

# Trade, Internal Migration, and Human Capital: Who Gains from India's IT Boom?

Devaki Ghose, World Bank, DECRG

June 1, 2026

# Trade and inequality

- Widespread interest in how trade affects regional inequality

# Trade and inequality

- Widespread interest in how trade affects regional inequality
- Aggregate growth masks inequalities: e.g. India (1980-2004)
  - Annual per capita GDP growth more than 6%
  - Trade as fraction of GDP more than doubled
  - Proportion of poor people unchanged at 80% (Kotwal et al.)

# Trade and inequality

- Widespread interest in how trade affects regional inequality
- Aggregate growth masks inequalities: e.g. India (1980-2004)
  - Annual per capita GDP growth more than 6%
  - Trade as fraction of GDP more than doubled
  - Proportion of poor people unchanged at 80% (Kotwal et al.)
- Margins of adjustment to trade shocks:
  - Human capital response to trade
  - Move for work and education

# Trade, mobility frictions, and inequality

- **Unequal access to education:**

e.g. Jarbidge, Nevada vs Charlottesville, Virginia

- **Costly migration:**

Out of all internal migrants, only **2.9%** moved for education (Indian Census, 2001)

# Trade, internal migration, and human capital

## **Overarching Question:**

How does the response of human capital to trade shocks affect inequality?

**Context:** IT boom and engineering expansion in India

# Trade, internal migration, and human capital

## Overarching Question:

How does the response of **human capital** to trade shocks affect inequality?

**Context:** IT boom and engineering expansion in India

- **Endogenous-skill model:** Welfare depends on access to education
- **Fixed-skill model:** Ignores this opportunity for welfare gain
- **Evidence that trade affects human capital:** Blanchard et al., 2017

# Trade, internal migration, and human capital

## **Overarching Question:**

How does the response of human capital to trade shocks affect **regional inequality**?

Welfare depends on regional access to education and jobs separately

- Mobility costs for education
- Mobility costs for work

# Trade, internal migration, and human capital

## Overarching Question:

How does the response of human capital to trade shocks affect **regional inequality**?

Welfare depends on regional access to jobs and education separately

- Mobility costs for education
- Mobility costs for work

**Example:** If Bangalore exports software, how does that affect other regions of India?

# Challenges

- **Modeling:** How to disentangle the different margins of adjustments to trade shocks?
  - Choose sector and location of work as well as field and location of higher education
  - Interdependence between the work and study decisions, as skills determine job opportunities

# Challenges

- **Modeling:** How to disentangle the different margins of adjustments to trade shocks?
  - Choose sector and location of work as well as field and location of higher education
  - Interdependence between the work and study decisions, as skills determine job opportunities
- **Data:** Migration flows across micro-regions, disaggregated by reason for migration.

# Challenges

- **Modeling:** How to disentangle the different margins of adjustments to trade shocks?
  - Choose sector and location of work as well as field and location of higher education
  - Interdependence between the work and study decisions, as skills determine job opportunities
- **Data:** Migration flows across micro-regions, disaggregated by reason for migration.
- **Empirical Setting:** Skill-intensive sector exposed to international demand shocks

# This paper

- ① **Quantitative general equilibrium model** of multiple locations and sectors
  - **New features:** Endogenous education + separate measures of access to jobs and education
  - **First migration cost estimates disaggregated by reason for migration:** Using unique Indian Census data, 2001

# This paper

- ① **Quantitative general equilibrium model** of multiple locations and sectors
  - **New features:** Endogenous education + separate measures of access to jobs and education
  - **First migration cost estimates disaggregated by reason for migration:** Using unique Indian Census data, 2001
- ② **New Stylized Facts:** Heterogeneous education response to IT boom across regions

# This paper

- ① **Quantitative general equilibrium model** of multiple locations and sectors
  - **New features:** Endogenous education + separate measures of access to jobs and education
  - **First migration cost estimates disaggregated by reason for migration:** Using unique Indian Census data, 2001
- ② **New Stylized Facts:** Heterogeneous education response to IT boom across regions
- ③ **Novel policy alternatives to moving jobs**
  - Education policies as tools to address regional welfare inequality from trade

## Contribution to literature

- Trade liberalization and educational attainment:
  - Edmonds et al. (2010), Shastry (2012), Jensen (2012), Oster et al. (2013), Atkin (2016), Blanchard et al. (2017), Danziger (2017), Greenland et al. (2016), Li (2018)

**Contribution:** Add spatial GE effects

## Contribution to literature

- Trade liberalization and educational attainment:
  - Edmonds et al. (2010), Shastri (2012), Jensen (2012), Oster et al. (2013), Atkin (2016), Blanchard et al. (2017), Danziger (2017), Greenland et al. (2016), Li (2018)

**Contribution:** Add spatial GE effects

- Empirical evidence for dynamic supply response (HO)
  - Findlay and Kierzkowski (1983), Harris and Robertson (2013)

**Contribution:** Regional inequality in endogenous comparative advantage

## Contribution to literature

- Trade liberalization and educational attainment:
  - Edmonds et al. (2010), Shastry (2012), Jensen (2012), Oster et al. (2013), Atkin (2016), Blanchard et al. (2017), Danziger (2017), Greenland et al. (2016), Li (2018)

**Contribution:** Add spatial GE effects

- Empirical evidence for dynamic supply response (HO)
  - Findlay and Kierzkowski (1983), Harris and Robertson (2013)

**Contribution:** Regional inequality in endogenous comparative advantage

- Quantitative economic geography models:
  - Multiple sectors: Artuc et al. (2010), Fajgelbaum et al. (2014), Kucheryavyi (2016), Fuchs (2018)
  - Migration Costs: Bryan and Morten (2018), Allen et al. (2016, 2019), Tombe and Zhu (2019)

**Contribution:** Worker field of education and location choice, taking into account long-term occupation choice

# Three main takeaways

Aggregate and distributional consequences of trade:

- ① IT boom increased welfare by 2.39% but with massive inequality

# Three main takeaways

Aggregate and distributional consequences of trade:

- ① IT boom increased welfare by 2.39% but with massive inequality
- ② **Endogenous education response to trade is quantitatively and qualitatively important**
  - Without endogenous education, aggregate welfare from the IT boom is over one-third lower and regional inequality more than twice as large

# Three main takeaways

Aggregate and distributional consequences of trade:

- 1 IT boom increased welfare by 2.39% but with massive inequality
- 2 **Endogenous education response to trade is quantitatively and qualitatively important**
  - Without endogenous education, aggregate welfare from the IT boom is over one-third lower and regional inequality more than twice as large
- 3 **Important to account for education mobility**
  - Without education mobility, aggregate welfare gains from the IT boom would be 25% lower and regional inequality 1.5 times larger

# Outline

- ① Background of the IT boom
- ② Stylized facts
- ③ A GE spatial model with endogenous human capital accumulation
- ④ Data and estimation
- ⑤ Quantification and Counterfactuals

