Trade, Migration, and Productivity: A Quantitative Analysis of China

Trevor Tombe and Xiaodong Zhu

November 27, 2023

Outline

Introduction

- Export growth has often been cited as one of the key driving forces of China's growth since 2001.
- It is also widely believed that the export growth has been supported by cheap labour from rural migrant workers.

Introduction

- Export growth has often been cited as one of the key driving forces of China's growth since 2001.
- It is also widely believed that the export growth has been supported by cheap labour from rural migrant workers.
- What equally dramatic is the growth of China's internal economic integration.
- Trade between provinces has increased more than trade between China and the rest of the world, while policy-induced migration cost has been substantially decreased. (Poncet 2006; Cai, Park, and Zhao 2008)

Introduction

- Export growth has often been cited as one of the key driving forces of China's growth since 2001.
- It is also widely believed that the export growth has been supported by cheap labour from rural migrant workers.
- What equally dramatic is the growth of China's internal economic integration.
- Trade between provinces has increased more than trade between China and the rest of the world, while policy-induced migration cost has been substantially decreased. (Poncet 2006; Cai, Park, and Zhao 2008)
- RQ: How important have internal trade, international trade and rural-urban migration been to China's growth?

Introduction

- Export growth has often been cited as one of the key driving forces of China's growth since 2001.
- It is also widely believed that the export growth has been supported by cheap labour from rural migrant workers.
- What equally dramatic is the growth of China's internal economic integration.
- Trade between provinces has increased more than trade between China and the rest of the world, while policy-induced migration cost has been substantially decreased. (Poncet 2006; Cai, Park, and Zhao 2008)
- RQ: How important have internal trade, international trade and rural-urban migration been to China's growth?

Migration

- In 1958, China formally instituted a household registration system to control population mobility.
- This prohibition was relaxed in the 1980s but, prior to 2003, workers without local hukou still had to apply for a temporary residence permit.
- After 2003, Migration cost remain high: limited access to local public services.

Migration

- In 1958, China formally instituted a household registration system to control population mobility.
- This prohibition was relaxed in the 1980s but, prior to 2003, workers without local hukou still had to apply for a temporary residence permit.
- After 2003, Migration cost remain high: limited access to local public services.
- Data: individual population census 1% sample in 2000 and 1% mini census in 2005.
- inter-province migrant: not living in original province.
- intra-province migrant: living in original province but not working in original sector.

Internal Trade

- High internal trade costs in China in the 1990s due to local government protection. (Young 2000, Poncet 2005).
- Since 2000, internal trade barriers have fallen significantly.
  - Zhu Rongji' Prohibition in 2001.
  - Various SOE reforms.
  - Improved transport infrastructure.

Internal Trade

- High internal trade costs in China in the 1990s due to local government protection. (Young 2000, Poncet 2005).
- Since 2000, internal trade barriers have fallen significantly.
  - Zhu Rongji' Prohibition in 2001.
  - Various SOE reforms.
  - Improved transport infrastructure.
- Data: input-ouput table from NBS in 2000 and 2007.

Model in Brief

- Two-sector multi-region general equilibrium model featuring internal trade, international trade, and worker migration.
- N+1 regions: N provinces within China + rest of the world, (n)
- Two sectors: Agriculture and Nonagriculture, (j)
- Endowments:
  - Initia/Hukou registrants, by home province and sector, $\bar{L}_j^n$
  - Fixed factors (structures) used for housing and production, S_j^n,
  - Fundamental productivity in each region/sector, T_j^n
- Trade:
  - Internal and international trade
  - Eaton-Kortum structure as in Redding (2015)
- Migration:
  - Worker heterogeneity (differences in productivity)
  - Some optimally choose to migrate, others don 't
  - Effective workers in each region/sector, H_j^n

Worker Preference

- The worker maximizes the Cobb-Douglas utility

$$u_n^j = \varepsilon_n^j \left[\left(C_n^{j,ag} \right)^{\psi^{ag}} \left(C_n^{j,na} \right)^{\psi^{na}} \right]^{\alpha} \left(S_n^{j,h} \right)^{1-\alpha} \tag{1}$$

- s.t.

$$P_n^{j,ag} C_n^{j,ag} + P_n^{j,na} C_n^{j,na} + r_n^j S_n^{j,h} \leq v_{in}^{kj},$$

