individual fixed effects:* Derivation

$$\ln y_{ij} = \sum_{k,l=1}^{n} s_{ij\leftarrow kl}(x_{kl}'\beta + \alpha_l + \eta_k + error_{kl}),$$

where $s_{ij\leftarrow kl}$ represents the signal from pair (k,l) to pair (i,j) as an element in $\mathrm{S}_N^{-1} := (\mathrm{I}_N - \lambda_d \mathrm{W}_N - \lambda_o \mathrm{M}_N - \lambda_w \mathrm{R}_N)^{-1}$.

Issue 2: Absence of Heteroskedasticity and Spatial Correlations in Errors

- Consider dominant economic positions that exerted substantial influence on other countries.
  - E.g., In 1986, the United States accounted for approximately 31.5% of global GDP, ranking first worldwide, and 2.44% of global trade as an origin, ranking third.
- Note that this is a common feature in the real world.
  - As a result, the error terms become correlated.
  - Moreover, the extent of this influence varies across countries depending on their unique characteristics (i.e., heteroskedasticity in errors).
- The traditional/spatial gravity equations do *not* account for these.
  - Failing to address these features affects the *efficiency* of estimators. (i.e., incorrect estimates of the variance-covariance matrix of estimators.)
  - Unreliable inferences on confidence intervals and hypothesis tests.

Issue 3: Invalid Inference with the Log-transformation

(i) The log-transformed error may fail to preserve the moment conditions of the error in the (original) constant elasticity model. (Silva and Tenreyro 2006)

To see this, consider a constant elasticity model:

$$y_{ij} = \exp(\beta_0^0 + \beta_1^0 x_{ij}) \cdot \xi_{ij} \Leftrightarrow y_{ij} = \exp(\beta_0^0 + \beta_1^0 x_{ij}) + u_{ij}, \tag{1}$$

where

$$\begin{cases} \xi_{ij} \text{ is the error with } \mathrm{E}(\xi_{ij}|x_{ij}) = 1 \\ \mathrm{u}_{ij} = \exp(x_{ij}'\beta^0) \cdot (\xi_{ij} - 1) \text{ with } \mathrm{E}(u_{ij}|x_{ij}) = 0. \end{cases}$$

- The following moment conditions are

$$[\beta_0]: \quad \mathrm{E}(u_{ij}) = \mathrm{E}(y_{ij} - \exp(\beta_0^0 + \beta_1^0 x_{ij})) = 0, \text{ and}$$
$$[\beta_1]: \quad \mathrm{E}(x_{ij}u_{ij}) = \mathrm{E}(x_{ij}(y_{ij} - \exp(\beta_0^0 + \beta_1^0 x_{ij}))) = 0.$$

Issue 3: Invalid Inference with the Log-transformation

▸ Back

(i) The log-transformed error may fail to preserve the moment conditions of the error in the (original) constant elasticity model. (Silva and Tenreyro 2006)

To see this, consider a constant elasticity model:

$$y_{ij} = \exp(\beta_0^0 + \beta_1^0 x_{ij}) \cdot \xi_{ij} \Leftrightarrow y_{ij} = \exp(\beta_0^0 + \beta_1^0 x_{ij}) + u_{ij}, \tag{2}$$

where

$$\begin{cases} \xi_{ij} \text{ is the error with } \mathrm{E}(\xi_{ij}|x_{ij}) = 1 \\ \mathrm{u}_{ij} = \exp(x_{ij}'\beta^0) \cdot (\xi_{ij} - 1) \text{ with } \mathrm{E}(u_{ij}|x_{ij}) = 0. \end{cases}$$

- Now consider the log-transformation of (2) as

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

  where $v_{ij} = \ln \xi_{ij}$.
  - By Jensen's inequality, $\mathrm{E}(\xi_{ij}|x_{ij}) = 1$ does not imply $\mathrm{E}(v_{ij}|x_{ij}) = 0$.
    Why? $\mathrm{E}(v_{ij}|x_{ij}) = \mathrm{E}(\ln \xi_{ij}|x_{ij}) < \ln \mathrm{E}(\xi_{ij}|x_{ij}) = 0$.

▸ Eg

- Specifically, the Maclaurin series expansion for $\mathrm{E}\ln(\xi_{ij})$ gives

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

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

where $\xi_{ij}^{-} := \xi_{ij} - 1$ with $\mathrm{E}(\xi_{ij}^{-}|x_{ij}) = 0$, followed by $\mathrm{E}(\xi_{ij}^{-}) = 0$ and $\mathrm{E}(x_{ij}\xi_{ij}^{-}) = 0$.

  - Note that $\mathrm{E}(v_{ij})$ could deviate from zero when the higher-order moments of ξ_{ij}^{-} are non-zero (e.g., large variance, heavy tails, or high skewness).
  - $\mathrm{E}(x_{ij} v_{ij})$ could deviate from zero if the interaction between x_{ij} and $(\xi_{ij}^{-1})^{p}$ is non-zero (e.g., heteroskedasticity).
  - This is even more severe in the presence of spillovers due to the interactions among x_{ij}'s. ▸ Detail

Issue 3: Invalid Inference with the Log-transformation

(ii) Log transformation by shifting zero by some random constant is problematic. (Mullahy and Norton 2024)

- Intuition: Log-transformation distorts data structure due to the sharp slope of log function around zero.
- Consider $\frac{d\ln(y_{ij}+c)}{dy_{ij}} = \frac{1}{y_{ij}+c}$ for some $c > 0$ such that

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

  - A small change around $y_{ij} = 0$ ($c \to 0$) produces significantly distorted log-transformed outcome ($\ln(y_{ij}+c)$).
  - A large change around $y_{ij} = 0$ ($c \to \infty$) produces log-transformed outcome similar to the non-transformed one. However, adding $c \to \infty$ involves an *asymptotic bias* that grows to infinity for y_{ij} close to zero.

To see this, let $\hat{\beta}^{+}(c)$ be the OLS estimator when we employ $\ln(y_{ij}+c)$ as the dependent variable. The asymptotic bias is characterized as

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

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

and $\varepsilon_y > 0$ denotes an infinitesimal number.

Issue 3: Invalid Inference with the Log-transformation

Since

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

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

(iii) Log-transformed specification is *sensitive* to the error distributions, due to the specific feature of spatial econometric models.
To see this, recall the spatial gravity equation (LeSage and Pace 2008):

$$\ln(y_{ij}) = \sum_{k,l=1}^{n} s_{ij\leftarrow kl}(x'_{kl}\beta + v_{kl}),$$

or equivalently,

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

Observe how $\mathrm{E}\left(y_{ij}|\mathrm{z}\right)$ changes by the distributional assumption on $\{v_{kl}\}$:

1. If $v_{ij}|\mathrm{z} \overset{i.i.d.}{\sim} N(0,\sigma^2)$

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

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

(iii) Log-transformed specification is *sensitive* to the error distributions, due to the specific feature of spatial econometric models.
To see this, recall the spatial gravity equation (LeSage and Pace 2008):

$$\ln(y_{ij}) = \sum_{k,l=1}^{n} s_{ij\leftarrow kl}(x'_{kl}\beta + v_{kl}),$$

or equivalently,

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

2. If $v_{ij}|\mathrm{z} \overset{i.i.d.}{\sim} logGamma(\alpha,\rho)$ with $\frac{\alpha}{\rho}=1$,

$$\mathbb{E}(y_{ij}|\mathrm{z}) = \exp\left(\sum_{k,l=1}^{n} s_{ij,kl}x'_{kl}\beta\right)\frac{\alpha^{\sum_{k,l=1}^{n}s_{ij,kl}}\prod_{k,l=1}^{n}\Gamma(\alpha+s_{ij,kl})}{\Gamma(\alpha)^{n^2}},$$

since $\mathbb{E}(\exp(v_{kl})^{s_{ij,kl}}|\mathrm{z}) = \frac{\Gamma(\alpha+s_{ij,kl})}{\Gamma(\alpha)\rho^{s_{ij,kl}}} = \frac{\Gamma(\alpha+s_{ij,kl})\alpha^{s_{ij,kl}}}{\Gamma(\alpha)}$.

Recap: Poisson Regression

Consider a set of nonnegative discrete OD flows. Suppose $y_i|x_i \sim Poisson$.

- Regression equation:

$$\ln(\mathbb{E}y_i|x_i) = x_i'\beta$$

  so that

$$\begin{cases} \mu_i := \mathbb{E}y_i|x_i = \exp(x_i'\beta) & \text{(conditional mean)} \\ p(y_i|x_i) = \mu_i^{y_i}\exp(-\mu_i)/y_i! & \text{(prob. mass function)} \end{cases}$$

- Issue: Overdispersion
  - Mean=Variance in Poisson, but the variance of the observed data may be greater in the real world.

Poisson Pseudo Maximum Likelihood Estimation (PPMLE)

Gourieroux et al. (1984) ▸ Details

- Introduced the *specification error* in Poisson regression as

$$\mathbb{E}(y_i|x_i,\varepsilon_i) = \exp(x_i'\beta + \varepsilon_i) \quad \text{s.t.} \begin{cases} \varepsilon_i \perp\!\!\!\perp x_i \\ \mathbb{E}_{\varepsilon_i} \exp \varepsilon_i = 1 \\ \mathbb{V}_{\varepsilon_i} \exp \varepsilon_i = \nu^2, \end{cases}$$

so that

$$\begin{cases} \mathbb{E}_y y_i|x_i = \exp(x_i'\beta) \\ \mathbb{V}_y y_i|x_i = \exp(x_i'\beta) + \nu^2 \exp(2x_i'\beta). \end{cases}$$

i.e., Overdispersion is resolved!

- We introduce a PPMLE for spatial models with individual fixed effects:

$$y_{ij} = \mu_{ij}\xi_{ij},$$

Specification

PPMLE for Spatial Interaction Models

- Note on the individual fixed effects
  - Due to the non-linearity, can't get rid of them
  - Need to estimate
  - Solution: Introduce a scalar restriction in the likelihood

Asymptotics in Non-spatial Linear Model

- Consider
$$\mathbf{y}_N = \mathbf{X}_N\beta + \mathbf{D}_N\phi + \boldsymbol{\epsilon}_N,$$
  - Boldface indicates an N-dimensional vector or matrix.
  - $\mathbf{y}_N = (y_{11}, y_{21}, \ldots, y_{n1}, \ldots, y_{1n}, y_{2n}, \ldots, y_{nn})'$,
  - $\mathbf{X}_N = [x_{ij,k}]$ is an $N \times K$ matrix of regressors,
  - $\mathbf{D}_N = [\mathbf{I}_n \otimes 1_n, 1_n \otimes \mathbf{I}_n]$ is an $N \times 2n$ matrix of dummy variables, and
  - $\boldsymbol{\epsilon}_N = (\epsilon_{11}, \epsilon_{21}, \ldots, \epsilon_{n1}, \ldots, \epsilon_{1n}, \epsilon_{2n}, \ldots, \epsilon_{nn})'$ is an N-dimensional vector of disturbances.

