Estimating the Effects of Globalization
Lecture 2: Trade and Inequality

2014 CREi Lectures in Macroeconomics
Dave Donaldson (MIT)

3 Lectures, 3 "What If?" Trade Questions

- **Lecture #1:** What would have happened to aggregate welfare if China hadn't entered global trade?
- **Lecture #2:** What would happen to inequality if trade were to disappear?
- **Lecture #3:** What would have happened to US welfare if Trump hadn't started his trade war?

- But major focus on methodology: what can economists do to improve their answers to questions like these?
- 100% joint work with Rodrigo Adao (Chicago) and Arnaud Costinot (MIT)

These Lectures: 2 Strategies for Improving Credibility of Structural Estimation

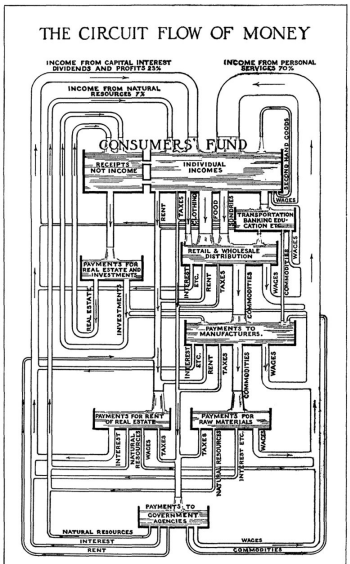
- **Strategy #1: Reduce what needs to be estimated**
  - Power of data is limited, so use it for what matters for causal question of interest
  - Related: "Marschak's Maxim" (Heckman, 2010), "Sufficient Statistics" (Chetty, 2009)
- **Strategy #2: Guess and "verify"**
  - Tools from program evaluation may not be able to answer the desired question
  - But they can still be used to check that the model's causal responses (of interest) align with those in the data
- **Key point:** both depend intimately on the question and the available data

Today's Theme: Unequal Effects of Trade

- Based on Adao, Rodrigo, Paul Carrillo, Arnaud Costinot, Dave Donaldson, and Dina Pomeranz (2022). "Imports, exports, and earnings inequality: Measures of exposure and estimates of incidence." *Quarterly Journal of Economics* 137(3): 1553-1614.

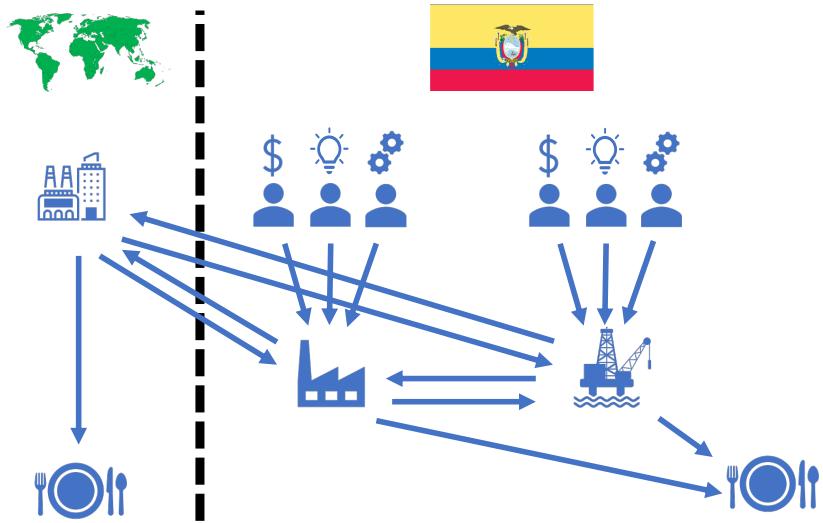
- Context:
  - **Question:** What would happen to earnings inequality if trade were to disappear?
  - **Data:** Rich administrative microdata (from Ecuador)—thanks to collaboration with tax authorities started by my coauthors: Paul Carrillo (GWU), Dina Pomeranz (Zurich), and Monica Singhal (UC Davis)