- $C_n^{j,ag}$ and $C_n^{j,na}$ Consumption in Two sectors with price $P_n^{j,ag}$ and $P_n^{j,na}$
- $S_n^{j,h}$ housing structures with a price r_n^j.
- ε_n^j is an idiosyncratic preference shifter that is i.i.d. across workers, sectors, and regions.
- v_{in}^{kj} is the nominal income of a worker in region i and sector k, but works in region n and sector j.
- L_{in}^{kj} be the number of such workers, $L_n^j = \sum_{k \in \{ag,na\}} \sum_{i=1}^N L_{in}^{kj}$ is the total number of workers.
- $v_n^j = \sum_{k \in \{ag,na\}} \sum_{i=1}^N v_{in}^{kj} L_{in}^{kj} / L_n^j$ is the average income in region n, sector j

Worker Preference

- The worker maximizes the Cobb-Douglas utility

$$u_n^j = \varepsilon_n^j \left[\left(C_n^{j,ag} \right)^{\psi^{ag}} \left(C_n^{j,na} \right)^{\psi^{na}} \right]^{\alpha} \left(S_n^{j,h} \right)^{1-\alpha} \tag{2}$$

- s.t.

$$P_n^{j,ag} C_n^{j,ag} + P_n^{j,na} C_n^{j,na} + r_n^j S_n^{j,h} \leq v_{in}^{kj},$$

- Final demand for good j by workers in region n is

$$D_n^j = \alpha \psi^j \sum_{k \in \{ag,na\}} v_n^k L_n^k$$

Similarly, final demand for housing there is $(1-\alpha) \sum_{k \in \{ag,na\}} v_n^k L_n^k$.

Production, Trade, and Goods Prices

- Agricultural and non-agricultural goods are composites of a continuum of horizontally differentiated varieties $y_n^j(\nu)$, where $j = \{ag, na\}$ and $v \in [0,1]$.
- A perfectly competitive firm produces good j using the (constant elasticitity of substitution) CES technology

$$Y_n^j = \left(\int_0^1 y_n^j(\nu)^{(\sigma-1)/\sigma} d\nu \right)^{\sigma/(\sigma-1)}$$

- Each variety may be sourced from local producers or imported, whichever minimizes costs.

$$c_n^j(\varphi) \propto \frac{1}{\varphi} \left[\left(w_n^j\right)^{\beta^j} \left(r_n^j\right) \eta^{\eta^j} \left(\prod_{k \in \{ag,na\}} \left(P_n^k\right)^{\sigma^{jk}} \right) \right],$$

- β^j and η^j are labor and land shares, and σ^{jk} share for intermediate input
- w_n^j is the wage, r_n^j the rental cost of land, and P_n^k the price of intermediate input from sector k

Production, Trade, and Goods Prices

- Consumer choose the cheapest price: $\tau_{ni}^{j} c_{i}^{j}(\varphi)$. As in Eaton and Kortum (2002), they assume that φ is distributed Fréchet with cumulative distribution function (CDF) $F_{n}^{j}(\varphi) = e^{-T_{n}^{j}\varphi^{-\theta}}$. Then equilibrium trade shares are

$$\pi_{ni}^{j} = \frac{T_{i}^{j}\left(\tau_{ni}^{j} c_{i}^{j}\right)^{-\theta}}{\sum_{m=1}^{N+1} T_{m}^{j}\left(\tau_{nm}^{j} c_{m}^{j}\right)^{-\theta}},$$

- where π_{ni}^{j} is the fraction of region n spending allocated to sector j goods produced in region i (trade shares), and final goods prices are

$$P_{n}^{j} = \gamma\left[\sum_{i=1}^{N+1} T_{i}^{j}\left(\tau_{ni}^{j} c_{i}^{j}\right)^{-\theta}\right]^{-1/\theta},$$

where γ is a constant, and T_{i}^{j} the productivity parameter.