- The log-likelihood function is
$$\ell_N(\beta, \phi) = -\frac{1}{2}(\mathbf{y}_N - \mathbf{X}_N\beta - \mathbf{D}_N\phi)'(\mathbf{y}_N - \mathbf{X}_N\beta - \mathbf{D}_N\phi) - \frac{1}{2}(v_{2n}'\phi)^2,$$
where $v_{2n} = (1_n', -1_n')'$.

Asymptotics in Non-spatial Linear Model

- The first-order conditions are
  - $[\beta] : \mathrm{X}_N'(\mathrm{y}_N - \mathrm{X}_N\beta - \mathrm{D}_N\phi) = 0$,
  - $[\phi] : \mathrm{D}_N'(\mathrm{y}_N - \mathrm{X}_N\beta - \mathrm{D}_N\phi) - v_{2n}v_{2n}'\phi = 0$.

- The second-order derivatives are

$$\partial_{\psi\psi}\ell_N(\beta,\phi) = -\begin{bmatrix}\mathrm{X}_N'\mathrm{X}_N & \mathrm{X}_N'\mathrm{D}_N \\ \mathrm{D}_N'\mathrm{X}_N & \mathrm{D}_N'\mathrm{D}_N\end{bmatrix} - \begin{bmatrix}0 & 0 \\ 0 & \begin{bmatrix}1 & -1 \\ -1 & 1\end{bmatrix} \otimes 1_n1_n'\end{bmatrix}$$

$$=: \begin{bmatrix}\mathrm{X}_N'\mathrm{X}_N & \mathrm{X}_N'\mathrm{D}_N \\ \mathrm{D}_N'\mathrm{X}_N & \widetilde{\mathrm{D}_N'\mathrm{D}_N}\end{bmatrix},$$

  - $\mathrm{D}_N'\mathrm{D}_N = \begin{bmatrix}n\mathrm{I}_n & 1_n1_n' \\ 1_n1_n' & n\mathrm{I}_n\end{bmatrix}$, $\widetilde{\mathrm{D}_N'\mathrm{D}_N} = \begin{bmatrix}n\mathrm{I}_n - 1_n1_n' & 2\cdot 1_n1_n' \\ 2\cdot 1_n1_n' & n\mathrm{I}_n - 1_n1_n'\end{bmatrix}$, and $\psi := (\beta', \phi')'$.
  - Note that
    (i) $\mathrm{rank}(\boldsymbol{D}_N'\boldsymbol{D}_N) = 2n - 1$ (*not* full) and $\mathrm{rank}(\widetilde{\boldsymbol{D}_N'\boldsymbol{D}_N}) = 2n$ (full).
    (ii) $\partial_{\psi\psi}\ell_N(\beta,\phi)$ does *not* depend on the parameters.

Asymptotics in Non-spatial Linear Model

- The quadratic expansion of $\ell_N(\boldsymbol{\psi})$ gives

$$\partial_{\boldsymbol{\psi}}\ell_N(\hat{\boldsymbol{\psi}}_N) = 0 = \partial_{\boldsymbol{\psi}}\ell_N(\boldsymbol{\psi}^0) + \partial_{\boldsymbol{\psi}\boldsymbol{\psi}}\ell_N(\boldsymbol{\psi}^0)(\hat{\boldsymbol{\psi}}_N - \boldsymbol{\psi}^0)$$
$$\Rightarrow (\hat{\boldsymbol{\psi}}_N - \boldsymbol{\psi}^0) = (-\partial_{\boldsymbol{\psi}\boldsymbol{\psi}}\ell_N(\boldsymbol{\psi}^0))^{-1}\partial_{\boldsymbol{\psi}}\ell_N(\boldsymbol{\psi}^0).$$

- Issue: Different convergence rate of β and ϕ
  - $\beta:\ N, \quad \phi:\ n$
  - Let $\boldsymbol{\Gamma} = \begin{pmatrix} N\mathrm{I}_K & 0 \\ 0 & n\mathrm{I}_{2n} \end{pmatrix}$ so that

$$(\hat{\boldsymbol{\psi}}_N - \boldsymbol{\psi}^0) = \left(\boldsymbol{\Gamma}^{-1/2}(-\partial_{\boldsymbol{\psi}\boldsymbol{\psi}}\ell_N(\boldsymbol{\psi}^0))\boldsymbol{\Gamma}^{-1/2}\right)^{-1}\left(\boldsymbol{\Gamma}^{-1/2}\partial_{\boldsymbol{\psi}}\ell_N(\boldsymbol{\psi}^0)\boldsymbol{\Gamma}^{-1/2}\right).$$

  - Then

$$\hat{\boldsymbol{\psi}}_N - \boldsymbol{\psi}^0 = \begin{bmatrix} \frac{1}{N}\mathrm{X}_N'\mathrm{X}_N & \frac{1}{n\sqrt{n}}\underbrace{\mathrm{X}_N'\mathrm{D}_N} \\ \frac{1}{n\sqrt{n}}\mathrm{D}_N'\mathrm{X}_N & \frac{1}{n}\mathrm{D}_N'\mathrm{D}_N \end{bmatrix}^{-1} \times \begin{pmatrix} \frac{1}{N}\mathrm{X}_N'\boldsymbol{\epsilon}_N \\ \frac{1}{n}\mathrm{D}_N'\boldsymbol{\epsilon}_N \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} \sqrt{N}(\hat{\beta}_N - \beta^0) \\ \sqrt{n}(\hat{\boldsymbol{\phi}}_{2n,N} - \boldsymbol{\phi}^0) \end{pmatrix} = \begin{bmatrix} \frac{1}{N}\mathrm{X}_N'\mathrm{X}_N & \frac{1}{n\sqrt{n}}\underbrace{\mathrm{X}_N'\mathrm{D}_N} \\ \frac{1}{n\sqrt{n}}\mathrm{D}_N'\mathrm{X}_N & \frac{1}{n}\mathrm{D}_N'\mathrm{D}_N \end{bmatrix}^{-1} \times \begin{pmatrix} \frac{1}{\sqrt{N}}\mathrm{X}_N'\boldsymbol{\epsilon}_N \\ \frac{1}{\sqrt{n}}\mathrm{D}_N'\boldsymbol{\epsilon}_N \end{pmatrix}.$$

- Thus, the approximated variance of $\begin{pmatrix} \sqrt{N}(\hat{\beta}_N - \beta^0) \\ \sqrt{n}(\hat{\phi}_{2n,N} - \phi^0) \end{pmatrix}$ is

$$\begin{bmatrix} \frac{1}{N}X'_N X_N & \frac{1}{n\sqrt{n}}X'_N D_N \\ \frac{1}{n\sqrt{n}}D'_N X_N & \frac{1}{n}\widetilde{D'_N D_N} \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{N}X'_N E(\epsilon_N \epsilon'_N) X_N & \frac{1}{n\sqrt{n}}X'_N E(\epsilon_N \epsilon'_N) D_N \\ \frac{1}{n\sqrt{n}}D'_N E(\epsilon_N \epsilon'_N) X_N & \frac{1}{n}D'_N E(\epsilon_N \epsilon'_N) D_N \end{bmatrix}$$
$$\times \begin{bmatrix} \frac{1}{N}X'_N X_N & \frac{1}{n\sqrt{n}}X'_N D_N \\ \frac{1}{n\sqrt{n}}D'_N X_N & \frac{1}{n}\widetilde{D'_N D_N} \end{bmatrix}^{-1}.$$

  - When the likelihood is correctly specified, the variance reduces to $\begin{bmatrix} \frac{1}{N}X'_N X_N & \frac{1}{n\sqrt{n}}X'_N D_N \\ \frac{1}{n\sqrt{n}}D'_N X_N & \frac{1}{n}\widetilde{D'_N D_N} \end{bmatrix}^{-1}.$

Asymptotics in Spatial Models

1. Due to our model's *nonlinearity*, the second-order derivatives rely on θ and ϕ.

   - Thus, we need consistent estimates of θ^0 and ϕ^0 for

$$\mathsf{E}\big(\Gamma^{-1/2}(-\partial_{\psi\psi}\ell_N(\theta^0))\Gamma^{-1/2}\big|\mathsf{z}\big)$$

$$=\begin{bmatrix} \frac{1}{N}\mathsf{G}_N'\mathsf{S}_N^{-1\prime}\mathrm{diag}(\boldsymbol{\mu}_N)\mathsf{S}_N^{-1}\mathsf{G}_N & \frac{1}{n\sqrt{n}}\mathsf{G}_N'\mathsf{S}_N^{-1\prime}\mathrm{diag}(\boldsymbol{\mu}_N)\mathsf{S}_N^{-1}\mathsf{D}_N \\ \frac{1}{n\sqrt{n}}\mathsf{D}_N'\mathsf{S}_N^{-1\prime}\mathrm{diag}(\boldsymbol{\mu}_N)\mathsf{S}_N^{-1}\mathsf{G}_N & \frac{1}{n}\mathsf{D}_N'\mathsf{S}_N^{-1\prime}\mathrm{diag}(\boldsymbol{\mu}_N)\mathsf{S}_N^{-1}\mathsf{D}_N + \frac{1}{n}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \otimes 1_n 1_n' \end{bmatrix},$$

with $\mathsf{G}_N := \mathsf{G}_N(\theta^0) = [\mathsf{W}_N\mathsf{S}_N^{-1}\mathsf{z}_N, \mathsf{M}_N\mathsf{S}_N^{-1}\mathsf{z}_N, \mathsf{R}_N\mathsf{S}_N^{-1}\mathsf{z}_N, \mathsf{X}_N]$,
$\boldsymbol{\mu}_N = \boldsymbol{\mu}_N(\theta^0) = (\exp(\tilde{\mu}_{11}), \ldots, \exp(\tilde{\mu}_{n1}), \ldots, \exp(\tilde{\mu}_{1n}), \ldots, \exp(\tilde{\mu}_{nn}))$,
$\tilde{\boldsymbol{\mu}}_N = \tilde{\boldsymbol{\mu}}_N(\theta^0) = (\tilde{\mu}_{11}, \ldots, \tilde{\mu}_{n1}, \ldots, \tilde{\mu}_{1n}, \ldots, \tilde{\mu}_{nn})$,
$\tilde{\boldsymbol{\mu}}_N = \mathsf{S}_N^{-1}(\mathsf{X}_N\beta^0 + \boldsymbol{\alpha}^0 \otimes 1_n + 1_n \otimes \boldsymbol{\eta}^0) = \mathsf{S}_N^{-1}\mathsf{z}_N$,
and $\boldsymbol{\psi} := (\theta', \boldsymbol{\phi}')'$.