# Background of the IT boom

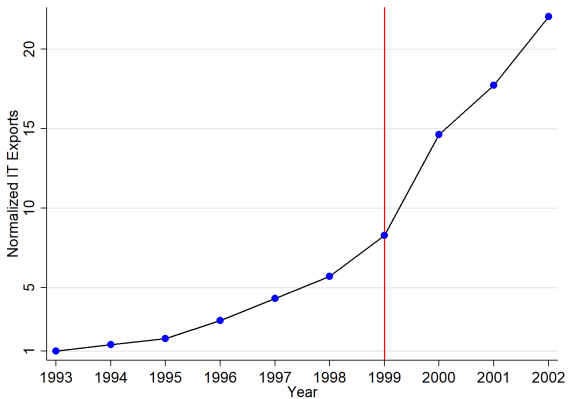
## Setting: Indian IT boom

- India is the world's 2nd largest exporter of IT and IT-enabled services
- More than 80% of IT output exported Exports to Domestic IT ratio

## Setting: Indian IT boom

- India is the world's 2nd largest exporter of IT and IT-enabled services
- More than 80% of IT output exported [Exports to Domestic IT ratio](#)
- International demand shocks for IT in the late 1990s and early 2000s: the Y2K shock, dot-com boom and bust (1998-2001) [Details](#)
- Persistent growth in the IT sector starting 1998

# Growth in IT exports

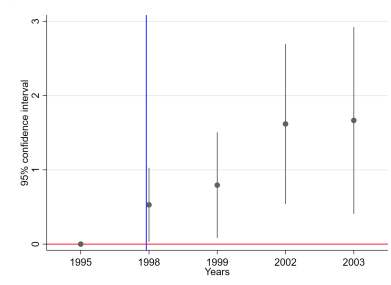


Growth in IT Employment and Engineering Enrollment

# Stylized Facts

- ① **Endogenous human capital response**
- ② **Separate migration costs for work and education**

# 1. Employment and education respond



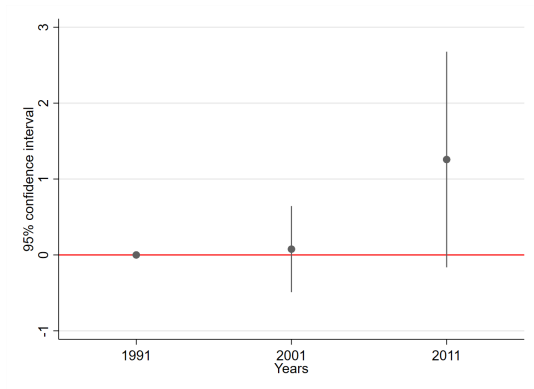
## Event Study:

$$IT \text{ Emp}_{dt} = \alpha_t + \gamma_d + \chi_d * t + \beta_t * I_t * \text{Any Exports}_{1995,d} + \epsilon_{dt}$$

**Clustering:** State, **Data Source:** Economic Census, NASSCOM

Alternative Specifications

# 1. Employment and education respond



$$Eng \ Enrollment_{dt} = \alpha_t + \gamma_d + \chi_d t + \beta_t * I_t * Any \ Exports_{1995,d} + \epsilon_{dt}$$

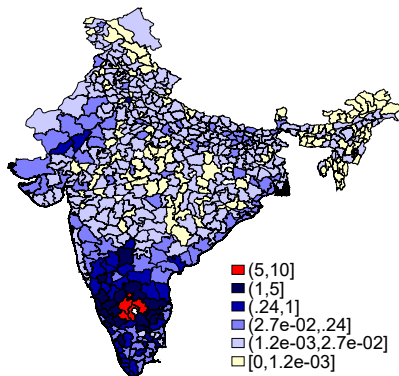
**Data Source:** Decennial Census

## 2. Migration reduces over distance and people migrate more for work than for education

- Out of all migrants in 2001, **48%** move for work and only **3%** move for education (**Source:** 2001 Census) [Histogram Migration Flows](#)

## 2. Migration reduces over distance and people migrate more for work than for education

- Out of all migrants in 2001, **48%** move for work and only **3%** move for education (**Source:** 2001 Census) Histogram Migration Flows



Percentage distribution of in-migrants into Bangalore

A GE spatial model with  
endogenous human capital  
accumulation

## Model environment

**Geography** Many discrete locations  $d$  (including Rest of the World (ROW))  $\in D$  that differ in:

- Distances from all other locations
- Population who can attend college

## Model environment

**Geography** Many discrete locations  $d$  (including Rest of the World (ROW))  $\in D$  that differ in:

- Distances from all other locations
- Population who can attend college

### Individuals

- Stage 1: Choose whether or not to go to college, field of study in college (if yes) and location
- Stage 2: Choose occupation and location

# Model environment

**Geography** Many discrete locations  $d$  (including Rest of the World (ROW))  $\in D$  that differ in:

- Distances from all other locations
- Population who can attend college

## Individuals

- Stage 1: Choose whether or not to go to college, field of study in college (if yes) and location
- Stage 2: Choose occupation and location

## Firms

- Representative firm in each location and in each sector produces a variety subject to costly trade
- Nested CES production function

## Model intuition

- Rise in IT demand  $\rightarrow$  IT prices rise

## Model intuition

- Rise in IT demand  $\rightarrow$  IT prices rise  $\rightarrow$  IT incomes rise

## Model intuition

- Rise in IT demand  $\rightarrow$  IT prices rise  $\rightarrow$  IT incomes rise  $\rightarrow$  Cost of living rises differentially depending on distance

## Model intuition

- Rise in IT demand  $\rightarrow$  IT prices rise  $\rightarrow$  IT incomes rise  $\rightarrow$  Cost of living rises differentially depending on distance
- Real wages rise more in some regions and sectors  $\rightarrow$  move to attractive regions and sectors

## Model intuition

- Rise in IT demand  $\rightarrow$  IT prices rise  $\rightarrow$  IT incomes rise  $\rightarrow$  Cost of living rises differentially depending on distance
- Real wages rise more in some regions and sectors  $\rightarrow$  move to attractive regions and sectors  $\rightarrow$  constrained by access to work and skill-level

## Model intuition

- Rise in IT demand  $\rightarrow$  IT prices rise  $\rightarrow$  IT incomes rise  $\rightarrow$  Cost of living rises differentially depending on distance
- Real wages rise more in some regions and sectors  $\rightarrow$  move to attractive regions and sectors  $\rightarrow$  constrained by access to work and skill-level
- Short-run rise in college skill premium  $\rightarrow$  more enrollment in higher education (esp. engineering) in long-run  $\rightarrow$  depends on access to education

**Distance Matters!**

## Stage 2: Work decision

**Utility:** An individual  $i$  with a degree  $s$  from location  $j$  chooses the sector of work  $S$  and region  $d$  :

$$\text{Max}_{d,S}(U_{ijsdS}^2) = \text{Max}_{d,S} \left( \underbrace{\frac{W_{dsS} U_{dsS} \mu_{jd} \eta_i}{P_d}}_{\text{Stage 2}} \right)$$

$W_{dsS}$ : District ( $d$ ), field of education ( $s$ ) and industry ( $S$ ) specific wage

## Stage 2: Work decision

**Utility:** Utility of an individual  $i$  with a degree in  $s$  from location  $j$  who decides to work in sector  $S$  in region  $d$  is given by:

$$Max_{d,S}(U_{ijsdS}^2) = Max_{d,S} \underbrace{\left( \frac{W_{dsS} U_{dsS} \mu_{jd} \eta_i}{P_d} \right)}_{\text{Stage 2}}$$