Production, Trade, and Goods Prices

- Let X_n^j be total expenditures on good j by region n. Total revenue is then

$$R_n^j = \sum_{i=1}^{N+1} \pi_{in}^j X_i^j$$

- Combined with demand for intermediates by producers, we have

$$X_n^j = D_n^j + \sum_k \sigma^{kj} R_n^k$$

Production, Trade, and Goods Prices

- Let X_n^j be total expenditures on good j by region n. Total revenue is then

$$R_n^j = \sum_{i=1}^{N+1} \pi_{in}^j X_i^j$$

- Combined with demand for intermediates by producers, we have

$$X_n^j = D_n^j + \sum_k \sigma^{kj} R_n^k$$

Incomes of Workers

- Consumer preferences and production technologies are Cobb-Douglas, so total spending on the fixed factor is

$$(1-\alpha)v_n^j L_n^j + \eta^j R_n^j = (1-\alpha)v_n^j L_n^j + \eta^j \beta^{j-1} w_n^j L_n^j$$

.

- Given a total fixed-factor endowment of $\bar{S}_n^j$, the market clearing condition for the fixed-factor is

$$r_n^j \bar{S}_n^j = (1-\alpha)v_n^j L_n^j + \eta^j \beta^{j-1} w_n^j L_n^j$$

.

- Add fixed-factor income to labor income we have $v_n^j L_n^j = (1-\alpha)v_n^j L_n^j + \eta^j \beta^{j-1} w_n^j L_n^j + w_n^j L_n^j$. And the total fixed-factor income in region n sector j is

$$r_n^j \bar{S}_n^j = \left[\frac{(1-\alpha)\beta^j + \eta^j}{\alpha\beta^j}\right] w_n^j L_n^j$$

Incomes of Workers

- As only workers with local hukou receive fixed-factor income, they define

$$\delta_{ni}^{jk} = \begin{cases} 1 + \left(\frac{(1-\alpha)\beta^j + \eta^j}{\alpha\beta^j}\right)\frac{L_n^j}{L_{nn}^{jj}} & \text{if } n = i \text{ and } j = k \\ 1 & \text{if } n \neq i \text{ or } j \neq k \end{cases},$$

  as the effective fixed-factor "rebate rate" to workers
- Then we can write the incomes of workers registered in region n and sector j as $v_{ni}^{jk} = \delta_{ni}^{jk} w_i^k$.

Internal Migration

- m_{ni}^{jk}: the share of workers registered in (n, j) who migrated to (i, k).
- migration cost: (1)land returns, (2)utility discount factor μ_{ni}^{jk}. (3) workers differ in their location preferences ε_{n}^{j}.
- Real wages $V_i^k \equiv w_i^k / \left(P_i^{ag\psi^{ag}} P_i^{na\psi^{na}} \right)^{\alpha} \left(r_i^k \right)^{1-\alpha}$, workers from (n, j) choose (i, k) to maximize their welfare $\varepsilon_i^k \delta_{ni}^{jk} V_i^k / \mu_{ni}^{jk}$.
- The law of large numbers implies that the proportion of workers who migrate to region (i, k) is

$$m_{ni}^{jk} = \Pr \left(\varepsilon_i^k \delta_{ni}^{jk} V_i^k / \mu_{ni}^{jk} \geq \max_{i', k'} \left\{ \varepsilon_{i'}^{k'} \delta_{ni'}^{jk'} V_{i'}^{k'} / \mu_{ni'}^{jk'} \right\} \right).$$

- Location preferences follow a Fréchet distribution with $\mathrm{CDF}\ F_{\varepsilon}(x) = e^{-x^{-k}}$, where κ governs the degree of dispersion across individuals. A large κ implies small dispersion.

Internal Migration

- m_{ni}^{jk}: the share of workers registered in (n, j) who migrated to (i, k).
- migration cost: (1)land returns, (2)utility discount factor μ_{ni}^{jk}. (3) workers differ in their location preferences ε_n^j.
- Real wages $V_i^k \equiv w_i^k / \left(P_i^{ag\psi^{ag}} P_i^{na\psi^{na}} \right)^{\alpha} \left(r_i^k \right)^{1-\alpha}$, workers from (n, j) choose (i, k) to maximize their welfare $\varepsilon_i^k \delta_{ni}^{jk} V_i^k / \mu_{ni}^{jk}$.
- $$m_{ni}^{jk} = \frac{\left(V_i^k \delta_{ni}^{jk} / \mu_{ni}^{jk} \right)^{\kappa}}{\sum_{k'} \sum_{i'=1}^{N} \left(V_{i'}^{k'} \delta_{ni'}^{jk'} / \mu_{ni'}^{jk'} \right)^{\kappa}},$$
- Total employment at (i, k) is $L_i^k = \sum_j \sum_{n=1}^{N} m_{ni}^{jk} \bar{L}_n^j$.