2. Observe that

$$\begin{pmatrix} \frac{1}{\sqrt{N}}\partial_\theta \ell_N(\theta,\boldsymbol{\phi}) \\ \frac{1}{\sqrt{n}}\partial_{\boldsymbol{\phi}} \ell_N(\theta,\boldsymbol{\phi}) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{N}}(\mathsf{S}_N^{-1}(\lambda)\mathsf{G}_N(\theta,\boldsymbol{\phi}))'\boldsymbol{\epsilon}_N(\theta,\boldsymbol{\phi}) \\ \frac{1}{\sqrt{n}}(\mathsf{S}_N^{-1}(\lambda)\mathsf{D}_N)'\boldsymbol{\epsilon}_N(\theta,\boldsymbol{\phi}) \end{pmatrix},$$

where
$\mathsf{G}_N(\theta,\boldsymbol{\phi}) = [\mathsf{W}_N\mathsf{S}_N^{-1}(\lambda)\mathsf{z}_N(\beta,\boldsymbol{\phi}), \mathsf{M}_N\mathsf{S}_N^{-1}(\lambda)\mathsf{z}_N(\beta,\boldsymbol{\phi}), \mathsf{R}_N\mathsf{S}_N^{-1}(\lambda)\mathsf{z}_N(\beta,\boldsymbol{\phi}), \mathsf{X}_N]$,
and $\mathsf{z}_N(\beta,\boldsymbol{\phi}) = \mathsf{X}_N\beta + \boldsymbol{\alpha}\otimes 1_n + 1_n \otimes \boldsymbol{\eta}$.

- The variance is then

$$\mathsf{V}\left(\begin{pmatrix} \frac{1}{\sqrt{N}}(\mathsf{S}_N^{-1}\mathsf{G}_N)'\boldsymbol{\epsilon}_N \\ \frac{1}{\sqrt{n}}(\mathsf{S}_N^{-1}\mathsf{D}_N)'\boldsymbol{\epsilon}_N \end{pmatrix} \middle| \mathsf{z} \right) = \begin{bmatrix} \frac{1}{N}\mathsf{G}_N'\mathsf{S}_N^{-1\prime}\mathsf{E}(\boldsymbol{\epsilon}_N\boldsymbol{\epsilon}_N')\mathsf{S}_N^{-1}\mathsf{G}_N & \frac{1}{n\sqrt{n}}\mathsf{G}_N'\mathsf{S}_N^{-1\prime}\mathsf{E}(\boldsymbol{\epsilon}_N\boldsymbol{\epsilon}_N')\mathsf{S}_N^{-1}\mathsf{D}_N \\ \frac{1}{n\sqrt{n}}\mathsf{D}_N'\mathsf{S}_N^{-1\prime}\mathsf{E}(\boldsymbol{\epsilon}_N\boldsymbol{\epsilon}_N')\mathsf{S}_N^{-1}\mathsf{G}_N & \frac{1}{n}\mathsf{D}_N'\mathsf{S}_N^{-1\prime}\mathsf{E}(\boldsymbol{\epsilon}_N\boldsymbol{\epsilon}_N')\mathsf{S}_N^{-1}\mathsf{D}_N \end{bmatrix}.$$

Heteroscedastic and Spatially Correlated Errors

- Since we do *not* impose a specification structure on $\text{Cov}(\epsilon_{ij}, \epsilon_{kl})$ but weak correlation among $\{\epsilon_{ij}\}$, our goal is to consistently estimate the variance.

- Observe that

$$\mathrm{E}(\boldsymbol{\epsilon}_N \boldsymbol{\epsilon}_N') = \begin{bmatrix} \sigma_{11,11} & \cdots & \sigma_{11,n1} & \cdots & \sigma_{11,1n} & \cdots & \sigma_{11,nn} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \sigma_{nn,11} & \cdots & \sigma_{nn,n1} & \cdots & \sigma_{nn,1n} & \cdots & \sigma_{nn,nn} \end{bmatrix},$$

  where $\sigma_{ij,kl} = \text{Cov}(\epsilon_{ij}, \epsilon_{kl})$.

  - $\mathrm{E}(\boldsymbol{\epsilon}_N \boldsymbol{\epsilon}_N')$ involves $N + \frac{N(N-1)}{2}$ unknown variance and covariance parameters, exceeding our N observation.

- Problem: The variance does *not* converge very well.
  - Solution: Reduce the number of effective terms in $\mathrm{E}(\boldsymbol{\epsilon}_N \boldsymbol{\epsilon}_N')$ for convergence.

Recap: **heteroskedasticity-Autocorrelation-Consistent (HAC)** Standard Errors in Time-series Literature (Newey & West 1987; Andrews 1991)

- Consider $y_t = x_t'\beta + \varepsilon_t, \quad t = 1, \ldots, T.$
  - Let $\hat{\beta} = (X'X)^{-1}X'y.$
- Suppose $\mathrm{V}(\varepsilon|X) = \sigma^2\Omega$, where $\Omega \neq I_T$.
  - i.e., variances differ across observations (heteroskedasticity) and non-zero correlation across observations (autocorrelation).
- Denote

$$\mathrm{plim}(X'X/T) = Q_{XX},$$
$$\mathrm{plim}(X'\Omega X/T) = Q_{X\Omega X} =: Q^*.$$

- The true variance of $\hat{\beta}$ is then

$$\mathrm{V}_T(\hat{\beta}|X) = \sigma^2(X'X)^{-1}(\underbrace{X'\Omega X}_{:=TQ^*})(X'X)^{-1}.$$

Recap: **heteroskedasticity-Autocorrelation-Consistent (HAC)** Standard Errors in Time-series Literature (Newey & West 1987; Andrews 1991)

- Need to estimate

$$Q^* := \frac{1}{T} X'\Omega X = \frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \sigma_{ts} x_t x_s'.$$

  - Use the residuals to estimate covariances, i.e., $\hat{\varepsilon}_t \hat{\varepsilon}_s$ to estimate σ_{ts}.

- Problem: This sum has T^2 terms. Difficult to get convergence.
  - Solution: <u>Cut short the sum.</u> Usually, use weights in the sum that imply that the process becomes *less* autocorrelated as time goes by.

- **HAC estimator:**

$$\hat{Q}^* = \underbrace{\frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t^2 x_t x_t'}_{\text{Robust to heteroskedasticity}} + \underbrace{\frac{1}{T} \sum_{\ell=1}^{L} \sum_{t=\ell+1}^{T} w_\ell \hat{\varepsilon}_t \hat{\varepsilon}_{t-\ell} (x_t x_{t-\ell}' + x_{t-\ell} x_t'),}_{\text{Robust to autocorrelations}}$$

where $w_\ell = 1 - \frac{\ell}{L+1}$ is the Bartlett kernel..

heteroskedasticity-Spatial-Correlations-Robust Standard Errors

In our context, we leverage the sparsity of spatial weight matrices.

- i.e., regard spatial weight matrices as similar to a kernel function for truncation.

- n=49; Replication=1,000.
  - Cross-sectional units: US states
- True parameters
  - $\beta^{0\prime} = (0.95, -0.85, 0.65, -0.75)'$
  - X_1: Continuous; X_2, X_3: Binary; X_4: Continuous (Distance)
  - Spatial dependence parameters
    - (i) Absence: $(\lambda_d^0, \lambda_o^0, \lambda_w^0) = (0, 0, 0)$ or
    - (ii) Presence: $(\lambda_d^0, \lambda_o^0, \lambda_w^0) = (0.25, 0.25, 0.15)$.
- Spatial weight matrices
  1. Constructed by *geographic proximity*
  2. Constructed by proximity in some trigger *dependent* on X
  3. Composite of 1 & 2

Spatial gravity equation (LeSage and Pace 2008) with individual fixed effects

Back

- Augmenting spatial terms,

$$\ln y_{ij} = x_{ij}'\beta + \alpha_j + \eta_i + error_{ij},$$

$$+ \underbrace{\lambda_d \sum_{k=1}^{n} w_{ik} \ln y_{kj}}_{\text{Destination-based dependence}} + \underbrace{\lambda_o \sum_{l=1}^{n} m_{lj} \ln y_{il}}_{\text{Origin-based dependence}} + \underbrace{\lambda_w \sum_{k,l=1}^{n} w_{ik} m_{lj} \ln y_{kl}}_{\text{origin-to-destination dependence}}$$

- This reduces to

Derivation

$$\ln y_{ij} = \sum_{k,l=1}^{n} s_{ij \leftarrow kl}(x_{kl}'\beta + \alpha_l + \eta_k + error_{kl}),$$

where $s_{ij \leftarrow kl}$ represents the signal from pair (k, l) to pair (i, j) as an element in $S_N^{-1} := (\mathrm{I}_N - \lambda_d W_N - \lambda_o M_n - \lambda_w R_N)^{-1}$.

- Introduced the *specification error* in Poisson regression as

$$\mathbb{E} Y_i | X_i, \varepsilon_i = \exp(X_i \beta + \varepsilon_i) \quad \text{s.t.} \begin{cases} \varepsilon_i \perp\!\!\!\perp X_i \\ \mathbb{E}_{\varepsilon_i} \exp \varepsilon_i = 1 \\ \mathbb{V}_{\varepsilon_i} \exp \varepsilon_i = \nu^2, \end{cases}$$

so that

$$\begin{cases} \mathbb{E}_Y Y_i | X_i & = \mathbb{E}_\varepsilon (\mathbb{E}_Y Y_i | X_i, \varepsilon_i) | X_i \\ & = \mathbb{E}_\varepsilon \exp(X_i \beta + \varepsilon_i) | X_i \\ & = \exp(X_i \beta) \mathbb{E}_\varepsilon \underbrace{\exp \varepsilon_i | X_i}_{\varepsilon_i \perp\!\!\!\perp X_i} \\ & = \exp(X_i \beta) \underbrace{\mathbb{E}_\varepsilon \exp \varepsilon_i}_{=1} \\ & = \exp(X_i \beta), \end{cases}$$

Poisson Pseudo MLE (Gourieroux et al. 1984) Back

- Introduced the specification error in Poisson regression as

$$\mathbb{E}Y_i|X_i,\varepsilon_i = \exp(X_i\beta+\varepsilon_i) \quad \text{s.t.} \quad \begin{cases} \varepsilon_i \perp\!\!\!\perp X_i \\ \mathbb{E}_{\varepsilon_i}\exp\varepsilon_i = 1 \\ \mathbb{V}_{\varepsilon_i}\exp\varepsilon_i = \nu^2, \end{cases}$$

so that

$$\begin{cases} \mathbb{V}_Y Y_i|X_i & = \mathbb{E}_\varepsilon(\underbrace{\mathbb{V}_Y Y_i|X_i,\varepsilon_i}_{=\exp(X_i\beta+\varepsilon_i)})|X_i + \mathbb{V}_\varepsilon(\underbrace{\mathbb{E}_Y Y_i|X_i,\varepsilon_i}_{=\exp(X_i\beta+\varepsilon_i)})|X_i \\ & = \mathbb{E}_\varepsilon\exp(X_i\beta+\varepsilon_i)|X_i + \mathbb{V}_\varepsilon\exp(X_i\beta+\varepsilon_i)|X_i \\ & = \exp(X_i\beta)\underbrace{\mathbb{E}_\varepsilon\exp\varepsilon_i|X_i}_{=1} + \exp(2X_i\beta)\underbrace{\mathbb{V}_\varepsilon\exp\varepsilon_i|X_i}_{=\nu^2} \\ & = \exp(X_i\beta) + \nu^2\exp(2X_i\beta). \end{cases}$$

Poisson Pseudo MLE (Gourieroux et al. 1984) ▸ Back

- ML estimator for β: Consider a likelihood

$$L(\beta) = \prod_{i=1}^{n} \frac{\exp((x_i\beta)y_i)\exp(-\exp(x_i\beta))}{y_i!}$$

  - Log-likelihood:

$$\ln L(\beta) = \sum_{i=1}^{n} y_i x_i \beta - \sum_{i=1}^{n} \exp(x_i\beta) - \sum_{i=1}^{n} \ln y_i!$$

  - Pseudo likelihood equations:

$$\sum_{i=1}^{n} x_i(-\exp(x_i\hat{\beta}) + y_i) \overset{\text{set}}{=} 0.$$

- Consistent estimator for ν^2: Use $\mathbb{V}Y_i = \exp(X_i\beta) + \nu^2\exp(2X_i\beta)$ & $\hat{\beta}$.