$U_{dsS}$ : District ( $d$ ), field of education ( $s$ ) and industry ( $S$ ) specific amenities

## Stage 2: Work decision

**Utility:** Utility of an individual  $i$  with a degree in  $s$  from location  $j$  who decides to work in sector  $S$  in region  $d$  is given by:

$$\text{Max}_{d,S}(U_{ijsdS}^2) = \text{Max}_{d,S} \left( \underbrace{\frac{W_{dsS} U_{dsS} \mu_{jd} \eta_i}{P_d}}_{\text{Stage 2}} \right)$$

$\mu_{jd}$ : Destination district ( $d$ ) and origin district ( $j$ ) specific migration cost for work

## Stage 2: Work decision

**Utility:** Utility of an individual with a degree in **s** from location **j** who decides to work in sector **S** in region **d** is given by:

$$\text{Max}_{d,S}(U_{ijsdS}^2) = \text{Max}_{d,S} \left( \underbrace{\frac{W_{dsS} U_{dsS} \mu_{jd} \eta_i}{P_d}}_{\text{Stage 2}} \right)$$

$\eta_i$ : Individual specific idiosyncratic shock from Frechet  $\sim \theta$

## Stage 2: Work decision

**Utility:** Utility of an individual  $i$  with a degree in  $s$  from location  $j$  who decides to work in sector  $S$  in region  $d$  is given by:

$$\text{Max}_{d,S}(U_{ijsdS}^2) = \text{Max}_{d,S} \left( \underbrace{\frac{W_{dsS} U_{dsS} \mu_{jd} \eta_i}{P_d}}_{\text{Stage 2}} \right)$$

$P_d$ : Composite price index of district  $d$  (cost-of-living)

## Gravity migration equation for work

- Given stage 1 choice, real wages, amenities, migration costs and Fréchet productivity shocks, choose sector and destination of work  $(S, d)$

## Gravity migration equation for work

- Given stage 1 choice, real wages, amenities, migration costs and Fréchet productivity shocks, choose sector and destination of work  $(S, d)$
- **Migration equations for work:**

$$m_{jds} = \frac{\left( \frac{W_{ds} u_{ds}}{P_d} \mu_{jd} \right)^\theta}{\Phi_{js}}$$

## Gravity migration equation for work

- Given stage 1 choice, real wages, amenities, migration costs and Fréchet productivity shocks, choose sector and destination of work ( $S, d$ )
- **Migration equations for work:**

$$m_{jdsS} = \frac{\left( \frac{W_{dsS} u_{dsS}}{P_d} \mu_{jd} \right)^\theta}{\Phi_{js}}$$

$m_{jdsS}$ : Proportion with degree  $s$  from  $j$  moving to  $d$  to work in sector  $S$

## Gravity migration equation for work

- Given stage 1 choice, real wages, amenities, migration costs and Fréchet productivity shocks, choose sector and destination of work ( $S, d$ )
- Migration equations for work:**

$$m_{jdsS} = \frac{\left( \frac{W_{dsS} u_{dsS}}{P_d} \mu_{jd} \right)^\theta}{\Phi_{js}}$$

$m_{jdsS}$ : Proportion with degree  $s$  from  $j$  moving to  $d$  to work in sector  $S$

$\Phi_{js}$ : Access to jobs for an individual from  $j$  with degree  $s$

## Stage 1: Education decision

**Utility:** Utility of an individual living in  $\mathbf{o}$  who chooses to study field  $\mathbf{s}$  in location  $\mathbf{j}$  and then decides to work in sector  $\mathbf{S}$  in region  $\mathbf{d}$  is given by:

$$U_{ijsdS} = \underbrace{\left( \frac{\mu_{oj} a_{js} w_{u,j} \zeta_{ijs}}{P_j} \right)^\beta}_{\text{Stage 1}} \underbrace{\left( U_{ijsdS}^2 \right)^{(1-\beta)}}_{\text{Stage 2}}$$

**Stage 1:** Pay-off from education

**Stage 2:** Pay-off from work

## Stage 1: Education decision

**Utility:** Utility of an individual living in  $\mathbf{o}$  who chooses to study field  $\mathbf{s}$  in location  $\mathbf{j}$  and then decides to work in sector  $\mathbf{S}$  in region  $\mathbf{d}$  is given by:

$$U_{ijsdS} = \underbrace{\left( \frac{\mu_{oj} a_{js} \zeta_{ijs}}{P_j} \right)^\beta}_{\text{Stage 1}} \underbrace{\left( U_{ijsdS}^2 \right)^{(1-\beta)}}_{\text{Stage 2}}$$

$\mu_{oj}$  : Migration cost for education

## Stage 1: Education decision

**Utility:** Utility of an individual living in  $\mathbf{o}$  who chooses to study field  $\mathbf{s}$  in location  $\mathbf{j}$  and then decides to work in sector  $\mathbf{S}$  in region  $\mathbf{d}$  is given by:

$$U_{ijsdS} = \underbrace{\left( \frac{\mu_{oj} a_{js} \zeta_{ijs}}{P_j} \right)^\beta}_{\text{Stage 1}} \underbrace{\left( U_{ijsdS}^2 \right)^{(1-\beta)}}_{\text{Stage 2}}$$

$a_{js}$ : Amenity from education, net of time and money cost

## Stage 1: Education decision

**Utility:** Utility of an individual living in  $\mathbf{o}$  who chooses to study field  $\mathbf{s}$  in location  $\mathbf{j}$  and then decides to work in sector  $\mathbf{S}$  in region  $\mathbf{d}$  is given by:

$$U_{ijsdS} = \underbrace{\left( \frac{\mu_{oj} a_{js} \zeta_{ijs}}{P_j} \right)^\beta}_{\text{Stage 1}} \underbrace{\left( U_{ijsdS}^2 \right)^{(1-\beta)}}_{\text{Stage 2}}$$

$\zeta_{ijs}$ : Idiosyncratic preference shock for education with Fréchet  $\gamma$

## Stage 1: Education decision

- Taking into account stage 2 choice, maximize expected life-time utility by choosing where and what to study ( $j, s$ )

## Stage 1: Education decision

- Taking into account stage 2 choice, maximize expected life-time utility by choosing where and what to study ( $j, s$ )

### Example:

Should I move to the Bay area to study engineering today or should I move tomorrow for jobs?