Solving the Model

- Given changes in migration and real incomes, the change in aggregate welfare is

$$\hat{W} = \sum_{j} \sum_{n=1}^{N} \omega_n^j \hat{V}_n^j \hat{\delta}_{nn}^{jj} \left(\hat{m}_{nn}^{jj}\right)^{-1/\kappa},$$

  where $\omega_n^j \propto \bar{L}_n^j V_n^j \delta_{nn}^{jj} \left(\hat{m}_{nn}^{ij}\right)^{-1/\kappa}$ is region n and sector j 's initial contribution to welfare.

- Similarly, the change in real GDP is

$$\hat{Y} = \sum_{j} \sum_{n=1}^{N} \phi_n^j \hat{V}_n^j \hat{L}_n^j$$

  where $\phi_n^j \propto V_n^j L_n^j$ is the contribution of region n and sector j to initial real GDP.

Calibration

Parameter	Value	Description
(β^{ag}, β^{na})	$(0.29, 0.22)$	Labor's share of output
(η^{ag}, η^{na})	$(0.28, 0.03)$	Land's share of output
$(\sigma^{ag,na}, \sigma^{na,ag})$	$(0.60, 0.06)$	Intermediate input shares
ψ^{ag}	0.095	Agriculture's share of final demand
α	0.87	Goods' expenditure share
θ	4	Elasticity of trade
κ	1.5	Elasticity of migration
π_{ni}^{j}	Data	Bilateral trade shares
m_{ni}^{j}	Data	Bilateral migration shares
$\bar{L}_{n}^{j}$	Data	Hukou registrations

Cost-Elasticity of Trade

- There is a large literature on the productivity dispersion parameter .
- For between countries, Simonovska and Waugh (2014) use cross-country price data to estimate 4. Parro (2013) estimates [4.5,5.2] for manufacturing using trade and tariff data.
- Within countries, however, there is little evidence to draw upon.
- $\theta = 4$

Cost-Elasticity of Migration

- $m_{ni}^{jk}/m_{nn}^{jj} = \left(V_i^k/\delta_{nn}^{jj}\mu_{ni}^{jk}V_n^j\right)^{\kappa}$.
- Two alternative assumptions about migration costs:
  - (1) $\mu_{ni}^{jk} = \bar{\mu}_n^j d_{ni}^o \xi_{ni}^{jk}$, where d_{ni} is the distance between province n and i;
  - (2) $\mu_{ni}^{jk} = \bar{\mu}_n^j \bar{\mu}_{ni} \xi_{ni}^{jk}$, where $\hat{\mu}_{ni}$ is symmetric with respect to n and i.
- Under these assumptions and given data on migration shares and real incomes, we estimate κ using the fixed effect regressions:

Cost-Elasticity of Migration

- $$\ln\left(\frac{m_{ni}^{jk}}{m_{nn}^{jj}}\right) = \kappa \ln\left(V_i^k\right) + \gamma_{ni} + \gamma_n^j + \varsigma_{ni}^{jk}, \quad for(n,i) \neq (i,k)$$

- $\gamma_n^j = -\kappa \ln \bar{\mu}_n^j - \kappa \ln\left(\delta_{nn}^{jj} V_n^j\right)$ is an origin province-sector fixed effect, $\gamma_{ni} = -\kappa \ln \bar{\mu}_{ni}$ is an origin-destination province-pair fixed effect, $\varsigma_{ni}^{jk} = -\kappa \ln \xi_{ni}^{jk} + \vartheta_{ni}^{jk}$, and ϑ_{ni}^{jk} is a measurement error term.
- Instrument:
  - (1) the distance weighted average income of neighboring provinces.
  - (2) Bartik-style expected income instrument.

Cost-Elasticity of Migration

metries are exporter-specific such that $\tau^j_{ni} = t^j_{ni} t^j_i$, where t^j_{ni} are symmetric
$(t^j_{ni} = t^j_{in})$ and t^j_i are costs of exporting. This and equation (15) imply an Adj
Head-Ries Index $\tau^j_{ni} = \bar{\tau}^j_{ni}\sqrt{t^j_i / t^j_n}$, as in Tombe (2015).