New Datasets on Economic Linkages

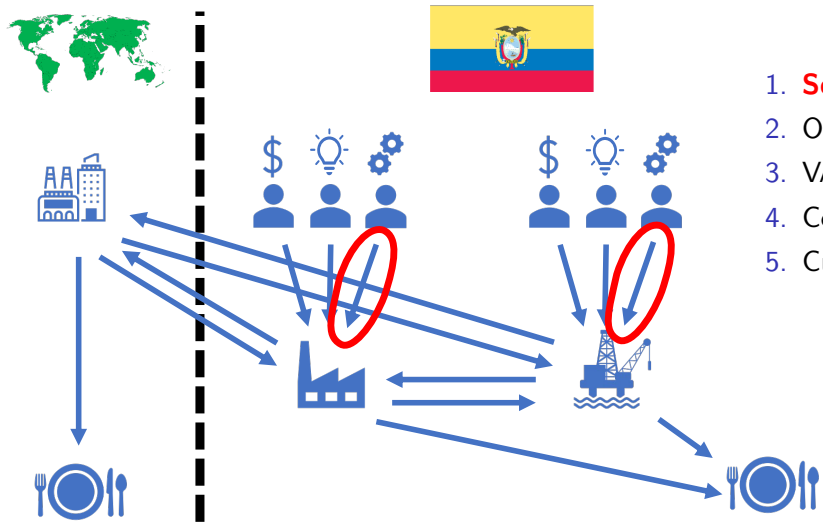


W.T. Foster (*AER*, 1922): "Unfortunately, the statistics upon which the most important conclusions...must be based are not at hand and are not likely to be for a long time to come."

5 Administrative Datasets on Economic Linkages

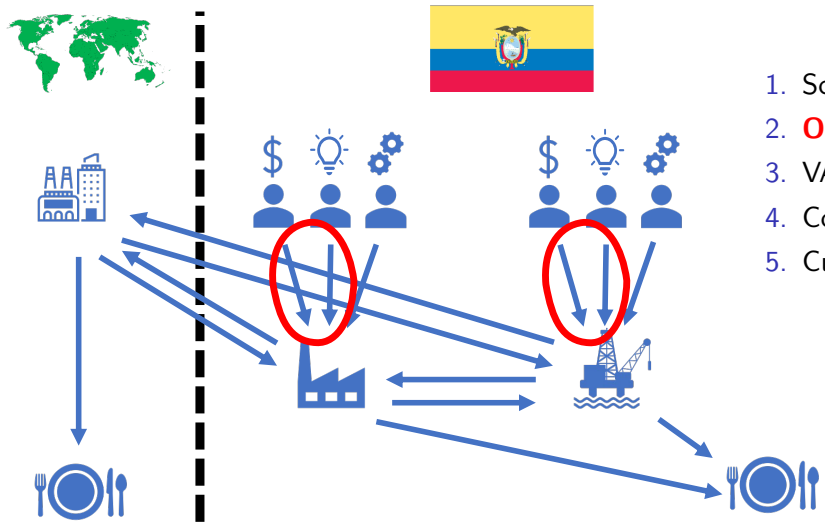


5 Administrative Datasets on Economic Linkages



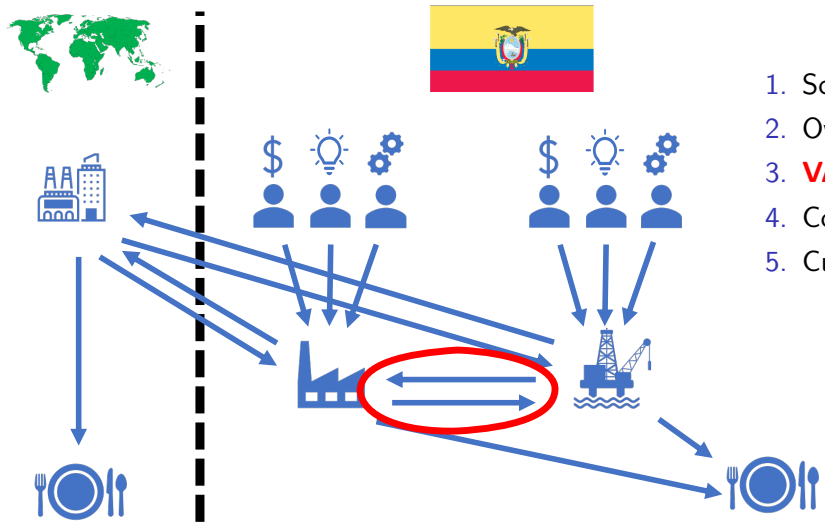
1. **Social security**
2. Ownership registry
3. VAT records
4. Corp. income tax
5. Customs records