**Trade-off:** High out of state costs today vs better access to job tomorrow

## Gravity migration equation for education

$$l_{ojs} = \frac{(a_{js}\mu_{oj} \mathbb{E}(U_{jSD}^2))^\gamma}{\Phi_o}$$

$l_{ojs}$ : Proportion of people migrating from  $o$  to  $j$  to study  $s$

$a_{js}$ : Net amenity from education

$\mu_{oj}$ : Migration cost for education

$U_{jSD}^2$ : Expected period 2 utility

$\Phi_o$ : Access to education for a person from  $o$

# Firms

## Nested CES production function

- Firms in industry **S** in location **d** produce output using high-skilled labor  $L_{hdS}$  and low-skilled labor  $L_{ldS}$ :

$$Q_{dS} = \left( Q_{hdS}^{\frac{\rho_S-1}{\rho_S}} + Q_{ldS}^{\frac{\rho_S-1}{\rho_S}} \right)^{\frac{\rho_S}{\rho_S-1}}$$

- $Q_{hdS}$ : Composite of workers with engineering and other college degrees
- There is costly trade between districts

## Trade with the ROW

- ROW demands IT from different districts:

$$E_{d,IT} = \underbrace{\left( \frac{\tau_d p_d}{\sum_{d'} (\tau_{d'}, p_{d'})^{1-\sigma_{IT}}} \right)^{(1-\sigma_{IT})}}_{\text{Gravity}} E_{ROW,IT}$$

- Determinants of exports:
  - (-) Price ( $p_d$ )
  - (+) Export-cost shifters: e.g, Historical exports (Banerjee and Duflo 2001), Linguistic distance from English (Shastry, 2012)

## General equilibrium

Given the distribution of population, education and work amenities, productivities across regions, bilateral migration and trade costs, and model parameters, equilibrium is defined as a set of {sectoral employment  $\{L_{dsS}\}$ , field-wise college enrollment  $\{L_{js}\}$ , wages  $\{W_{dsS}\}$ , prices  $\{P_d\}$ , quantities  $\{Q_{dS}\}$ } such that in each region

1. Markets clear for each sector  $S$
2. Trade balances for each sector  $S$
3. Labor demand for each type  $s$  of labor in each sector  $S$  equals labor supply

**A Competitive general equilibrium exists**

Data

# Data

## ① Wages and employment

- Unit: District-industry-year-field of education
- Source: Economic Census of India, NSSO

# Data

## ① Wages and employment

- Unit: District-industry-year-field of education
- Source: Economic Census of India, NSSO

## ② Enrollment

- Unit: District-year-field of education
- Source: Population Census of India, NSSO

# Data

## ① Wages and employment

- Unit: District-industry-year-field of education
- Source: Economic Census of India, NSSO

## ② Enrollment

- Unit: District-year-field of education
- Source: Population Census of India, NSSO

## ③ IT export values

- Unit: District-year
- Source: National Association of Software Companies in India (NASSCOM)

# Data

## ① Wages and employment

- Unit: District-industry-year-field of education
- Source: Economic Census of India, NSSO

## ② Enrollment

- Unit: District-year-field of education
- Source: Population Census of India, NSSO

## ③ IT export values

- Unit: District-year
- Source: National Association of Software Companies in India (NASSCOM)

## ④ Internal migration

- Unit: Origin-destination-reason for Migration
- Source: Unique bilateral migration data from 2001 Population Census

## Summary statistics: District level

Variable	Pre Boom		Post Boom	
	Mean	Std. Dev.	Mean	Std. Dev.
<b>Labor market: Wages (in 1994 CPI INR)</b>				
<b>High-skill services</b>				
College-educated	20,295	3,494	51,875	25,228
Non college-educated	8,450	3,369	30,279	17,587
<b>Manufacturing</b>				
College-educated	17,632	6,034	49,475	30,942
Non college-educated	6,391	3,546	15,012	7,248

Data

## Summary statistics: District level

Variable	Pre Boom		Post Boom	
	Mean	Std. Dev.	Mean	Std. Dev.
<b>Labor market: Employment</b>				
<b>High-skill services</b>				
College-educated	1,384	1,256	3,940	9,700
Non college-educated	5,279	8,031	12,614	25,407
<b>Manufacturing</b>				
College-educated	6,840	5,591	6,016	8,794
Non college-educated	18,043	30,379	32,618	58,059

Data

## Summary statistics: District level

	<b>Pre Boom</b>	<b>Post Boom</b>
<b>Education</b>		
Engineers as a percentage of college graduates	4.90%	10.40%
Total college graduates	25,132	61,266

Data

# Identification and Estimation

# Estimation

- Migration costs:  $\mu_{jds}$  (for work),  $\mu_{oj}$  (for education)
- Migration elasticities:  $\theta$  (for work),  $\gamma$  (for education)
- Trade costs:  $\tau_{d,ROW,IT}$
- Trade elasticity:  $\sigma_{IT}$

Estimation: Migration costs

## Identification challenges

- Is labor supply in a region high due to better unobserved amenities or because the region is closer to other populous regions?
- Do some regions have higher college enrollment because they offer particularly good education or they are closer to regions with more young people?

## Estimation: Stage 1 migration costs

- **Education migration flow in model:**

$$l_{ojs} = \frac{(a_{js}\mu_{oj} \mathbb{E}(U_{jSD}^2))^\gamma}{\Phi_o}$$

- **Education migration flow in data:** Proportion of people moving from  $o$  to  $j$  for education, aggregated across all fields of education
- **Sample:** Cross-section of people who migrate only for education (including the option to remain unskilled) in 2001

# Estimation: Stage 1 migration costs

## PPML estimation:

$$\log(I_{oj}) = f_j + f_o + \beta \log \text{dist}_{oj} + \epsilon_{oj}$$

where  $\gamma \ln(\mu_{oj}) = \beta \ln \overrightarrow{\text{Distance}}_{oj}$

Distance vector:

- 1 Distance centroid
- 2 Common language overlap
- 3 Same-state neighbors
- 4 Same-state not-neighbors
- 5 Different-state neighbors

## Estimation: Stage 1 migration costs

### PPML estimation:

$$\log(l_{oj}) = f_j + f_o + \beta \log dist_{oj} + \epsilon_{oj}$$

**Solve the model for amenities:** Use estimated migration costs, enrollment by degrees, to recover field-of-education specific net amenities

## Estimation: Stage 2 migration costs

- **Work migration flow in model:**

$$m_{jdsS} = \frac{\left( \frac{W_{dsS} u_{dsS}}{P_d} \mu_{jd} \right)^\theta}{\Phi_{js}}$$

## Estimation: Stage 2 migration costs

- **Work migration flow in model:**

$$m_{jds} = \frac{\left( \frac{W_{ds} u_{ds}}{P_d} \mu_{jd} \right)^\theta}{\Phi_{js}}$$

- **Work migration flow in data:** Proportion of people moving from  $j$  to  $d$  for work, aggregated across all fields of education and sectors of work

## Estimation: Stage 2 migration costs

- **Work migration flow in model:**

$$m_{jds} = \frac{\left( \frac{W_{ds} u_{ds}}{P_d} \mu_{jd} \right)^\theta}{\Phi_{js}}$$

- **Work migration flow in data:** Proportion of people moving from  $j$  to  $d$  for work, aggregated across all fields of education and sectors of work

- **Log aggregated:**

$$\ln(m_{jd}) = \underbrace{\theta \ln(\mu_{jd})}_{\text{Migration Costs}} \quad \overset{\text{destination f.e.}}{\underbrace{-\theta \ln P_d}} \quad + \underbrace{\theta \ln \left( \sum_{sS} \frac{W_{dsS}}{\Phi_{js}} \right) + \theta \left( \ln \sum_s u_{dsS} \right)}_{\text{Unobserved}}$$