Migration Cost

$$\mu_{ni}^{jk} = \frac{1}{\delta_{nn}^{jj}} \left(\frac{V_i^k}{V_n^j}\right) \left(\frac{m_{nn}^{jj}}{m_{ni}^{jk}}\right)^{1/\kappa}, forn \neq i$$

	Initial share of employment	Average migration costs μ_{in}^{kj} Level in 2000	Level in 2005	Relative change
Overall	0.174	2.82	2.31	0.82
Agriculture to non-agriculture migration cost changes				
Overall	0.16	2.63	2.16	0.82
Within province	0.13	2.21	1.83	0.83
Between provinces	0.03	25.21	15.43	0.61
Between provinces migration cost changes				
Overall	0.04	24.75	15.08	0.61
Within agriculture	0.003	47.67	42.22	0.89
Within non-agriculture	0.01	21.02	12.2	0.58

Notes: Displays migration-weighted harmonic means of migration costs in 2000 and 2005. The migrant share of employment summarizes m_{ni}^{jk} in 2000. We use initial period weights to average the 2005 costs to capture only changes in costs and not migration patterns.

Migration Cost

	Initial share of employment	Average migration costs μ_{in}^{kj}		
		Level in 2000	Level in 2005	Relative change
Overall	0.174	2.82	2.31	0.82
Agriculture to non-agriculture migration cost changes				
Overall	0.16	2.63	2.16	0.82
Within province	0.13	2.21	1.83	0.83
Between provinces	0.03	25.21	15.43	0.61
Between provinces migration cost changes				
Overall	0.04	24.75	15.08	0.61
Within agriculture	0.003	47.67	42.22	0.89
Within non-agriculture	0.01	21.02	12.2	0.58

Notes: Displays migration-weighted harmonic means of migration costs in 2000 and 2005. The migrant share of employment summarizes m_{ni}^{jk} in 2000. We use initial period weights to average the 2005 costs to capture only changes in costs and not migration patterns.

- the estimated migration costs are large, but broadly consistent with evidence from individual survey data.
- location preferences and amenities vary systematically across provinces
- It's the relative change that matters.

Trade Cost

$$\bar{\tau}^j_{ni} \equiv \sqrt{\tau^j_{ni}\tau^j_{in}} = \left(\frac{\pi^j_{nn}\pi^j_{ii}}{\pi^j_{ni}\pi^j_{in}}\right)^{1/2\theta}$$

- Trade costs systematically differ depending on the direction of trade
- $\tau^j_{ni} = t^j_{ni} t^j_i$, where t^j_{ni} are symmetric costs $\left(t^j_{ni} = t^j_{in}\right)$ and t^j_i are costs of exporting.
- djusted Head-Ries Index $\tau^j_{ni} = \bar{\tau}^j_{ni}\sqrt{t^j_i / t^j_n}$, as in Tombe (2015).

Trade Cost

Importer	Exporter: Northeast	Beijing Tianjin	North Coast	Central Coast	South Coast	Central region	Northwest	Southwest	Abroad
Average trade cost levels in 2002									
Northeast		2.61	2.89	3.65	2.71	3.32	2.57	3.36	3.43
Beijing/Tianjin	2.60		1.92	3.13	2.44	3.08	2.66	3.44	2.84
North Coast	2.79	1.87		2.69	2.51	2.58	2.53	3.61	3.30
Central Coast	3.80	3.27	2.89		2.21	2.27	2.70	3.34	2.43
South Coast	3.74	3.39	3.59	2.91		3.03	3.08	2.93	2.62
Central region	3.18	2.94	2.53	2.15	2.06		2.46	3.12	4.08
Northwest	3.02	3.07	2.96	2.94	2.50	2.95		2.89	4.61
Southwest	3.10	3.20	3.47	2.95	1.96	3.08	2.38		4.25
Abroad	4.94	4.10	4.75	3.37	2.63	6.05	5.79	6.32	
Average trade cost changes from 2002 to 2007									
Northeast		0.91	0.90	0.84	0.83	0.88	0.92	0.88	0.81
Beijing/Tianjin	0.84		0.89	0.91	0.89	0.80	0.75	0.85	0.78
North Coast	0.87	0.93		1.00	0.87	0.78	0.72	0.77	0.80
Central Coast	0.76	0.88	0.92		0.88	0.82	0.74	0.85	0.81
South Coast	0.77	0.93	0.87	0.92		0.81	0.72	0.80	0.92
Central region	0.88	0.91	0.85	0.96	0.90		0.78	0.84	0.75
Northwest	0.99	0.92	0.85	0.96	0.88	0.86		0.87	0.68
Southwest	0.89	0.94	0.83	0.97	0.85	0.82	0.78		0.74
Abroad	0.88	0.93	0.92	0.98	1.05	0.77	0.64	0.79	

Notes: Displays the aggregate average trade costs in 2002 and the relative changes from 2002 to 2007. We aggregate the sectoral trade costs using expenditure weights, but use the sector-specific estimates in the quantitative analysis.

