From Shipments to Supply Chains: Mining Input-Output Links from Firm-level Trade Flows

Authors: Karsten Mau, Antonio Vicencio, Mingzhi Xu, Yawen Zheng

Discussant: Fabrizio Leone

5th Banca d'Italia and European Central Bank Conference on
"Trade, value chains and financial linkages in the global economy"

October 29-30, 2026

Summary

- **Goal**: Build *product-level* I-O links from firm-level customs data
- **Why**: Trade shocks (e.g. tariffs) hit products. Sector-level I-O tables dilute this granularity
- **How**: Machine learning + economic restrictions
- **Applications**: Product-level propagation of (*i*) China's WTO access and (*ii*) Trump 1 tariffs

My discussion: methodological recap and ideas for further development

Bidirectional Frequency Association (BFA)

- BFA measures co-occurrences in customs data

$$S_{AB} = \frac{N_{AB}}{N_B} \cdot \frac{N_{AB}}{N_A} \in [0, 1]$$

$\Rightarrow$ Share of B's exporters that import A $\times$ share of A's importers that export B

- **Def^n of link**: A is an input of B iff $S_{AB} \geq$ threshold (data driven)

- Penalizes ubiquitous inputs (used by many) and rewards specific ones (used by few)

- **Caveat**: links are as good as the info contained in the customs data
  - No info on domestic sourcing/sales $\rightarrow$ biases links estimates proportionally to RCA
  - Links contaminated by carry-along trade and other forms of intermediation

Bidirectional Frequency Association (BFA)

- BFA measures co-occurrences in customs data

$$S_{AB} = \frac{N_{AB}}{N_B} \cdot \frac{N_{AB}}{N_A} \in [0, 1]$$

$\Rightarrow$ Share of B's exporters that import A $\times$ share of A's importers that export B

- **Def^n of link**: A is an input of B iff $S_{AB} \geq$ threshold (data driven)

- Penalizes ubiquitous inputs (used by many) and rewards specific ones (used by few)

- **Caveat**: links are as good as the info contained in the customs data
  - No info on domestic sourcing/sales $\rightarrow$ biases links estimates proportionally to RCA
  - Links contaminated by carry-along trade and other forms of intermediation

Empirical Implementation

- **Data**: Chinese customs data in 2004

- **Sample selection / economic restrictions**
  - A product cannot be its own input ($A \neq B$)
  - Keep only firms that import and export at least 1 product (two-way traders)
  - Focus on intermediate and capital goods (BEC classification). Discard consumption goods
  - Drop firms above the 90th pctile in terms of # of exported products (suspect intermediaries)

- **Rich validation**: AIPNET, aggregate IO tables, import-export co-movement, Panjiva GTI

Results

Table A.2: Summary statistics BFA-based mappings (HS4-level), China 2004

		Inputs per output (HS4 level)						
Method	count	avg.	sd	p5	p25	p50	p90	max
BFA baseline	1,218	47.4	50.8	5	13	25	144	197
BFA post	1,210	35.5	38.9	4	9	18	110	153
BFA pre- & post	1,168	17.7	20.8	2	5	10	68	121
		Outputs per input (HS4 level)						
Method	count	avg.	sd	p5	p25	p50	p90	max
BFA baseline	1,215	47.5	50.7	4	14	26	142	215
BFA post	935	46.0	49.3	4	13	25	135	213
BFA pre- & post	924	22.4	22.4	2	8	14	53	118

Note: Author's calculations based on Firm-level trade data from *China Customs* for the year 2004.

Potential Extensions

- **If available**: Domestic B2B data + PRODCOM for domestic transactions

- **Panel data**: check correlation between importing A at t and exporting B at $t+1$
  - Helps for lumpy purchases of capital goods

- **Merge with firm registry**
  - Use info on NACE/province to correct for sector- or province-wide sourcing
  - **Value-weighted BFA** $\rightarrow$ account for firm size (see next slide)

$$N_{AB} = \sum_f s_{fA}\, s_{fB}, \qquad N_A = \sum_f s_{fA}, \qquad N_B = \sum_f s_{fB}$$

s_{fA}: share of A in f's imports; s_{fB}: share of B in f's exports (BFA is the special case $s = 1$)

Potential Extensions

- **If available**: Domestic B2B data + PRODCOM for domestic transactions
- **Panel data**: check correlation between importing A at t and exporting B at $t+1$
  - Helps for lumpy purchases of capital goods
- **Merge with firm registry**
  - Use info on NACE/province to correct for sector- or province-wide sourcing
  - **Value-weighted BFA** $\rightarrow$ account for firm size (see next slide)

$$N_{AB} = \sum_f s_{fA}\, s_{fB}, \qquad N_A = \sum_f s_{fA}, \qquad N_B = \sum_f s_{fB}$$

s_{fA}: share of A in f's imports; s_{fB}: share of B in f's exports (BFA is the special case $s = 1$)

Potential Extensions

- **If available**: Domestic B2B data + PRODCOM for domestic transactions

- **Panel data**: check correlation between importing A at t and exporting B at $t+1$
  - Helps for lumpy purchases of capital goods

- **Merge with firm registry**
  - Use info on NACE/province to correct for sector- or province-wide sourcing
  - **Value-weighted BFA** $\rightarrow$ account for firm size (see next slide)

$$N_{AB} = \sum_f s_{fA}\, s_{fB}, \qquad N_A = \sum_f s_{fA}, \qquad N_B = \sum_f s_{fB}$$

  s_{fA}: share of A in f's imports; s_{fB}: share of B in f's exports (BFA is the special case $s = 1$)

BFA versus Weighted BFA — Italian Customs Data (2017–2023)

	Links	of which in AIPNET	Precision	Recall
BFA	11,163	3,804	34%	8%
Weighted BFA	11,163	4,592	41%	9%
Random pick among candidates			16%	

- Weighting alone adds 788 AIPNET links (+21%) at equal size
- Precision relative to random benchmark rises from 2.1 to 2.6