Derivation of Spatial Gravity Equation Back

Spatial gravity equation (LeSage and Pace 2008)

Stacking up,

$$\begin{bmatrix} \ln(y_{\mathbf{11}}+1) \\ \ln(y_{\mathbf{21}}+1) \\ \vdots \\ \ln(y_{n\mathbf{1}}+1) \\ \vdots \end{bmatrix} = \begin{bmatrix} x'_{\mathbf{11}}\beta + \lambda_d \sum_k w_{\mathbf{1}k} \ln(y_{k\mathbf{1}}+1) + \lambda_{\mathbf{0}} \sum_l m_{l\mathbf{1}} \ln(y_{\mathbf{1}l}+1) + \lambda_w \sum_{k,l}^n w_{\mathbf{1}k} m_{l\mathbf{1}} \ln(y_{kl}+1) + error_{\mathbf{11}} \\ x'_{\mathbf{21}}\beta + \lambda_d \sum_k w_{\mathbf{2}k} \ln(y_{k\mathbf{1}}+1) + \lambda_{\mathbf{0}} \sum_l m_{l\mathbf{1}} \ln(y_{\mathbf{12}}+1) + \lambda_w \sum_{k,l}^n w_{\mathbf{2}k} m_{l\mathbf{1}} \ln(y_{kl}+1) + error_{\mathbf{21}} \\ \vdots \\ x'_{n\mathbf{1}}\beta + \lambda_d \sum_k w_{nk} \ln(y_{k\mathbf{1}}+1) + \lambda_{\mathbf{0}} \sum_l m_{l\mathbf{1}} \ln(y_{nl}+1) + \lambda_w \sum_{k,l}^n w_{nk} m_{l\mathbf{1}} \ln(y_{kl}+1) + error_{n\mathbf{1}} \\ \vdots \end{bmatrix}$$

Moving terms,

$$\begin{bmatrix} \ln(y_{\mathbf{11}}+1) - \lambda_d \sum_k w_{\mathbf{1}k} \ln(y_{k\mathbf{1}}+1) - \lambda_{\mathbf{0}} \sum_l m_{l\mathbf{1}} \ln(y_{\mathbf{1}l}+1) - \lambda_w \sum_{k,l}^n w_{\mathbf{1}k} m_{l\mathbf{1}} \ln(y_{kl}+1) \\ \ln(y_{\mathbf{21}}+1) - \lambda_d \sum_k w_{\mathbf{2}k} \ln(y_{k\mathbf{1}}+1) - \lambda_{\mathbf{0}} \sum_l m_{l\mathbf{1}} \ln(y_{\mathbf{12}}+1) - \lambda_w \sum_{k,l}^n w_{\mathbf{2}k} m_{l\mathbf{1}} \ln(y_{kl}+1) \\ \vdots \\ \ln(y_{n\mathbf{1}}+1) - \lambda_d \sum_k w_{nk} \ln(y_{k\mathbf{1}}+1) - \lambda_{\mathbf{0}} \sum_l m_{l\mathbf{1}} \ln(y_{nl}+1) - \lambda_w \sum_{k,l}^n w_{nk} m_{l\mathbf{1}} \ln(y_{kl}+1) \\ \vdots \end{bmatrix} = \begin{bmatrix} x'_{\mathbf{11}}\beta + error_{\mathbf{11}} \\ x'_{\mathbf{21}}\beta + error_{\mathbf{21}} \\ \vdots \\ x'_{n\mathbf{1}}\beta + error_{n\mathbf{1}} \\ \vdots \end{bmatrix}$$

In a vectorized form,

$$\ln(\mathrm{Y}_N + 1) - \lambda_d(I_n \otimes W_n)\ln(\mathrm{Y}_N + 1) - \lambda_o(M_n' \otimes I_n)\ln(\mathrm{Y}_N + 1) - \lambda_w(M_n' \otimes W_n)\ln(\mathrm{Y}_N + 1)$$
$$= \mathrm{X}_N\beta + \mathrm{error}_N$$

$$\rightarrow (\mathrm{I}_N - \lambda_d \mathrm{W}_N - \lambda_o \mathrm{M}_N - \lambda_w \mathrm{R}_N)\ln(\mathrm{Y}_N + 1) = \mathrm{X}_N\beta + \mathrm{error}_N$$

Derivation of Spatial Gravity Equation

Spatial gravity equation (LeSage and Pace 2008)

In a vectorized form,

$$\ln(\mathrm{Y}_N + 1) - \lambda_d (I_n \otimes W_n) \ln(\mathrm{Y}_N + 1) - \lambda_o (M_n' \otimes I_n) \ln(\mathrm{Y}_N + 1) - \lambda_w (M_n' \otimes W_n) \ln(\mathrm{Y}_N + 1) = \mathrm{X}_N \beta + \mathrm{error}_N$$

$$\rightarrow \underbrace{(\mathrm{I}_N - \lambda_d \mathrm{W}_N - \lambda_o \mathrm{M}_n - \lambda_w \mathrm{R}_N)}_{=:\mathrm{S}_N} \ln(\mathrm{Y}_N + 1) = \mathrm{X}_N \beta + \mathrm{error}_N$$

$$\rightarrow \ln(\mathrm{Y}_N + 1) = \mathrm{S}_N^{-1} (\mathrm{X}_N \beta + \mathrm{error}_N), \quad \text{given } \mathrm{S}_N \text{ is invertible.}$$

Derivation of Spatial Gravity Equation ▸ Back

Spatial gravity equation (LeSage and Pace 2008)

Thus,

$$\begin{aligned}\ln(y_{ij}+1) = \lambda_d \sum_{k=1}^{n} w_{ik} \ln(y_{kj}+1) \;+\; \lambda_o \sum_{l=1}^{n} m_{lj} \ln(y_{il}+1) \\ +\; \lambda_w \sum_{k,l=1}^{n} w_{ik} m_{lj} \ln(y_{kl}+1) \;+\; x_{ij}'\beta \;+\; error_{ij},\end{aligned} \tag{1}$$

reduces to

$$\ln(y_{ij}+1) = \sum_{k,l=1}^{n} s_{ij\leftarrow kl}(x_{kl}'\beta + error_{kl}),$$

where $s_{ij\leftarrow kl}$ represents the signal from pair (k,l) to pair (i,j) as an element in S_N^{-1}.

Derivation of Spatial Gravity Equation ▸ Back

Spatial gravity equation (LeSage and Pace 2008) with individual fixed effects

Stacking up,

$$\begin{bmatrix} \ln y_{11} \\ \ln y_{21} \\ \vdots \\ \ln y_{n1} \\ \vdots \end{bmatrix} = \begin{bmatrix} x_{11}'\beta + \alpha_1 + \eta_1 + \lambda_d \sum_k w_{1k} \ln y_{k1} + \lambda_0 \sum_l m_{l1} \ln y_{1l} + \lambda_w \sum_{k,l}^n w_{1k} m_{l1} \ln y_{kl} + error_{11} \\ x_{21}'\beta + \alpha_1 + \eta_2 + \lambda_d \sum_k w_{2k} \ln y_{k1} + \lambda_0 \sum_l m_{l1} \ln y_{12} + \lambda_w \sum_{k,l}^n w_{2k} m_{l1} \ln y_{kl} + error_{21} \\ \vdots \\ x_{n1}'\beta + \alpha_1 + \boldsymbol{\eta} + \lambda_d \sum_k w_{nk} \ln y_{k1} + \lambda_0 \sum_l m_{l1} \ln y_{nl} + \lambda_w \sum_{k,l}^n w_{nk} m_{l1} \ln y_{kl} + error_{n1} \\ \vdots \end{bmatrix}$$

Moving terms,

$$\begin{bmatrix} \ln y_{11} - \lambda_d \sum_k w_{1k} \ln y_{k1} - \lambda_0 \sum_l m_{l1} \ln y_{1l} - \lambda_w \sum_{k,l}^n w_{1k} m_{l1} \ln y_{kl} \\ \ln y_{21} - \lambda_d \sum_k w_{2k} \ln y_{k1} - \lambda_0 \sum_l m_{l1} \ln y_{12} - \lambda_w \sum_{k,l}^n w_{2k} m_{l1} \ln y_{kl} \\ \vdots \\ \ln y_{n1} - \lambda_d \sum_k w_{nk} \ln y_{k1} - \lambda_0 \sum_l m_{l1} \ln y_{nl} - \lambda_w \sum_{k,l}^n w_{nk} m_{l1} \ln y_{kl} \\ \vdots \end{bmatrix} = \begin{bmatrix} x_{11}'\beta + \alpha_1 + \eta_1 + error_{11} \\ x_{21}'\beta + \alpha_1 + \eta_2 + error_{21} \\ \vdots \\ x_{n1}'\beta + \alpha_1 + \boldsymbol{\eta} + error_{n1} \\ \vdots \end{bmatrix}$$

Derivation of Spatial Gravity Equation ▸ Back

Spatial gravity equation (LeSage and Pace 2008) with individual fixed effects

In a vectorized form,

$$\begin{aligned}\ln \mathrm{Y}_N - \lambda_d (I_n \otimes W_n) \ln \mathrm{Y}_N - \lambda_o (M_n' \otimes I_n) \ln \mathrm{Y}_N - \lambda_w (M_n' \otimes W_n) \ln \mathrm{Y}_N \\ = \mathrm{X}_N \beta + \boldsymbol{\alpha} \otimes 1_n + 1_n \otimes \boldsymbol{\eta} + \mathrm{error}_N\end{aligned}$$

$$\rightarrow (\mathrm{I}_N - \lambda_d \mathrm{W}_N - \lambda_o \mathrm{M}_N - \lambda_w \mathrm{R}_N) \ln \mathrm{Y}_N = \mathrm{X}_N \beta + \boldsymbol{\alpha} \otimes 1_n + 1_n \otimes \boldsymbol{\eta} + \mathrm{error}_N$$

In a vectorized form,

$$\ln \mathsf{Y}_N - \lambda_d (I_n \otimes W_n) \ln \mathsf{Y}_N - \lambda_o (M_n' \otimes I_n) \ln \mathsf{Y}_N - \lambda_w (M_n' \otimes W_n) \ln \mathsf{Y}_N$$
$$= \mathsf{X}_N \beta + \boldsymbol{\alpha} \otimes 1_n + 1_n \otimes \boldsymbol{\eta} + \text{error}_N$$

$$\rightarrow (\underbrace{\mathsf{I}_N - \lambda_d \mathsf{W}_N - \lambda_o \mathsf{M}_N - \lambda_w \mathsf{R}_N}_{=:\mathsf{S}_N}) \ln \mathsf{Y}_N = \mathsf{X}_N \beta + \boldsymbol{\alpha} \otimes 1_n + 1_n \otimes \boldsymbol{\eta} + \text{error}_N$$

$$\rightarrow \ln \mathsf{Y}_N = \mathsf{S}_N^{-1} (\mathsf{X}_N \beta + \boldsymbol{\alpha} \otimes 1_n + 1_n \otimes \boldsymbol{\eta} + \text{error}_N), \quad \text{given } \mathsf{S}_N \text{ is invertible.}$$

Derivation of Spatial Gravity Equation ▸ Back

Spatial gravity equation (LeSage and Pace 2008) with individual fixed effects

Thus,

$$\begin{aligned} \ln y_{ij} &= x_{ij}'\beta \ + \ \alpha_j \ + \ \eta_i \ + \ error_{ij}, \\ &+ \lambda_d \sum_{k=1}^{n} w_{ik} \ln y_{kj} \ + \ \lambda_o \sum_{l=1}^{n} m_{lj} \ln y_{il} \ + \ \lambda_w \sum_{k,l=1}^{n} w_{ik} m_{lj} \ln y_{kl} \end{aligned} \tag{1}$$

reduces to

$$\ln y_{ij} = \sum_{k,l=1}^{n} s_{ij \leftarrow kl}(x_{kl}'\beta + \alpha_l + \eta_k + error_{kl}),$$

where $s_{ij \leftarrow kl}$ represents the signal from pair (k, l) to pair (i, j) as an element in S_N^{-1}.