Lower Migration Cost

	Trade shares (p.p. change)		Migrant stock (%)		Real GDP per worker (%)	Aggregate welfare (%)
	Internal	External	Within province	Between province		
All	0.1	0.1	14.5	80.8	4.8	11.1
No land inputs	0.1	0.2	14.4	85.6	5.3	8.4
And no housing	0.1	0.2	13.8	90.4	6.5	7.6
And $\theta \to \infty$	−0.2	0.1	23.2	119.2	11.8	6.2
Agriculture to non-agriculture migration cost changes						
Overall	0.1	0.1	15.2	52.9	4.3	9.1
Within provinces	−0.0	−0.1	22.8	−9.7	2.0	5.9
Between provinces	0.1	0.2	−7.0	69.9	2.8	3.5
Between provinces migration cost changes						
Overall	0.2	0.3	−7.8	97.9	3.2	5.5
Within agriculture	−0.0	0.0	−0.1	2.3	−0.0	0.1
Within non-agriculture	0.1	0.1	−1.0	30.9	0.7	2.2

Notes: Displays aggregate response to various migration cost changes. All use migration cost changes as measured, though set $\hat{\mu}_{ni}^{kj} = 1$ for certain (n, i, j, k) depending on the experiment. The change in internal and external trade shares are the expenditure weighted average changes in region's $\sum_{n \neq i} \pi_{ni}^{j}$ and $\pi_{n(N=1)}^{j}$. The migrant stock is the number of workers living outside their province of registration.

Lower Trade Cost

	Trade shares (p.p. change)		Migrant stock (%)		Real GDP per worker (%)	Aggregate welfare (%)
	Internal	External	Within province	Between province		
Internal trade	9.2	−0.7	0.8	−1.8	11.2	11.4
External trade	−0.7	3.9	1.8	2.4	4.0	2.9
All trade	8.2	2.8	2.5	0.5	15.2	14.1
No Change in migration						
Internal trade	9.1	−0.7	–	–	11.2	11.2
External trade	−0.7	3.9	–	–	3.4	3.4
All trade	8.2	2.8	–	–	14.5	14.5
No intermediate inputs						
Internal trade	8.6	−0.5	0.3	−1.4	3.0	3.3
External trade	−0.7	3.9	1.5	1.6	1.1	0.3
All trade	7.6	3.2	1.6	0.1	4.1	3.5
No intermediate inputs and no change in migration						
Internal trade	8.6	−0.5	–	–	3.1	3.1
External trade	−0.7	3.9	–	–	0.6	0.6
All trade	7.6	3.2	–	–	3.7	3.7

Notes: Displays aggregate response to various trade cost changes. All use trade cost changes as measured, though set $\hat{\tau}^j_{ni} = 1$ for certain (n,i,j) depending on the experiment. The change in internal and external trade shares are the expenditure weighted average changes in region's $\sum_{n\neq i}\pi^j_{ni}$ and $\pi^j_{n(N+1)}$. The migrant stock is the number of workers living outside their province of registration.

Decomposing China's Recent Economic Growth

	Marginal effects		
	Real GDP per worker growth (%)	Share of growth	Standard deviation (%)
Overall (all changes)	57.1	–	–
Productivity changes	36.9	0.64	1.3
Internal trade cost changes	10.2	0.18	0.3
External trade cost changes	4.5	0.08	0.7
Migration cost changes	5.6	0.10	0.9
Of the migration cost changes			
Between-province, within-non-agriculture	0.9	0.02	0.4
Between-province, within-agriculture	0.0	0.00	0.0
Between-province, agriculture-non-agriculture	3.2	0.06	0.9
Within-province, agriculture-non-agriculture	1.5	0.03	0.3

Notes: Decomposes the change in real GDP into contributions from productivity, internal trade cost changes, external trade cost changes, and migration cost changes. The bottom panel decomposes the change due to migration cost changes into various different types of migration. To attribute contributions from each component, we report the marginal contribution to aggregate growth of each component across all permutations. In the last column, we report the standard deviation of those growth rates across permutations. Shares may not sum to 1 due to rounding. The growth rates are continuously compounded rates.