## Estimation: Stage 2 migration costs

Log aggregated:

$$\ln(m_{jd}) = \underbrace{\theta \ln(\mu_{jd})}_{\text{Migration Costs}} \underbrace{-\theta \ln P_d}_{\text{destination f.e.}} + \underbrace{\theta \ln \left( \sum_{sS} \frac{W_{dsS}}{\Phi_{js}} \right) + \theta \left( \ln \sum_s u_{dsS} \right)}_{\text{Unobserved}}$$

Traditional gravity specification:

$$\ln(m_{jd}) = \gamma \ln \text{Distance}_{jd} + \beta_d + \alpha_j + \epsilon_{jd}$$

# Estimation: Stage 2 migration costs

## Log aggregated:

$$\ln(m_{jd}) = \underbrace{\theta \ln(\mu_{jd})}_{\text{Migration Costs}} \underbrace{-\theta \ln P_d}_{\text{destination f.e.}} + \underbrace{\theta \ln \left( \sum_{sS} \frac{W_{dsS}}{\Phi_{js}} \right) + \theta \left( \ln \sum_s u_{dsS} \right)}_{\text{Unobserved}}$$

## Traditional gravity specification:

$$\ln(m_{jd}) = \gamma \ln \text{Distance}_{jd} + \beta_d + \alpha_j + \epsilon_{jd}$$

## Problems:

1. Unobserved errors correlated
2. Migration costs correlated with unobserved errors

Traditional Gravity Specification

Intuition: Empirical Model Misspecification

Two Step Approach

# Estimation: Elasticity of migration flows to distance

**Sample:** Cross-section of people who migrate only for work in 2001

**Assumption:** After accounting for unobserved structural quantities, the remaining error is pure white noise

**Method: Nested Non-linear Least Squares**

- Inner loop: Use sectoral employment levels to back out amenities, given a guess of migration costs for work

# Estimation: Elasticity of migration flows to distance

**Sample:** Cross-section of people who migrate only for work in 2001

**Assumption:** After accounting for unobserved structural quantities, the remaining error is pure white noise

**Method: Nested Non-linear Least Squares**

- Inner loop: Use sectoral employment levels to back out amenities, given a guess of migration costs for work
- Outer loop: Choose parameters that minimize the distance between predicted and actual migration flows, given amenities

Further estimation: Dispersion parameter

## Migration costs: Result highlights

- **Average migration costs:**

Migrants on average have to be compensated **90%- 91%** more than non-migrants when moving for work and education, respectively

- **Huge state border effects:**

Crossing state borders adds about **19 percentage points** to migration costs

## Estimation Results

	(1)	(2)
	Education	Work
	PPML	NLS
log distance centroid	-.58***	-.90***
	(-24.70)	(-25.75)
log common language	.37***	.15
	(9.06)	(0.89)
Same state, neighbors	3.04***	5.53***
	(14.24)	(10.67)
Same state, not neighbors	1.97***	4.03**
	(10.26)	(9.03)
Different state, neighbors	2.02***	3.65*
	(11.62)	(7.45)
No. of observations	339889	339889

Column (1): PPML for education. Column (2): Nonlinear Least Squares for work.

Estimation: Trade costs

## External trade costs

Use the gravity equation for exports and take first difference,

$$\underbrace{\Delta \ln \left( \frac{E_{d,t}}{E_{ROW,t}} \right)}_{\text{Trade Flows}} = \Delta(1 - \sigma_{IT}) \underbrace{\ln(p_{d,IT,t})}_{\text{Marginal Costs}}$$

## External trade costs

Using  $P=MC$ ,

$$\Delta \ln \left( \frac{E_{d,t}}{E_{ROW,t}} \right) = \Delta (1 - \sigma_{IT}) \ln f(W_t, L_t) + \Delta \ln g(A_{d,IT,t})$$

- **Challenge:** Unobserved productivity changes move labor demand

## External trade costs

Using  $P=MC$ ,

$$\Delta \ln \left( \frac{E_{d,t}}{E_{ROW,t}} \right) = \Delta (1 - \sigma_{IT}) \underbrace{\ln f(W_t, L_t)}_{\text{Need Instrument}} + \Delta \ln g(A_{d,IT})$$

- **Challenge:** Unobserved productivity changes move labor demand
- **Approach:** Use the 2000 IT boom, combined with existing regional differences to instrument for changes in MC
- **Identifying Assumption:** Changes in productivity between the pre and post boom periods are uncorrelated with the instruments

## Three instruments

- ① A measure of labor supply access for each region, summarized by

$$LMA_d = \sum_o L_o(\text{distance}_{o,d})^{-1}$$

## Three instruments

- 1 A measure of labor supply access for each region, summarized by

$$LMA_d = \sum_o L_o(\text{distance}_{o,d})^{-1}$$

- 2 The share of historical software exports from a region, measured in 1995
- 3 The historical linguistic distance of each district from English, estimated using 1991 population Census

## Trade cost estimation: Results

	(1)	(2)
	Exports OLS	Exports IV
Change in observable MC	-0.06 (-1.14)	-4.09** (-2.30)
IV	No	Yes
State-time trend	Yes	Yes
First Stage F-stat:	-	58.19***
N	522	522

Standard errors are clustered at the state level. t-statistics reported in parentheses

Elasticity of substitution between IT products:  $\sigma_{IT} = 5.09$  [Full Results](#)

[First Stage Details](#): Instruments valid (Hansen overidentification test)

## Solve for unobserved productivities

- Firm profit maximization yields the ratio of unknown productivities as a function of observed wages and employment
- Use  $P=MC$ , along with observed portion of MC to recover the level of sectoral and skill specific productivity

## Summary of parameters

Parameter	Value	Source
Productivity dispersion ( $\theta$ )	2.62	Estimated
Education amenity dispersion ( $\gamma$ )	1.69	Estimated
IT Trade elasticity ( $\sigma_{IT}$ )	5.09	Estimated
$\rho_S^*$	1.41	Katz and Murphy (1992)
$\rho_{hS}^*$	2	Ryoo and Rosen (2004)
Internal Trade Elasticity ( $\sigma_S$ )	5	Simonovska and Waugh (2014)
Agri share	.38	Ministry of Statistics, Govt of India
Manu share	.16	-
High-skill Services	.07	-
Other-services	.37	-

$\rho_S^*$ : Elasticity of substitution between high and low skill workers

$\rho_{hS}^*$ : Elasticity of substitution between workers with engineering and other degrees

Internal Trade Cost Estimation

# Model validity

## 1 Model is consistent with reduced-form facts

- IT employment and engineering enrollment positively respond in districts with historical software exports over the pre and post boom periods Response of IT employment and engineering enrollment
- Employment response is heterogeneous and stronger in districts with higher historical engineering enrollment. Heterogeneous Response

## 2 Model fit: Non-targeted

Evolution of IT Employment and Engineering Enrollment in Data and Model

# Quantification and counterfactuals

# Aggregate and distributional consequences of the IT boom

How does the IT boom (1995-2000) change aggregate and regional welfare in the short and long run?