- Consider a univariate SAR model:

$$Y = \lambda WY + X\beta + U$$
$$\Rightarrow Y = S^{-1}(X\beta + U),$$

where $S^{-1} = (I - \lambda W)^{-1} = I + \lambda W + (\lambda W)^2 + \cdots \simeq (I + \lambda W)$.

- Eg. $n = 3$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \underbrace{\left(\begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix} - \lambda \begin{pmatrix} 0 & w_{12} & w_{13} \\ w_{21} & 0 & w_{23} \\ w_{31} & w_{32} & 0 \end{pmatrix} \right)^{-1}}_{=S^{-1}} \left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \beta + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \right)$$

$$\simeq \left(\begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix} + \lambda \begin{pmatrix} 0 & w_{12} & w_{13} \\ w_{21} & 0 & w_{23} \\ w_{31} & w_{32} & 0 \end{pmatrix} \right) \left(\begin{bmatrix} x_1\beta + u_1 \\ x_2\beta + u_2 \\ x_3\beta + u_3 \end{bmatrix} \right)$$

$$= \left(\begin{pmatrix} 1 & \lambda w_{12} & \lambda w_{13} \\ \lambda w_{21} & 1 & \lambda w_{23} \\ \lambda w_{31} & \lambda w_{32} & 1 \end{pmatrix} \right) \left(\begin{bmatrix} x_1\beta + u_1 \\ x_2\beta + u_2 \\ x_3\beta + u_3 \end{bmatrix} \right).$$

Hence

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \simeq \begin{bmatrix} 1\cdot(x_1\beta+u_1)+\lambda w_{12}(x_2\beta+u_2)+\lambda w_{13}(x_3\beta+u_3) \\ \lambda w_{21}(x_1\beta+u_1)+1\cdot(x_2\beta+u_2)+\lambda w_{23}(x_3\beta+u_3) \\ \lambda w_{31}(x_1\beta+u_1)+\lambda w_{32}(x_2\beta+u_2)+1\cdot(x_3\beta+u_3) \end{bmatrix}$$

$$= \begin{bmatrix} S^{-1}_{(1,1)}(x_1\beta+u_1)+S^{-1}_{(1,2)}(x_2\beta+u_2)+S^{-1}_{(1,3)}(x_3\beta+u_3) \\ S^{-1}_{(2,1)}(x_1\beta+u_1)+S^{-1}_{(2,2)}(x_2\beta+u_2)+S^{-1}_{(2,3)}(x_3\beta+u_3) \\ S^{-1}_{(3,1)}(x_1\beta+u_1)+S^{-1}_{(3,2)}(x_2\beta+u_2)+S^{-1}_{(3,3)}(x_3\beta+u_3) \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{k=1}^{3}[S^{-1}]_{(1,k)}(x_k\beta+u_k) \\ \sum_{k=1}^{3}[S^{-1}]_{(2,k)}(x_k\beta+u_k) \\ \sum_{k=1}^{3}[S^{-1}]_{(3,k)}(x_k\beta+u_k) \end{bmatrix}.$$

Data Structure of OD Flows ▸ Back

Jieun/PT (20m)/Figs/Lesage and Pace (2009) - Table

(i) Suppose y_{ij} with origin i and destination j. Then Y is defined as

Jieun/PT (20m)/Figs/Screenshot 2024-09-30 at 11.35.42âĂŕAM

Constant Elasticity Model (Recap) Back

Consider a Cobb-Dougals function

$$Y = AX^{\alpha} Z^{\beta}.$$

Elasticities are *constant*:

$$\begin{cases} \frac{\partial Y/Y}{\partial X/X} & = \frac{\partial Y}{\partial X} \cdot \frac{X}{Y} \\ & = \alpha \cdot (AX^{\alpha-1} Z^{\beta}) \cdot \frac{X}{Y} \\ & = \alpha \cdot \frac{\cancel{AX^{\alpha} Z^{\beta}}}{\cancel{Y}} = \alpha, \\ \frac{\partial Y/Y}{\partial Z/Z} & = \frac{\partial Y}{\partial Z} \cdot \frac{Z}{Y} \\ & = \beta \cdot (AX^{\alpha} Z^{\beta-1}) \cdot \frac{Z}{Y} \\ & = \beta \cdot \frac{\cancel{AX^{\alpha} Z^{\beta}}}{\cancel{Y}} = \beta. \end{cases}$$

Rank of $D_N'D_N$ and $\widetilde{D_N'D_N}$ Back

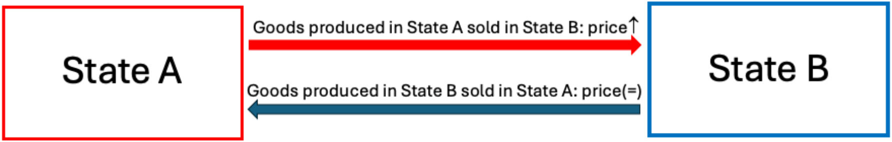
TBD

Background ‣ Back

Data Structure & Origin-Destination Flows

- Note. Usefulness and broad applicability of OD flows in applied micro
  - Well-suited for analyzing heterogeneous treatment effects of policy interventions that may *vary* across cross-sectional units.
  - Eg. How does a carbon tax policy in State A on goods produced locally drive *heterogeneous* economic activities between State A and State B, where State B does not have such a policy?

Figure 4: Directional heterogeneities in cross-state economic activities



Issue 3: Invalid Inference with the Log-transformed Specification

Examples

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

   Note. This gap is sensitive to the *unit* of the outcome. To see this, let $y_{ij}^* := 0.01 \cdot y_{ij}$.
   - Observe that $\ln(E(y_{ij})) = \ln(0.01 \cdot 1 \cdot 0.5 + 0.01 \cdot 0 \cdot 0.5) \approx -5.2983$.
   - Now consider $E \ln(y_{ij})$. As zero is not defined in the log function, we need to add some arbitrary constant, say 1.
     Then $E \ln(y_{ij} + 1) = 0.5 \cdot \ln(0.01 \cdot 1 + 1) + 0.5 \cdot \ln(0.01 \cdot 0 + 1) \approx 0.0050$.
   - The gap is now about 5.3033.

Violation of the moment condition in the presence of spillovers ▸ Back

For example, suppose $\mathrm{E}((\xi_{ij}^{-})^2|\mathrm{x}) = c_0 + c_1 x_{ij}^2 + c_2 x_{kj}^2 + c_3 x_{il}^2$, where $c_0, \ldots, c_3 > 0$, k is an i's neighbor, and l is a j's neighbor. In this case,

$$\begin{aligned}
&\mathrm{E}(x_{ij}(\xi_{ij}^{-}))^2 \\
&= \mathrm{E}(x_{ij}\mathrm{E}((\xi_{ij}^{-})^2|x_{ij}, x_{kj}, x_{il})) \\
&= c_o\mathrm{E}(x_{ij}) + c_1\mathrm{E}(x_{ij}^3) + c_2\mathrm{E}(x_{ij}x_{kj}^2) + c_3\mathrm{E}(x_{ij}x_{il}^2).
\end{aligned}$$

Note.

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

PPML: No Bias Due to Incidental Parameters ▸ Back

With $\mu_{ij} = \exp(x_{ij}'\beta)\exp(\alpha_i)\exp(\eta_j)$,

$$\begin{aligned}
\ell &= \sum_{i,j}(y_{ij}\ln\mu_{ij} - \mu_{ij}) \\
&= \sum_{i,j}[y_{ij}(x_{ij}'\beta + \alpha_i + \eta_j) - \exp(x_{ij}'\beta + \alpha_i + \eta_j)] \\
&= \sum_{i,j} y_{ij}x_{ij}'\beta + \sum_i \alpha_i \sum_j y_{ij} + \sum_j \eta_j \sum_i y_{ij} - \sum_{i,j}\mu_{ij},
\end{aligned}$$

FOCs are

$$\begin{cases}
\frac{\partial\ell}{\partial\alpha_i} = \sum_j y_{ij} - \sum_j \mu_{ij} = 0 \Rightarrow \sum_j \mu_{ij} = \sum_j y_{ij} & \forall i \\
\frac{\partial\ell}{\partial\eta_j} = \sum_i y_{ij} - \sum_i \mu_{ij} = 0 \Rightarrow \sum_i \mu_{ij} = \sum_i y_{ij} & \forall j.
\end{cases}$$

PPML: No Bias Due to Incidental Parameters (Cont'd) ▸ Back

Using the row FOC,

$$\underbrace{\sum_j y_{ij}}_{:=S_i} = \sum_j \mu_{ij} = \sum_j \exp(x_{ij}'\beta + \alpha_i + \eta_j) = \exp(\alpha_i) \sum_j \exp(x_{ij}'\beta + \eta_j)$$

$$\Rightarrow \mu_{ij} = S_i \cdot \frac{\exp(x_{ij}'\beta + \eta_j)}{\sum_j \exp(x_{ij}'\beta + \eta_j)} - (*).$$

Note that α_i is eliminated. Now using the column FOC,

$$\underbrace{\sum_i y_{ij}}_{=:C_j} = \sum_i \mu_{ij} \overset{(*)}{=} \sum_i S_i \cdot \frac{\exp(x_{ij}'\beta + \eta_j)}{\sum_j \exp(x_{ij}'\beta + \eta_j)}.$$

Solving for η_j as a function of β, η_j is profiled out.
Thus, μ_{ij} in $(*)$ is concentrated out with respect to β.
Thus, $\frac{\partial \ell}{\partial \beta} = \sum_{i,j} x_{ij}(y_{ij} - \mu_{ij})$ is a function of β, free from FE.