Logit Model

A decision maker chooses the alternative that maximizes utility

- A decision maker, n, faces a choice among J discrete alternatives
- Alternative j provides utility U_{nj} (where $j = 1, \ldots, J$)
- n chooses i if and only if $U_{ni} > U_{nj}\ \forall j \neq i$

We (the econometricians) do not observe utility U_{nj}, so we model it as being composed of

- V_{nj}: Utility from observed attributes
- ε_{nj}: Utility from unobserved attributes, which we treat as random

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

The probability that decision maker n chooses alternative i is

$$\begin{aligned} P_{ni} &= \Pr(U_{ni} > U_{nj}\ \forall j \neq i) \\ &= \int_{\varepsilon} \mathbb{1}(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj}\ \forall j \neq i) f(\boldsymbol{\varepsilon}_n) d\boldsymbol{\varepsilon}_n \end{aligned}$$

Logit Model

The logit model makes a simple (but sometimes overly strong) assumption about the joint density of unobserved utility, $f(\varepsilon_n)$

$$\varepsilon_{nj} \sim \text{i.i.d. type I extreme value (Gumbel) with } Var(\varepsilon_{nj}) = \frac{\pi^2}{6}$$

Why make this assumption about the unobserved component of utilities?

- It yields a simple closed-form expression for choice probabilities

Are there any downsides to making this assumption?

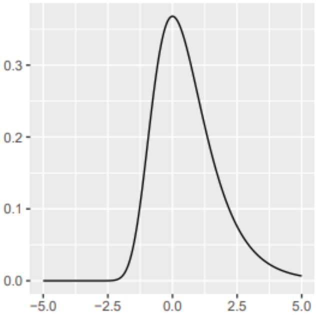
- It implies substitution patterns that may be unrealistic

Logit Model

Type I extreme value is similar to a normal distribution but with a fatter tail on one side

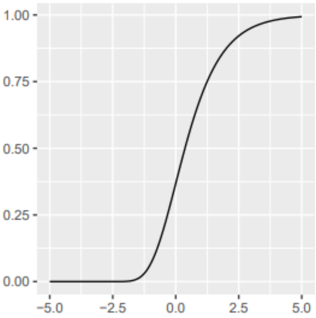
Probability density

$$f(\varepsilon_{nj}) = e^{-\varepsilon_{nj}} e^{-e^{-\varepsilon_{nj}}}$$



Cumulative distribution

$$F(\varepsilon_{nj}) = e^{-e^{-\varepsilon_{nj}}}$$

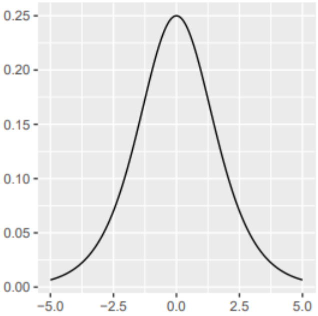


Logit Model

The difference of two type I extreme value draws, $\varepsilon^*_{nji} = \varepsilon_{nj} - \varepsilon_{ni}$, follows the logistic distribution

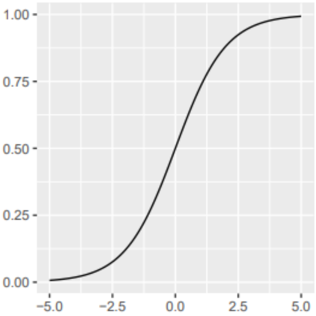
Probability density

$$f(\varepsilon^*_{nji}) = \frac{e^{\varepsilon^*_{nji}}}{\left(1 + e^{\varepsilon^*_{nji}}\right)^2}$$



Cumulative distribution

$$F(\varepsilon^*_{nji}) = \frac{e^{\varepsilon^*_{nji}}}{1 + e^{\varepsilon^*_{nji}}}$$



Logit Model

$$
\begin{aligned}
P_{ni} &= \Pr(U_{ni} > U_{nj} \; \forall j \neq i) \\
&= \Pr(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \; \forall j \neq i) \\
&= \Pr(\varepsilon_{nj} < \varepsilon_{ni} + V_{ni} - V_{nj} \; \forall j \neq i)
\end{aligned}
$$