5 Administrative Datasets on Economic Linkages



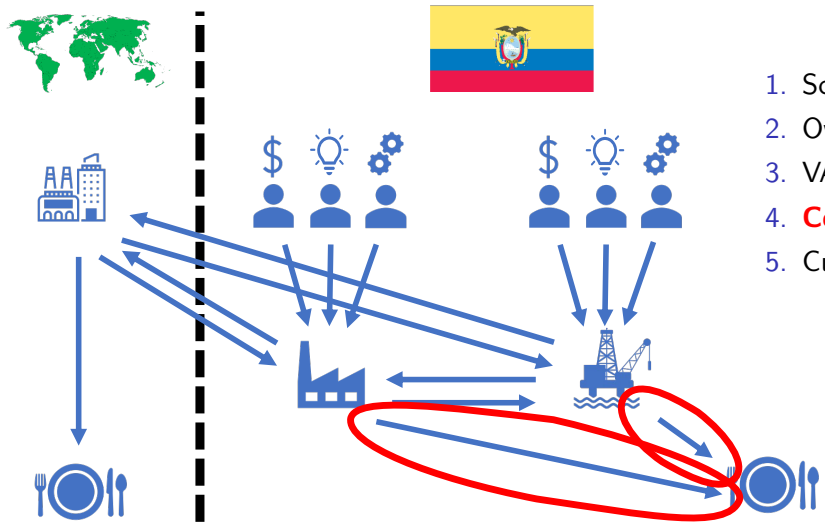
1. Social security
2. **Ownership registry**
3. VAT records
4. Corp. income tax
5. Customs records

5 Administrative Datasets on Economic Linkages



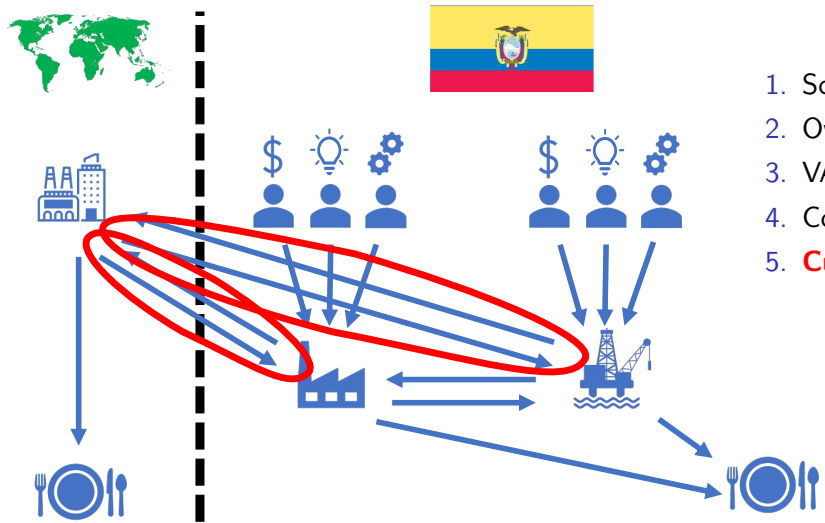
1. Social security
2. Ownership registry
3. **VAT records**
4. Corp. income tax
5. Customs records

5 Administrative Datasets on Economic Linkages



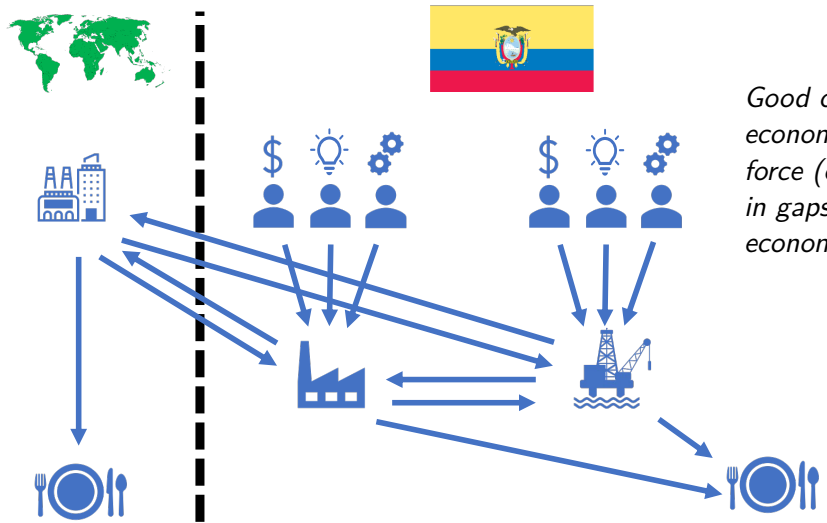
1. Social security
2. Ownership registry
3. VAT records
4. **Corp. income tax**
5. Customs records

5 Administrative Datasets on Economic Linkages



1. Social security
2. Ownership registry
3. VAT records
4. Corp. income tax
5. **Customs records**

5 Administrative Datasets on Economic Linkages



Good coverage of formal economy. But must use labor force (etc.) survey data to fill in gaps about informal economy.

How does trade affect within-country earnings inequality?