Number of Restrictions ▸ Back

Consider

$$Y = X\beta + F\Lambda' + \varepsilon.$$

- Identification issue on factors and their loadings
  - Consider $F\Lambda' = FAA^{-1}\Lambda'$ with an arbitrary invertible $R \times R$ matrix A.
    $\rightarrow$ This gives infinite solutions of (F, Λ) (a.k.a. *rotational indeterminacy*)
  - Idea: **Fix** A with R^2 free elements. Thus, we need R^2 restrictions.
- Two sets of restrictions to fix A (Bai 2009)
  (i) Control for **scale**: $F'F/N = I_R$. (#restrictions=$\frac{R(R+1)}{2}$)
  (ii) Control for **orthogonality**: $\Lambda'\Lambda$ =diagonal. (#restrictions=$\frac{R(R-1)}{2}$)

  $\rightarrow$ As total, they give R^2 restrictions.
  $\rightarrow$ Under these restrictions, the only admissible A are signed permutations.
  $\rightarrow$ Thus we can uniquely estimate (F, Λ) with respect to fixing A (i.e., unique up to signed permutation).

Multiplicative Form and Poisson Regression Specification

Back

Observe

$$\begin{aligned} Y &= \tilde{X}_1^{\alpha_1}\tilde{X}_2^{\alpha_2} \quad \text{where } \tilde{X}_1, \tilde{X}_2 \geq 0 \\ &= \exp(\ln(\tilde{X}_1^{\alpha_1}\tilde{X}_2^{\alpha_2})) \\ &= \exp(\alpha_1 \underbrace{\ln \tilde{X}_1}_{=:X_1} + \alpha_2 \underbrace{\ln \tilde{X}_2}_{=:X_2}) \\ &=: \exp(\alpha_1 X_1 + \alpha_2 X_2), \quad \text{where } X_1, X_2 \in \mathbb{R}. \end{aligned}$$

Thus, $Y = \exp(X'\alpha)$, where $X = (X_1, X_2)'$ and $\alpha = (\alpha_1, \alpha_2)'$.
Recall the Poisson regression model's specification is $\mathrm{E}Y|X = \exp(X'\theta)$.

- Outcomes don't need to be discrete; any non-negative values work.
  - Why? PPML only uses the conditional mean specification. The PPML estimator is consistent as long as

$$\mathrm{E}(y|x) = \exp(x'\beta),$$

    where $y := \exp(x'\beta)\xi \geq 0$ is allowed to be continuous with $\xi \geq 0$ and $\mathrm{E}(\xi|x) = 1$.

- Free of incidental parameters problem thanks to the objective functional form of PPML.
  - Why? In PPML, additive fixed effects in the linear predictor become multiplicative in the conditional mean, so the mean factorizes:

$$\mu_{ij} = \exp(x_{ij}'\beta + \alpha_i + \eta_j) = \exp(x_{ij}'\beta)\exp(\alpha_i)\exp(\eta_j).$$

  - Accordingly, the row and column sums—$\sum_j y_{ij}$ for each i and $\sum_i y_{ij}$ for each j—are jointly minimally sufficient for the FEs. After profiling, the β score contains no FEs, avoiding incidental parameter bias for β. ‣

Contraction mapping and unique identification ▸ Back

- **Contraction mapping:** unique solution
  Suppose we have an operator $T : X \to X$, and we want to find x^* such that $x^* = T(x^*)$. If T is a contraction on a complete metric space, i.e.,
  $$d(T(x), T(y)) \leq cd(x, y), \quad c < 1,$$
  then by the Banach Fixed Point Theorem:
  - A unique fixed point x^* exists.
  - Iterating $x_{n+1} = T(x_n)$ converges to x^*.

  So contraction implies a unique solution to the functional equation.

- **Identification:** unique parameter consistent with data
  In econometrics, identification usually refers to the mapping from parameters θ to data features (moments, likelihood, etc): $\mathrm{E}[m(Z, \theta_0)] = 0$.
  The parameter θ_0 is identified if
  $$\mathrm{E}[m(Z, \theta)] = 0 \quad \text{implies} \quad \theta = \theta_0.$$
  That is, only the true parameter θ_0 satisfies the model's restrictions.

Note. If the operator that maps parameters to $\underbrace{\text{conditional expectations (or moments)}}_{\mathrm{E}Y|X=g(X,\theta)\Leftrightarrow\mathrm{E}[m(Y,X,\theta)]=0}$ is a contraction in θ, then the true parameter θ_0 is uniquely identified.

Contraction mapping and unique identification ▸ Back

Example. Consider a dgp: $Y = X\beta_0 + \varepsilon, \quad \mathrm{E}\varepsilon|X = 0.$

- Think the OLS estimating equation as a fixed-point problem: OLS solves $\mathrm{E}X'(Y - X\beta) = 0$.
- Define an operator that updates β: $T(\beta) = \beta + \lambda\mathrm{E}[X'(Y - X\beta)]$ for some small positive constant λ (like a learning rate).
  - Think of this as a one-step update toward the true coefficient using the population moment. Then,

$$
\begin{aligned}
T(\beta) &= \beta + \lambda\mathrm{E}[X'(Y - X\beta)] \\
&= \beta + \lambda\mathrm{E}[X'(X\beta_0 + \varepsilon - X\beta)] \\
&= \beta + \lambda\mathrm{E}(X'X)(\beta_0 - \beta) \\
&= (I - \lambda(\mathrm{E}X'X))\beta + \lambda\mathrm{E}(X'X)\beta_0.
\end{aligned}
$$

$$
\begin{aligned}
&\Rightarrow T(\beta_1) - T(\beta_2) = (I - \lambda(\mathrm{E}X'X))(\beta_1 - \beta_2) \\
&\Rightarrow ||T(\beta_1) - T(\beta_2)|| \leq ||I - \lambda\mathrm{E}X'X||||\beta_1 - \beta_2||.
\end{aligned}
$$

- If we choose λ small enough so that all eigenvalues of $I - \lambda\mathrm{E}X'X$ lie in $(0, 1)$, then $c = ||I - \lambda\mathrm{E}X'X|| < 1$ and $T(\beta)$ is a contraction mapping on the parameter space. Note. $X'X$ having full rank is the sufficient condition of the existence of such λ.

Contraction mapping and unique identification ▸ Back

- Then the Banach fixed-point theorem says there is unique β^* such that

$$\beta^* = T(\beta^*).$$

Solve for it: $\beta^* = \beta^* = \lambda \mathrm{E} X'(Y - X\beta^*) \Rightarrow \mathrm{E}X'(Y - X\beta^*) = 0$.
That solution, β^*, is the true parameter β_0 (i.e., β^* is uniquely identified).
Because T is a contraction in β, the fixed point is unique – meaning there's only one parameter value that satisfies the model's moment condition. That's the definition of identification.

- **A sufficient condition to certify a contraction**: **"The row sum of the Jacobian is < 1"** is a sufficient way to certify a contraction. More precisely,
  - If you measure distances with the ∞-norm (max absolute component), then a mapping T is a contraction whenver $||J_T(\theta)||_\infty < 1$ uniformly on the domain, where $||A||_\infty := \max_i \sum_j |a_{ij}|$ (the maximum absolute row sum).
  - This guarantees $||T(\theta_1) - T(\theta_2)||_\infty \leq c||\theta_1 - \theta_2||_\infty$ with $c = ||J_T||_\infty < 1$.

Eigenvector basis of W Back

- Consider a symmetric matrix $\widetilde{W} = (\tilde{w}_{ij})$.
- Let W be a row-normalized matrix of $\widetilde{W}$, i.e.,
    - $W = \text{Diag}^{\text{sum}}(\widetilde{W})^{-1}\widetilde{W}$, with $\text{Diag}^{\text{sum}}(\widetilde{W}) = \text{diag}(\sum_{j=1}^{n} \tilde{w}_{ij}, \ldots, \sum_{j=1}^{n} \tilde{w}_{nj})$,
- Let $\widetilde{\widetilde{W}}$ be a symmetrically normalized matrix, i.e.,
    - $\widetilde{\widetilde{W}} := \text{Diag}^{\text{sum}}(\widetilde{W})^{-1/2}\widetilde{W}\text{Diag}^{\text{sum}}(\widetilde{W})^{-1/2}$.

Eigenvector basis of W ▸ Back

- Consider a symmetric matrix $\widetilde{W} = (\tilde{w}_{ij})$.
- Let W be a row-normalized matrix of $\widetilde{W}$, i.e.,
  - $W = \text{Diag}^{\text{sum}}(\widetilde{W})^{-1}\widetilde{W}$, with $\text{Diag}^{\text{sum}}(\widetilde{W}) = \text{diag}(\sum_{j=1}^{n} \tilde{w}_{ij}, \ldots, \sum_{j=1}^{n} \tilde{w}_{nj})$,
- Let $\widetilde{\widetilde{W}}$ be a symmetrically normalized matrix, i.e.,
  - $\widetilde{\widetilde{W}} := \text{Diag}^{\text{sum}}(\widetilde{W})^{-1/2}\widetilde{W}\text{Diag}^{\text{sum}}(\widetilde{W})^{-1/2}$.
- Observe that W and $\widetilde{\widetilde{W}}$ are similar because

$$\begin{aligned}
\widetilde{\widetilde{W}} :&= \text{Diag}^{\text{sum}}(\widetilde{W})^{-1/2}\widetilde{W}\text{Diag}^{\text{sum}}(\widetilde{W})^{-1/2} \\
&= \text{Diag}^{\text{sum}}(\widetilde{W})^{-1/2}(\text{Diag}^{\text{sum}}(\widetilde{W})W)\text{Diag}^{\text{sum}}(\widetilde{W})^{-1/2} \\
&= \text{Diag}^{\text{sum}}(\widetilde{W})^{1/2}W\text{Diag}^{\text{sum}}(\widetilde{W})^{-1/2}.
\end{aligned}$$

  - This implies $W = \text{Diag}^{\text{sum}}(\widetilde{W})^{-1/2}\widetilde{\widetilde{W}}\text{Diag}^{\text{sum}}(\widetilde{W})^{1/2}$.

Eigenvector basis of W Back

- Since $\widetilde{\widetilde{W}}$ is symmetric, by the spectral theorem, $\widetilde{\widetilde{W}} = \widetilde{Q} D \widetilde{Q}'$,
  - $\widetilde{Q}$ is orthogonal,
  - $D = \mathrm{diag}(\varphi_1, \ldots, \varphi_n)$ since W and $\widetilde{\widetilde{W}}$ are similar.

Eigenvector basis of W ‣ Back

- Since $\widetilde{\widetilde{W}}$ is symmetric, by the spectral theorem, $\widetilde{\widetilde{W}} = \widetilde{Q} D \widetilde{Q}'$,
  - $\widetilde{Q}$ is orthogonal,
  - $D = \text{diag}(\varphi_1, \ldots, \varphi_n)$ since W and $\widetilde{\widetilde{W}}$ are similar.
- Thus, W is diagonalizable because

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

  - i.e., W is diagonalizable and thus, its eigenvectors form basis.
  - Q is the eigenvector basis of W.

If $A \in \mathbb{R}^{n \times n}$ is symemtric, then $A = Q \Lambda Q'$,

- Λ: diagonals are the eigenvalues
- Q: columns are the corresponding orthonormal eigenvectors.