## Long-run welfare (2002-2007):

$$\Gamma\left(1 - \frac{1}{\gamma}\right) \left( \underbrace{\Phi_o}_{\text{Access to education}} \right)^{\frac{1}{\gamma}}$$

where  $\Phi_o = \sum_{j',s'} \left( (a_{j's'} \mu_{oj'})^{IU} \Phi_{j's'}^{\frac{(1-IU)}{\theta}} \right)^{\frac{\gamma}{IU}}$

# Aggregate and distributional consequences of the IT boom

How does the IT boom (1995-2000) change aggregate and regional welfare in the short and long run?

## Long-run welfare (2002-2007):

$$\Gamma\left(1 - \frac{1}{\gamma}\right) \left( \underbrace{\Phi_o}_{\text{Access to education}} \right)^{\frac{1}{\gamma}}$$

where  $\Phi_o = \sum_{j',s'} \left( (a_{j's'} \mu_{oj'})^{IU} \Phi_{j's'}^{\frac{(1-IU)}{\theta}} \right)^{\frac{\gamma}{IU}}$

$\Phi_{j's'}$ : Access to jobs for individual with skill  $s'$  from  $j'$

# Aggregate and distributional consequences of the IT boom

How does the IT boom (1995-2000) change aggregate and regional welfare in the short and long?

## Long-run welfare (2002-2007):

$$\Gamma\left(1 - \frac{1}{\gamma}\right) \left( \underbrace{\Phi_o}_{\text{Access to education}} \right)^{\frac{1}{\gamma}}$$

where  $\Phi_o = \sum_{j',s'} \left( (a_{j's'} \mu_{oj'})^{IU} \Phi_{j's'}^{\frac{(1-IU)}{\theta}} \right)^{\frac{\gamma}{IU}}$

$\mu_{oj'}$ : Migration cost from  $o$  to  $j'$

# Aggregate and distributional consequences of the IT boom

How does the IT boom (1995-2000) change aggregate and regional welfare in the short and long-run?

## Long-run welfare (2002-2007):

$$\Gamma\left(1 - \frac{1}{\gamma}\right) \left( \underbrace{\Phi_o}_{\text{Access to education}} \right)^{\frac{1}{\gamma}}$$

where  $\Phi_o = \sum_{j',s'} \left( (a_{j's'} \mu_{oj'})^{IU} \Phi_{j's'}^{\frac{(1-IU)}{\theta}} \right)^{\frac{\gamma}{IU}}$

$a_{j's'}$ : Amenities of studying  $s'$  at  $j'$

# Aggregate and distributional consequences of the IT boom

How does the IT boom (1995-2000) change aggregate and regional welfare in the short and long run?

**Long-run welfare (2002-2007) (Endogenous education):**

$$\Gamma\left(1 - \frac{1}{\gamma}\right) \left( \underbrace{\Phi_o}_{\text{Access to Education}} \right)^{\frac{1}{\gamma}}$$

**Short-run welfare (1995-2000) (Fixed skill):**

$$\Gamma\left(1 - \frac{1}{\theta}\right) \left( \Phi_{Eng}^{1/\theta} propEng + \Phi_{NonEng}^{1/\theta} propNonEng + \Phi_{Un}^{1/\theta} propUn \right)$$

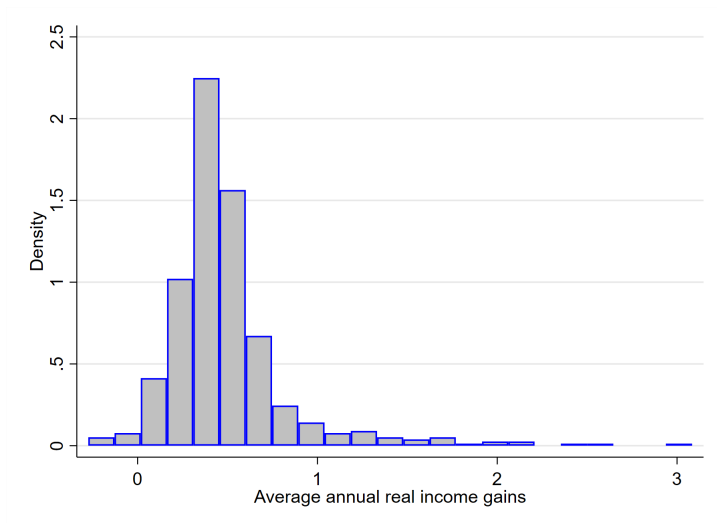
# Aggregate and distributional consequences of the IT boom

How does the IT boom (1995-2000) change aggregate and regional welfare?

Table: Welfare gains from the IT boom

	1) Baseline	2) No edu mobility	3) No endogenous edu
<b>Aggregate Welfare (EV)</b>			
Welfare	2.39%	1.80%	1.09%
<b>Inequality (CV)</b>			
CV	0.018	0.027	0.06

## Histogram of welfare gains from the IT boom



Welfare gains are uneven across regions. Without educational mobility, regional inequality is 1.5 times higher

# Heterogeneity in welfare gains

Table: Welfare gains from the IT boom

	1) Baseline	2) No edu mobility	3) No endogenous edu
<b>Aggregate Welfare (EV)</b>			
Welfare	2.39%	1.80%	<b>1.09%</b>
<b>Inequality (CV)</b>			
CV	0.018	0.027	<b>0.06</b>

Without endogenous education, welfare gains are below half and inequality more than doubles

# How important is mobility for education?

## **Welfare gains from the IT Boom in a model without education mobility:**

- Regional inequality:
  - 1.5 times higher regional inequality (coefficient of variation: 0.027 vs 0.018)

# How important is mobility for education?

## **Welfare gains from the IT Boom in a model without education mobility:**

- Regional inequality:
  - 1.5 times higher regional inequality (coefficient of variation: 0.027 vs 0.018)
- Aggregate welfare gains:
  - 25% lower (1.80% vs 2.39%)

## Reductions in work vs education migration costs

- Reducing education migration costs by 72% raises welfare to 8.19%

## Reductions in work vs education migration costs

- Reducing education migration costs by 72% raises welfare to 8.19%
- Same reduction in work migration costs yields only 2.50%—reducing education mobility costs is  $3.3\times$  more effective
- Without higher education, workers find it harder to find jobs in high-skill sectors such as IT

## Conclusion

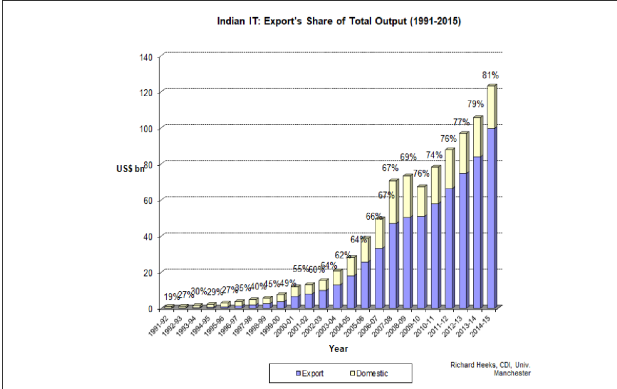
- **New quantitative spatial equilibrium framework:** Skill choice and location of education in a classic economic geography model
- **Role of frictions:** Separate estimates of migration costs by reasons for migration
- **Work vs education migration costs:** For high-skill shocks, education costs bind more because workers must first acquire skills to access expanding jobs
- **Education response:** Significantly different aggregate and distributional effects of trade across regions
- **Policies:** Education + labor market policies to mitigate regional inequality due to trade