Suppose we know V_{ni}, V_{nj}, and ε_{ni}

- We know the right-hand side of the inequality inside the probability

For a single ε_{nj}, this probability is the cumulative distribution of a type I extreme value random variable

$$\Pr(\varepsilon_{nj} < \varepsilon_{ni} + V_{ni} - V_{nj} \mid \varepsilon_{ni}) = e^{-e^{-(\varepsilon_{ni} + V_{ni} - V_{nj})}}$$

We need to know this probability $\forall j \neq i$, not just a single j

- ε_{nj} is i.i.d., so we can take the product of the probability for each ε_{nj}

$$P_{ni} \mid \varepsilon_{ni} = \prod_{j \neq i} e^{-e^{-(\varepsilon_{ni} + V_{ni} - V_{nj})}}$$

Logit Model

Conditional on knowing ε_{ni}, the choice probability of alternative i is

$$P_{ni} \mid \varepsilon_{ni} = \prod_{j \neq i} e^{-e^{-(\varepsilon_{ni}+V_{ni}-V_{nj})}}$$

But ε_{ni} is random, so we have to integrate over the density of ε_{ni}

$$\begin{aligned} P_{ni} &= \int \left(\prod_{j \neq i} e^{-e^{-(\varepsilon_{ni}+V_{ni}-V_{nj})}} \right) e^{-\varepsilon_{ni}} e^{-e^{-\varepsilon_{ni}}} d\varepsilon_{ni} \\ &= \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} \end{aligned}$$

- See the textbook for the proof of this equivalence

The probability of n choosing i is a closed-form expression that depends on the representative utility (or observable attributes) of all alternatives

Nested Logit Model

We partition the J alternatives into K nonoverlapping subsets denoted $B_1, B_2, \dots, B_K$ and called "nests"

The utility for each alternative is again $U_{nj} = V_{nj} + \varepsilon_{nj}$ where the vector of unobserved utility, $\boldsymbol{\varepsilon}_n = (\varepsilon_{n1}, \varepsilon_{n2}, \dots, \varepsilon_{nJ})$, has cumulative distribution

$$F(\boldsymbol{\varepsilon}_n) = \exp\left(-\sum_{k=1}^{K}\left(\sum_{j \in B_k} e^{-\varepsilon_{nj}/\lambda_k}\right)^{\lambda_k}\right)$$

which is a type of generalized extreme value (GEV) distribution

- The marginal distribution of ε_{nj} is extreme value
- $Cov(\varepsilon_{ni}, \varepsilon_{nm}) = 0$ if $i \in B_k$ and $m \in B_\ell$ with $k \neq \ell$
- $Cov(\varepsilon_{ni}, \varepsilon_{nm}) \geq 0$ if $i \in B_k$ and $m \in B_k$
- λ_k is a measure of independence in nest k
  - $\lambda_k = 1 \; \forall k$ gives the logit model

Mixed Logit Model

The utility that decision maker n obtains from alternative j is

$$U_{nj} = \boldsymbol{\beta}_n' \mathbf{x}_{nj} + \varepsilon_{nj}$$

But we (the researchers) do not observe $\boldsymbol{\beta}_n$ for any individual

- We model $\boldsymbol{\beta}_n$ as a random variable with density $f(\boldsymbol{\beta} \mid \boldsymbol{\theta})$

If we did know $\boldsymbol{\beta}_n$, then the model would be a standard logit with the *conditional* choice probability

$$L_{ni}(\boldsymbol{\beta}_n) = \frac{e^{\boldsymbol{\beta}_n' \mathbf{x}_{ni}}}{\sum_{j=1}^{J} e^{\boldsymbol{\beta}_n' \mathbf{x}_{nj}}}$$

But we do not know $\boldsymbol{\beta}_n$, so we have to integrate over the density of the random coefficients to obtain the *unconditional* choice probability

$$P_{ni} = \int \frac{e^{\boldsymbol{\beta}' \mathbf{x}_{ni}}}{\sum_{j=1}^{J} e^{\boldsymbol{\beta}' \mathbf{x}_{nj}}} f(\boldsymbol{\beta} \mid \boldsymbol{\theta}) d\boldsymbol{\beta}$$