When do eigenvectors form a basis? Back

Eigenvectors form a basis if and only if the matrix A is diagonalizable, i.e., $A = Q\Lambda Q^{-1}$,

- Q is the matrix of eigenvectors
- Λ is the diagonal matrix of eigenvalues.

Then Q's columns (the eigenvectors) span the whole space – they are a basis.

Stage 2: Given π_{ij} from **Stage 1**, the optimal trade flows are determined in the second stage as Anderson and van Wincoop (2003).

Stage 2.1: Demand Function

- A representative consumer in country i chooses $\{c_{i1}, \dots, c_{in}\}$ by solving

$$\max_{\{c_{ij}\}_{j=1}^{n}} U_i = \left(\sum_{j=1}^{n} \chi_j^{\frac{1}{\rho}} \cdot c_{ij}^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}} \quad \text{s.t.} \ \sum_{j=1}^{n} p_{ij} c_{ij} = G_i,$$

  - χ_j denotes a preference parameter for country j's good,
  - p_{ij} is the price of country i of consuming one unit from country j,
  - $\rho > 1$ is the elasticity of substitution between all goods,
  - and G_i is the GDP of country i, with $G_1, \dots, G_n$ exogenously given.

Stage 2: Given π_{ij} from **Stage 1**, the optimal trade flows are determined in the second stage as Anderson and van Wincoop (2003).

Stage 2.1: Demand Function

- A representative consumer in country i chooses $\{c_{i1}, \ldots, c_{in}\}$ by solving

$$\max_{\{c_{ij}\}_{j=1}^{n}} U_i = \left(\sum_{j=1}^{n} \chi_j^{\frac{1}{\rho}} \cdot c_{ij}^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}} \quad \text{s.t.} \sum_{j=1}^{n} p_{ij} c_{ij} = G_i,$$

  - χ_j denotes a preference parameter for country j's good,
  - p_{ij} is the price of country i of consuming one unit from country j,
  - $\rho > 1$ is the elasticity of substitution between all goods,
  - and G_i is the GDP of country i, with $G_1, \ldots, G_n$ exogenously given.

- Solving the Lagrangian, $c_{ij}^* = \frac{C_i^{\frac{\rho}{\rho-1}} \chi_j}{(\lambda_i p_{ij})^\rho}$, where
  - $C_i = \sum_{j=1}^{n} \chi_j^{\frac{1}{\rho}} \cdot c_{ij}^{\frac{\rho-1}{\rho}}$
  - and λ_i is the Lagrange multiplier for $i = 1, \ldots, n$.

- Applying it to the budget constraint,

$$G_i = \sum_{j=1}^{n} p_{ij} c_{ij}^* = C_i^{\frac{\rho}{\rho-1}} \lambda_i^{-\rho} \sum_{j=1}^{n} \chi_j p_{ij}^{1-\rho} = C_i^{\frac{\rho}{\rho-1}} \lambda_i^{-\rho} P_i^{1-\rho},$$

  - P_i is the CES price index, $P_i = \left(\sum_{j=1}^{n} \chi_j p_{ij}^{1-\rho} \right)^{\frac{1}{1-\rho}}$.
  - P_i solves the cost minimization problem:
    $\min_{\{c_{ij}\}_{j=1}^{n}} \sum_{j=1}^{n} p_{ij} c_{ij} \quad \text{s.t.} \left(\sum_{j=1}^{n} \chi_j^{\frac{1}{\rho}} \cdot c_{ij}^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}} = \bar{U}_i$ for some $\bar{U}_i$.

- Applying it to the budget constraint,

$$G_i = \sum_{j=1}^{n} p_{ij} c_{ij}^* = C_i^{\frac{\rho}{\rho-1}} \lambda_i^{-\rho} \sum_{j=1}^{n} \chi_j p_{ij}^{1-\rho} = C_i^{\frac{\rho}{\rho-1}} \lambda_i^{-\rho} P_i^{1-\rho},$$

  - P_i is the CES price index, $P_i = \left(\sum_{j=1}^{n} \chi_j p_{ij}^{1-\rho}\right)^{\frac{1}{1-\rho}}$.
  - P_i solves the cost minimization problem:
    $\min_{\{c_{ij}\}_{j=1}^{n}} \sum_{j=1}^{n} p_{ij} c_{ij} \quad \text{s.t.} \left(\sum_{j=1}^{n} \chi_j^{\frac{1}{\rho}} \cdot c_{ij}^{\frac{\rho-1}{\rho}}\right)^{\frac{\rho}{\rho-1}} = \bar{U}_i$ for some $\bar{U}_i$.

- Thus, $\lambda_i = C_i^{\frac{1}{\rho-1}} P_i^{\frac{1-\rho}{\rho}} G_i^{-\frac{1}{\rho}}$ and the demand function is

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

Stage 2.2: Market Clearing

- The existence of trade costs lead to heterogeneous prices. We assume

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

where p_j is the exporter's supply price and exogenously given.

Stage 2.2: Market Clearing

- The existence of trade costs lead to heterogeneous prices. We assume

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

where p_j is the exporter's supply price and exogenously given.

- The nominal value of exports from country j to country i is

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

Stage 2.2: Market Clearing

- The existence of trade costs lead to heterogeneous prices. We assume

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

  where p_j is the exporter's supply price and exogenously given.

- The nominal value of exports from country j to country i is

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

- The market clearing condition imposes

$$G_j = \sum_{i=1}^{n} \mu_{ij}^* = \chi_j p_j^{1-\rho} \sum_{i=1}^{n} \pi_{ij}^{1-\rho} P_i^{-(1-\rho)} G_i,$$

  which, with normalization of $p_1 = p_2 = \cdots = p_n = 1$, leads to

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

  where $G^W := \sum_{k=1}^n G_k$ represents the world GDP.

- Let $\Pi_j := \left(\sum_{i=1}^{n} \frac{G_k}{G^W} \left(\frac{\pi_{ij}}{P_i}\right)^{1-\rho}\right)^{\frac{1}{1-\rho}}$ represents the overall trade cost from j that shows how exporter j faces trade barriers across all potential export destinations. Then,

$$\chi_j = \frac{G_j}{G^W} \Pi_j^{-(1-\rho)}$$

- Let $\Pi_j := \left(\sum_{i=1}^{n} \frac{G_k}{G^W}\left(\frac{\pi_{ij}}{P_i}\right)^{1-\rho}\right)^{\frac{1}{1-\rho}}$ represents the overall trade cost from j that shows how exporter j faces trade barriers across all potential export destinations. Then,

$$\chi_j = \frac{G_j}{G^W}\Pi_j^{-(1-\rho)}$$

- Then, the CES price index $P_i = (\sum_{j=1}^{n} \chi_j p_{ij}^{1-\rho})^{\frac{1}{1-\rho}} = \left(\sum_{j=1}^{n} \frac{G_j}{G^W} \cdot \left(\frac{\pi_{ij}}{\Pi_j}\right)^{1-\rho}\right)^{\frac{1}{1-\rho}}$ represents the overall trade cost to i, capturing how importer i experiences trade barriers across all possible foreign suppliers.

- Let $\Pi_j := \left(\sum_{i=1}^{n} \frac{G_k}{G^W} \left(\frac{\pi_{ij}}{P_i} \right)^{1-\rho} \right)^{\frac{1}{1-\rho}}$ represents the overall trade cost from j that shows how exporter j faces trade barriers across all potential export destinations. Then,

$$\chi_j = \frac{G_j}{G^W} \Pi_j^{-(1-\rho)}$$

- Then, the CES price index $P_i = (\sum_{j=1}^{n} \chi_j p_{ij}^{1-\rho})^{\frac{1}{1-\rho}} = \left(\sum_{j=1}^{n} \frac{G_j}{G^W} \cdot \left(\frac{\pi_{ij}}{\Pi_j} \right)^{1-\rho} \right)^{\frac{1}{1-\rho}}$ represents the overall trade cost to i, capturing how importer i experiences trade barriers across all possible foreign suppliers.

- We refer Π_j and P_i to *multilateral resistance* terms.
  - Trade barriers across all trade partners.
  - Note. Since $\pi_{ij} = \pi_{ij}(\boldsymbol{\mu})$, our multilateral resistance terms are implicit functions of $\boldsymbol{\mu}$.

- Hence, $\mu^*_{ij} = p_{ij} c^*_{ij} = \chi_j \pi_{ij}^{1-\rho} P_i^{-(1-\rho)} G_i$ becomes

$$\mu^*_{ij} = \frac{G_i G_j}{G^W} \left(\frac{\pi_{ij}}{\Pi_j P_i} \right)^{1-\rho}.$$

- Hence, $\mu_{ij}^* = p_{ij} c_{ij}^* = \chi_j \pi_{ij}^{1-\rho} P_i^{-(1-\rho)} G_i$ becomes

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

- Plugging π_{ij} from Stage 1,

$$\mu_{ij}^* = \underbrace{\mu_{ij}^{*\text{recursive}}}_{\substack{\text{explicitly}\\ \text{endogenous}}} \cdot \underbrace{P_i^{\rho-1}(\boldsymbol{\mu}^*) \Pi_j^{\rho-1}(\boldsymbol{\mu}^*)}_{\substack{\text{implicitly}\\ \text{endogenous}}} \cdot \underbrace{G_i G_j (G^W)^{-1} \mu_{ij}^{*\text{exo}}}_{\substack{\text{purely}\\ \text{exogenous}}},$$

  - $\mu_{ij}^{\text{recursive}}$: a discounting factor for the trade cost s.t.

$$\mu_{ij}^{\text{recursive}} = \left(\prod_{k=1}^{n} \mu_{kj}^{w_{ik}} \right)^{\lambda_d} \left(\prod_{l=1}^{n} \mu_{il}^{w_{jl}} \right)^{\lambda_o} \left(\prod_{k,l=1}^{n} \mu_{kl}^{w_{ik} w_{jl}} \right)^{\lambda_w},$$

    where $\lambda_d = (\varrho - 1)\tilde{\lambda}_d$, $\lambda_o = (\varrho - 1)\tilde{\lambda}_o$, $\lambda_w = (\varrho - 1)\tilde{\lambda}_w$.
  - $P_i^{\rho-1}(\boldsymbol{\mu})$, $\Pi_j^{\rho-1}(\boldsymbol{\mu})$: *Network* multilateral resistance terms
  - $\mu_{ij}^{\text{exo}} = \mu_{ij}^B \cdot \mu_i^D \cdot \mu_j^O$: exogenous characteristic components of the bilateral (B), origin-specific (O), and destination-specific (D), respectively.

- In our framework, the fixed effects components have their own structures:

$$\begin{cases} \tilde{\alpha}_j(\boldsymbol{\mu}) & = (G^W)^{-1/2} G_j \Pi_j^{\rho-1}(\boldsymbol{\mu})\mu_j^O \quad \text{for } j = 1, \dots, n, \\ \tilde{\eta}_i(\boldsymbol{\mu}) & = (G^W)^{-1/2} G_i P_i^{\rho-1}(\boldsymbol{\mu})\mu_i^D \quad \text{for } i = 1, \dots, n. \end{cases}$$

- In our framework, the fixed effects components have their own structures:

$$\begin{cases} \tilde{\alpha}_j(\boldsymbol{\mu}) & = (G^W)^{-1/2} G_j \Pi_j^{\rho-1}(\boldsymbol{\mu}) \mu_j^O \quad \text{for } j = 1, \dots, n, \\ \tilde{\eta}_i(\boldsymbol{\mu}) & = (G^W)^{-1/2} G_i P_i^{\rho-1}(\boldsymbol{\mu}) \mu_i^D \quad \text{for } i = 1, \dots, n. \end{cases}$$

- Semi-reduced form: Under some regularity conditions, there is a unique $\boldsymbol{\mu}^*$ s.t.