Thank You. For comments/ questions/ suggestions:  
dghose@worldbank.org

# Appendix

- 1 Motivating Facts
- 2 Reduced Form: Alternative Specifications
- 3 Estimation and Model Validation

# Export's share of IT Output



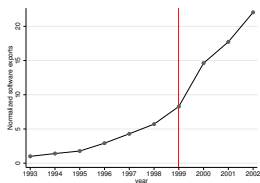
Back

## Details of Shocks

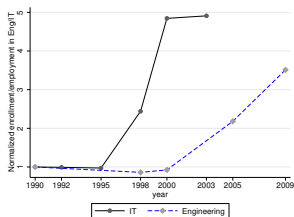
- NASSCOM (2000) estimated India's software exports at \$2.65 billion in 1998 - 1999, with Y2K-related projects accounting for 20% of Total Revenue, (about \$560 million)

[Back](#)

# Growth in IT Employment and Engineering Enrollment



Growth in IT exports over time, normalized to their 1993 levels.



Growth in IT employment and engineering enrollment over time, normalized to their 1990-1991 levels.

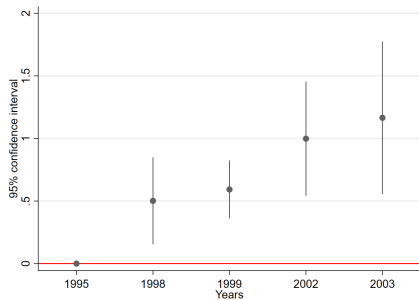
# Appendix

1 Motivating Facts

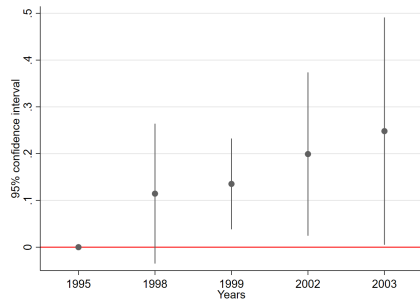
2 Reduced Form: Alternative Specifications

3 Estimation and Model Validation

# Alternative Specifications



Continuous measure: log number of Y2K projects received



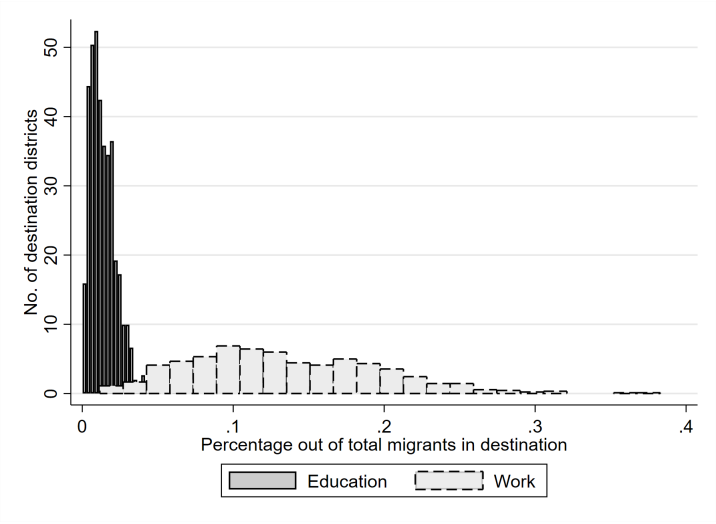
Continuous measure: log number of pre-existing IT specialization areas, 1992 and 1995

[Back](#)

# Appendix

- 1 Motivating Facts
- 2 Reduced Form: Alternative Specifications
- 3 Estimation and Model Validation

# Histogram of Migration Flows by Reason for Migration



## Traditional Gravity Specification

- Recover migration costs:  $\ln(m_{jd}) = \gamma \log(\mu_{jd}) + \alpha_j + \beta_d + \epsilon_{jd}$

# Traditional Gravity Specification

- Recover migration costs:  $\ln(m_{jd}) = \gamma \log(\mu_{jd}) + \alpha_j + \beta_d + \epsilon_{jd}$

**Proposition:**  $\gamma$  is biased and inefficient when

1. The costs of migration differ by reason for migration
2. Wages depend on skills.

# Intuition: Empirical Model Misspecification

## Work Migration Flow:

$$m_{jds} = \frac{\left(\frac{W_{ds} u_{ds}}{P_d} \mu_{jd}\right)^{(\theta)}}{\Phi_{js}}$$

where  $\Phi_{j,s}$  is the option value of education for person from  $j$  with degree  $s$

Back

# Intuition: Empirical Model Misspecification

## Work Migration Flow:

$$m_{jds} = \frac{\left(\frac{W_{ds} u_{ds}}{P_d} \mu_{jd}\right)^\theta}{\Phi_{js}}$$

where  $\Phi_{j,s}$  is the option value of education for person from  $j$  with degree  $s$

Taking logs and summing,

$$\ln(m_{jd}) = \underbrace{\theta \log(\mu_{jd})}_{\text{Migration Costs}} \underbrace{-\theta \log P_d}_{\text{destination f.e.}} + \underbrace{\theta \log\left(\sum_{sS} \frac{W_{dsS}}{\Phi_{js}}\right) + \theta \left(\log \sum_s u_{dsS}\right)}_{\text{Unobserved}}$$

**Traditional Gravity Specification:**  $\ln(m_{jd}) = \gamma \log(\mu_{jd}) + \alpha_j + \beta_d + \epsilon_{jd}$

## Intuition: Empirical Model Misspecification

Taking logs and summing,

$$\ln(m_{jd}) = \underbrace{\theta \log(\mu_{jd})}_{\text{Migration Costs}} \underbrace{-\theta \log P_d}_{\text{destination f.e.}} + \underbrace{\theta \log\left(\sum_{sS} \frac{W_{dsS}}{\Phi_{js}}\right) + \theta \left(\log \sum_s u_{dsS}\right)}_{\text{Unobserved}}$$

**Traditional Gravity Specification:**  $\ln(m_{jd}) = \gamma \log(\mu_{jd}) + \alpha_j + \beta_d + \epsilon_{jd}$

**Problems:** 1. Unobserved errors correlated

2. Migration costs correlated with unobserved errors [Back](#)

# Two Step Approach

## Migration Data

### Ideal Data: By Reason & Skill

Edu	Work	S1	S1
College		x1	x2
No College		x3	x4

### Actual Data: By degree

College	z1
No College	z2

Back

# Two Step Approach

## Migration Data

### Ideal Data: By Reason & Skill

Edu	Work	S1	S1
College		x1	x2
No College		x3	x4

### Actual Data: By Reason

Edu	y1
Work	y4

Back

# Two Step Approach

## Migration Data

### Ideal Data: By Reason & Skill

Edu	Work	S1	S1
College		x1	x2
No College		x3	x4

### Actual Data: By Reason

Edu	y1
Work	y4

- Use data dis-aggregated by reason for migration
- Use migration flow due to work and working population to jointly recover work migration costs and amenities

## Migration Costs: Estimate Dispersion $\theta$

**Challenge:** How to separate the elasticity of migration flows to migration costs (dispersion parameter  $\theta$ ) and the elasticity of migration costs to distances ( $\lambda$ )?