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

  - $x_{kl} = (\ln(D_{kl,1}), \cdots, \ln(D_{kl,K}))'$ and $\beta = (\beta_1, \cdots, \beta_K)'$.
  - $s_{ij,kl} \equiv ((j-1)n + i, (l-1)n + k)$-element of $\mathbf{S}^{-1}$.
    - ★ $\mathbf{S}^{-1}$: Network multiplier matrix, where $\mathbf{S}^{-1} = (I_N - (\underbrace{\lambda_d(I_n \otimes W)}_{\text{cross-destination linkage}} + \underbrace{\lambda_o(W \otimes I_n)}_{\text{cross-origin linkage}} + \underbrace{\lambda_w(W \otimes W)}_{\text{flows among third-party units}}))^{-1}$.
  - $\alpha_l(\boldsymbol{\mu}^*)$ and $\eta_k(\boldsymbol{\mu}^*)$: *Network* fixed effects
    - ★ Includes the multilateral resistance terms

Why $\mathbf{S}^{-1}\mathbf{1}_N$ homogeneous? ▸ Back

Recall

$$\mathbf{S}^{-1} = (\mathbf{I} - \mathbf{A})^{-1}, \quad \text{where } \mathbf{A} = \lambda_d(I \otimes W) + \lambda_o(W \otimes I) + \lambda_w(W \otimes W).$$

Observe that

$$\mathbf{A}(\underbrace{1_n \otimes 1_n}_{=\mathbf{1}_N}) = \cdots = (\underbrace{\lambda_d + \lambda_o + \lambda_w}_{=:\rho})(\underbrace{1_n \otimes 1_n}_{=\mathbf{1}_N}),$$

where the dots utilizes $W1_n = 1_n$ due to the row-normalization. That is, $\mathbf{1}_N$ is an eigenvector of $\mathbf{A}$ with eigenvalue ρ.
Thus,

$$(\mathbf{I} - \mathbf{A})\mathbf{1}_N = (\mathbf{I} - \rho)\mathbf{1}_N$$
$$\Rightarrow \mathbf{S}^{-1}\mathbf{1}_N = \frac{1}{1-\rho}\mathbf{1}_N,$$

i.e., proportional to a homogeneous vector of ones.

Characteristics of W ‣ Back

- Countries' connectivity matrix: Average total trade flows over the recent past years
  - $w_{ij}^{\text{Phase}} = \frac{\widetilde{w}_{ij}^{\textbf{Phase}}}{\sum_{k=1}^{n} \widetilde{w}_{ik}^{\textbf{Phase}}}$, where $\widetilde{w}_{ij}^{\text{Phase}} = \frac{1}{\#(\mathcal{T}^{\textbf{Phase}})} \sum_{t \in \mathcal{T}^{\textbf{Phase}}} \left(y_{ij,t} + y_{ji,t} \right)$
  - e.g., $\mathcal{T}^{\text{Phase}=2} = \{1987, 1988, \cdots, 1996\}$
  - W is undirected and row-normalized, with its diagonal elements being zero.

Characteristics of W ‣ Back

- Countries' connectivity matrix: Average total trade flows over the recent past years
  - $w_{ij}^{\textbf{Phase}} = \frac{\widetilde{w}_{ij}^{\textbf{Phase}}}{\sum_{k=1}^{n} \widetilde{w}_{ik}^{\textbf{Phase}}}$, where $\widetilde{w}_{ij}^{\text{Phase}} = \frac{1}{\#(\mathcal{T}^{\textbf{Phase}})} \sum_{t \in \mathcal{T}^{\textbf{Phase}}} (y_{ij,t} + y_{ji,t})$
  - e.g., $\mathcal{T}^{\text{Phase}=2} = \{1987, 1988, \cdots, 1996\}$
  - W is undirected and row-normalized, with its diagonal elements being zero.

- This implies $\varphi_{\text{max}} := \max\{\varphi_1, \ldots, \varphi_n\} = 1$ and $\varphi_{\text{min}} := \min\{\varphi_1, \ldots, \varphi_n\} < 0$.
  - $W\mathbf{1} = \mathbf{1}$. Thus, $\rho_{\text{spec}}(W) \geq 1$. Also, $\rho_{\text{spec}}(W) \leq ||W||_\infty = 1$.
  - $\text{tr}(W) = \sum_{i=1}^{n} \varphi_i = 0$.

Characteristics of W ▸ Back

- Countries' connectivity matrix: Average total trade flows over the recent past years
  - $w_{ij}^{\textbf{Phase}} = \frac{\widetilde{w}_{ij}^{\textbf{Phase}}}{\sum_{k=1}^{n} \widetilde{w}_{ik}^{\textbf{Phase}}}$, where $\widetilde{w}_{ij}^{\textbf{Phase}} = \frac{1}{\#(\mathcal{T}^{\textbf{Phase}})} \sum_{t \in \mathcal{T}^{\textbf{Phase}}} \left(y_{ij,t} + y_{ji,t} \right)$
  - e.g., $\mathcal{T}^{\text{Phase}=2} = \{1987, 1988, \cdots, 1996\}$
  - W is undirected and row-normalized, with its diagonal elements being zero.

- This implies $\varphi_{\text{max}} := \max\{\varphi_1, \ldots, \varphi_n\} = 1$ and $\varphi_{\text{min}} := \min\{\varphi_1, \ldots, \varphi_n\} < 0$.
  - $W\mathbf{1} = \mathbf{1}$. Thus, $\rho_{\text{spec}}(W) \geq 1$. Also, $\rho_{\text{spec}}(W) \leq ||W||_{\infty} = 1$.
  - $\text{tr}(W) = \sum_{i=1}^{n} \varphi_i = 0$.

- The averaging rate of W is governed by $(|\varphi_{(2)}|, |\varphi_{\text{min}}|)$.
  - Bipartite: $\varphi_{(2)} \simeq 0$, $\varphi_{\text{min}} = -1$.
    - ★ Polarized patterns
  - Linear-in-means: $\varphi_{(2)} = \varphi_{\text{min}} = -\frac{1}{n}$, which go to zero with large n.
    - ★ High leveraging rate (i.e., W averages out heterogeneity).

- The admissible parameter space varies by $|\varphi_{\text{min}}|$.

Characteristics of W ▸ Back

Note that $\varphi_{\max} = 1$ and $-1 \leq \varphi_i \leq 1$.

- Linear-in-mean

$$W = \frac{1}{n-1} \begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 0 \end{bmatrix}$$

  - $\varphi_2 = \varphi_{\min} \simeq 0$

Characteristics of W ▸ Back

Note that $\varphi_{\max} = 1$ and $-1 \leq \varphi_i \leq 1$.

- Linear-in-mean

$$W = \frac{1}{n-1}\begin{bmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 0 \end{bmatrix}$$

  - $\varphi_2 = \varphi_{\min} \simeq 0$

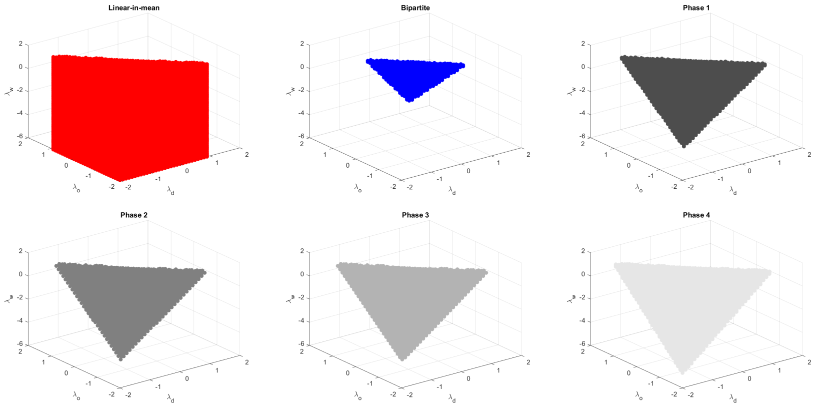
- Bipartite

$$W = \begin{bmatrix} 0 & \frac{1}{n_2 - 1} l_{n_1} l'_{n_2} \\ \frac{1}{n_1 - 1} l_{n_2} l'_{n_1} & 0 \end{bmatrix}$$

  - $\varphi_2 \simeq 0$ and $\varphi_{\min} = -1$

Characteristics of W ▸ Back

- Admissible parameter space



Other Network Statistics ▸ Back

A. Frobenius-norm-based distances ($\frac{1}{\sqrt{n(n-1)}}\|W^{\text{Phase}} - W^{\text{Phase}'}\|_F$ in the parentheses)

	$W^{\text{Phase}=1}$	$W^{\text{Phase}=2}$	$W^{\text{Phase}=3}$	$W^{\text{Phase}=4}$
$W^{\text{Phase}=2}$	2.3728 (0.0159)	0	*	*
$W^{\text{Phase}=3}$	2.8813 (0.0193)	1.9212 (0.0129)	0	*
$W^{\text{Phase}=4}$	3.6703 (0.0246)	2.8833 (0.0193)	1.9474 (0.0130)	0

B. Jaccard coefficients $J_{\text{Phase},\text{Phase}'} = \frac{\#(\mathcal{E}_{\textbf{Phase}} \cap \mathcal{E}_{\textbf{Phase}'})}{\#(\mathcal{E}_{\textbf{Phase}} \cup \mathcal{E}_{\textbf{Phase}'})}$

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

Other Network Statistics ▸ Back

A. Frobenius-norm-based distances ($\frac{1}{\sqrt{n(n-1)}}\|W^{\text{Phase}} - W^{\text{Phase}'}\|_F$ in the parentheses)

	$W^{\text{Phase}=1}$	$W^{\text{Phase}=2}$	$W^{\text{Phase}=3}$	$W^{\text{Phase}=4}$
$W^{\text{Phase}=2}$	2.3728 (0.0159)	0	*	*
$W^{\text{Phase}=3}$	2.8813 (0.0193)	1.9212 (0.0129)	0	*
$W^{\text{Phase}=4}$	3.6703 (0.0246)	2.8833 (0.0193)	1.9474 (0.0130)	0

B. Jaccard coefficients $J_{\text{Phase},\text{Phase}'} = \frac{\#(\mathcal{E}_{\textbf{Phase}} \cap \mathcal{E}_{\textbf{Phase}'})}{\#(\mathcal{E}_{\textbf{Phase}} \cup \mathcal{E}_{\textbf{Phase}'})}$

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

- Implications
  - Initial adjustment: Phase 1 $\mapsto$ Phase 2 (both the topology & the intensity of connections)
  - After Phase 2, trading relationships stablizes. Most of the subsequent adjustment occurs through re-weighting existing links.