## Migration Costs: Estimate Dispersion $\theta$

**Challenge:** How to separate the elasticity of migration flows to migration costs (dispersion parameter  $\theta$ ) and the elasticity of migration costs to distances ( $\lambda$ )?

**Sample:** Use only the sample of stayers

## Migration Costs: Estimate Dispersion $\theta$

**Challenge:** How to separate the elasticity of migration flows to migration costs (dispersion parameter  $\theta$ ) and the elasticity of migration costs to distances ( $\lambda$ )?

**Sample:** Use only the sample of stayers

**Procedure:**

- $\log(W_{isdS}^{\sim}) = F_{sdS} + \log\eta_{isdS}$
- Variance of exponentiated residuals identifies  $\theta=2.62$

Back

# PPML Gravity: Migration by Reason

Table: PPML gravity estimation on district-to-district migration by reason for migration

	(1)	(2)	(3)
	Education	Work	Other reasons
log distance centroid	-.58*** (-24.70)	-.53*** (-40.41)	-.71*** (-23.51)
log common language	.37*** (9.06)	.36*** (6.26)	.34*** (9.07)
Same state, neighbors	3.04*** (14.24)	2.48*** (8.61)	2.59*** (12.67)
Same state, not neighbors	1.97*** (10.26)	1.40*** (5.56)	1.18*** (7.01)
Different state, neighbors	2.02*** (11.62)	2.41*** (11.40)	2.50*** (15.88)
Observations	339889	339889	339889

PPML gravity by reason for migration. Standard errors clustered two-way by origin and destination states.

# Trade Cost Estimation: Results

Table: Trade Cost Estimation

	(1)	(2)	(3)	(4)
	Exports	Exports	Exports	Exports
	OLS	OLS	IV	IV
Change in observable MC	-0.06 (-1.03)	-0.06 (-1.14)	-4.53** (-2.32)	-4.09** (-2.30)
State-time Trend	No	Yes	No	Yes
IV	No	No	Yes	Yes
First Stage F-stat:	-	-	40.81***	58.19***
N	525	522	525	522

Standard errors are clustered at the state level. t-statistics reported in parentheses

Elasticity of substitution between IT products:  $\sigma_{IT} = 5.09$

[Back](#)

## Trade Cost Estimation: First Stage

- The Kleibergen-Paap rk LM statistic rejects the null that the instruments are underidentified
- The Kleibergen-Paap Wald rk F statistic rejects that the instruments are weak

Back

# Internal Trade Costs

$$P_i^{1-\sigma_S} = \underbrace{\sum_k \tau_{ik}^{1-\sigma_S} P_{kS}^{1-\sigma_S}}_{\text{Weighted Price Index}}$$

$$\underbrace{Y_{dS}}_{\text{Income of (S,d)}} = \sum_j \tau_{dj}^{1-\sigma_S} P_{dS}^{1-\sigma_S} P_j^{\sigma_S-1} \underbrace{(E_{jS})}_{\text{Expenditure of j on S}}$$

# Internal Trade Costs

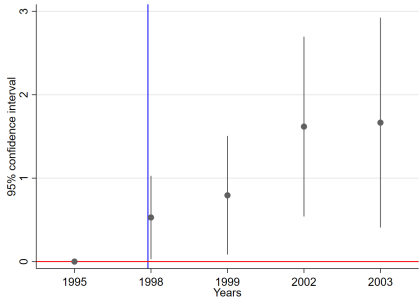
$$P_i^{1-\sigma_S} = \underbrace{\sum_k \tau_{ik}^{1-\sigma_S} P_{kS}^{1-\sigma_S}}_{\text{Weighted Price Index}}$$

$$\underbrace{Y_{dS}}_{\text{Income of (S,d)}} = \sum_j \tau_{dj}^{1-\sigma_S} P_{dS}^{1-\sigma_S} P_j^{\sigma_S-1} \underbrace{(E_{jS})}_{\text{Expenditure of j on S}}$$

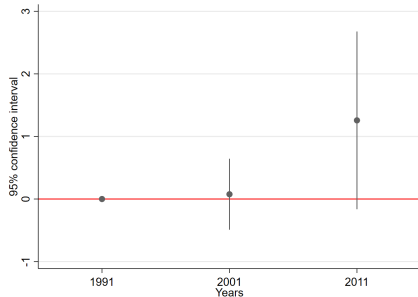
Solve set of equations for  $P_{dS}^{\sigma_S-1}$  where

- $E_{jS}$  and  $Y_{dS}$  are known from the data
- Calibrate  $\tau_{dj}^{1-\sigma_S} = \text{dist}_{d,j}^{-1}$  (Head and Mayer, 2013)

# Response of IT Employment and Engineering Enrollment



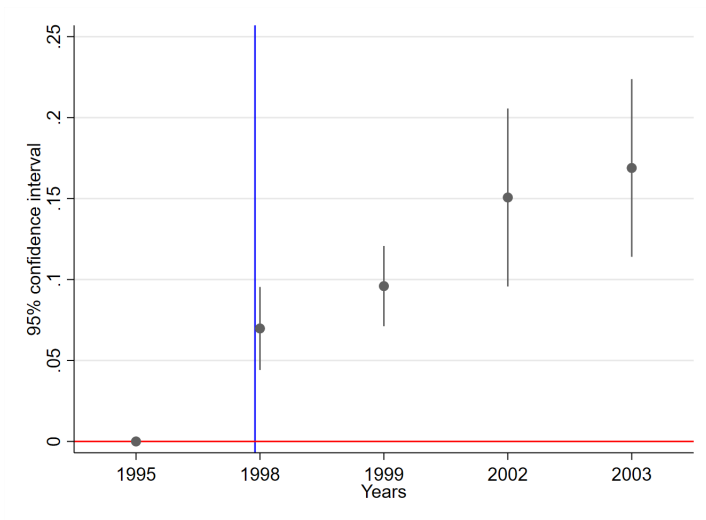
The year-by-year response of standardized IT employment.



The year-by-year response of standardized engineering enrollment.

[Back](#)

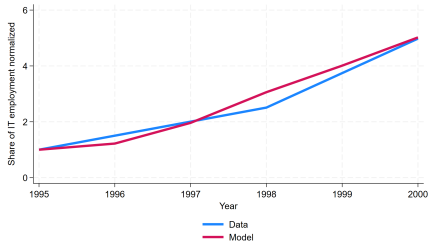
# Heterogeneous Response



The year-by-year heterogeneous response of IT employment to differences in historical engineering enrollment

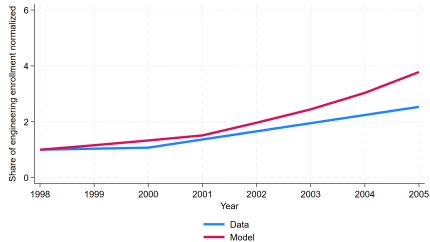
[Back](#)

# Evolution of IT Employment and Engineering Enrollment Over Time in Data and Model



The share of normalized IT employment year by year.

[Back](#)



The share of normalized engineering enrollment year by year.