- Recall framework from yesterday:
  - Arbitrary tastes and technologies
  - Competitive factor markets (competitive goods markets no longer needed today)

- This implies

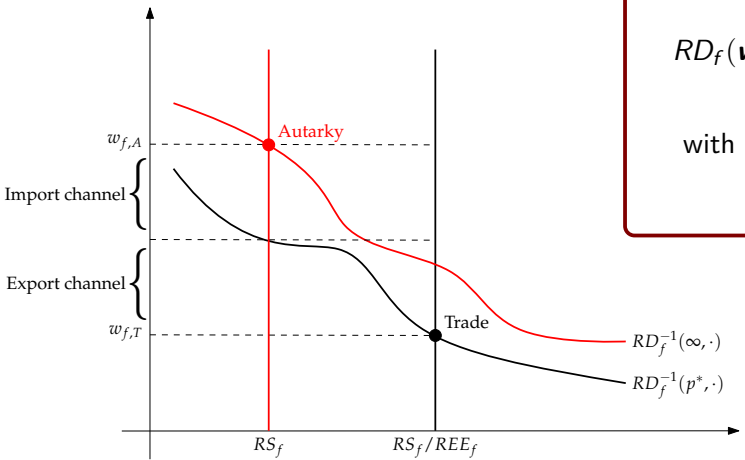
$$\sum_d L_{od}^f(\boldsymbol{\omega}_o) = \bar{L}_{of} \quad \text{for all } o \text{ and } f$$

- But given today's focus—studying one "Home" country and taking it to autarky—will express this as (where $FCE \equiv$ factor content of exports):

$$\underbrace{L_f(\boldsymbol{w}_T, \boldsymbol{p}^*)}_{\text{Home demand for } f} = \bar{L}_f - FCE_f \quad \text{for all } f \text{ in "Home" country}$$

$$\iff RD_f(\boldsymbol{w}_T, \boldsymbol{p}^*) = RS_f / REE_f \qquad \text{with} \qquad REE_f \equiv \frac{1 - \frac{FCE_0}{\bar{L}_0}}{1 - \frac{FCE_f}{\bar{L}_f}}$$

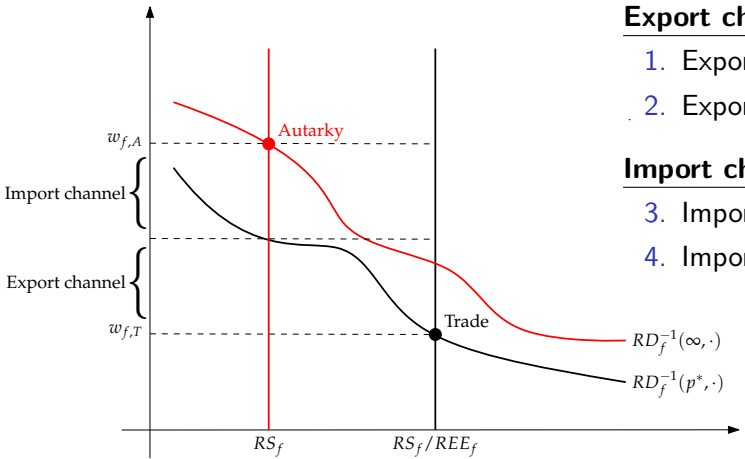
Exports, Imports, and Earnings Inequality



$$RD_f(\boldsymbol{w}_T, \boldsymbol{p}^*) = RS_f/REE_f$$

with $$REE_f \equiv \frac{1 - \frac{FCE_0}{L_0}}{1 - \frac{FCE_f}{L_f}}$$

Exports, Imports, and Earnings Inequality



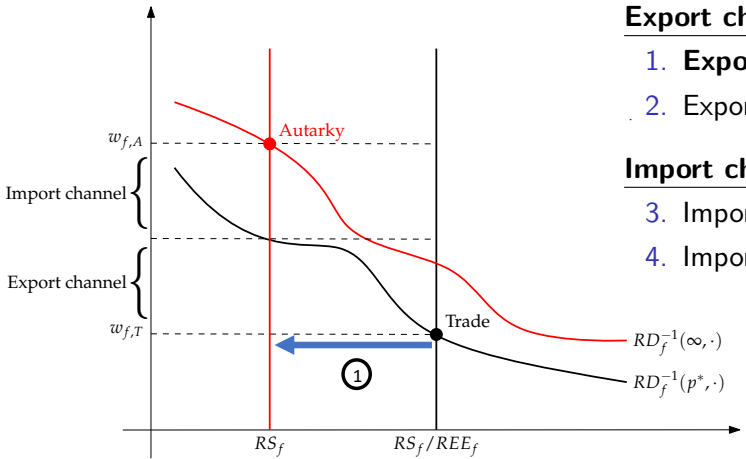
Export chanel:

1. Export exposure ($\boldsymbol{REE_f \to 1}$)
2. Export incidence

Import chanel:

3. Import exposure ($\boldsymbol{p^* \to \infty}$)
4. Import incidence

Exports, Imports, and Earnings Inequality



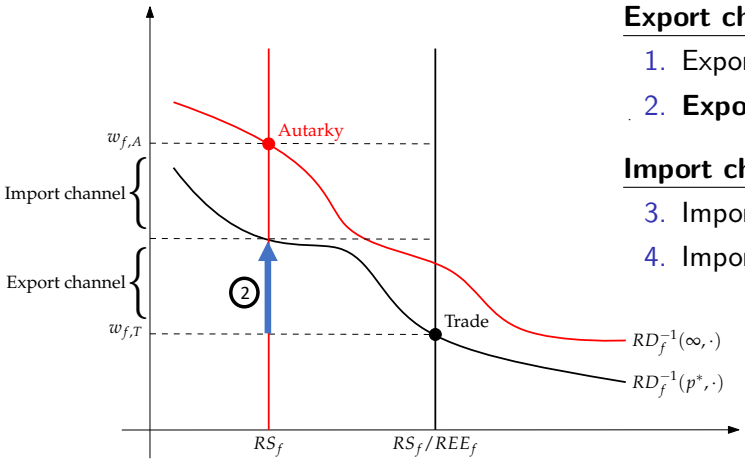
Export chanel:

1. **Export exposure ($REE_f \to 1$)**
2. Export incidence

Import chanel:

3. Import exposure ($p^* \to \infty$)
4. Import incidence

Exports, Imports, and Earnings Inequality



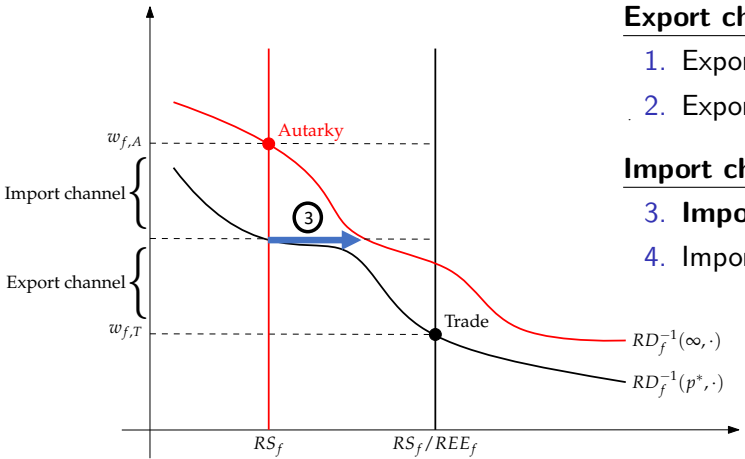
Export chanel:

1. Export exposure ($\boldsymbol{REE_f} \to \mathbf{1}$)
2. **Export incidence**

Import chanel:

3. Import exposure ($\boldsymbol{p^*} \to \infty$)
4. Import incidence

Exports, Imports, and Earnings Inequality



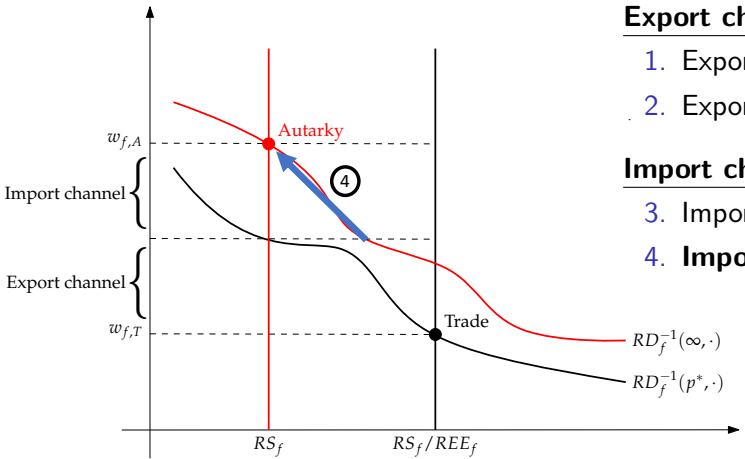
Export chanel:

1. Export exposure ($\boldsymbol{REE_f \to 1}$)
2. Export incidence

Import chanel:

3. **Import exposure ($\boldsymbol{p^* \to \infty}$)**
4. Import incidence

Exports, Imports, and Earnings Inequality



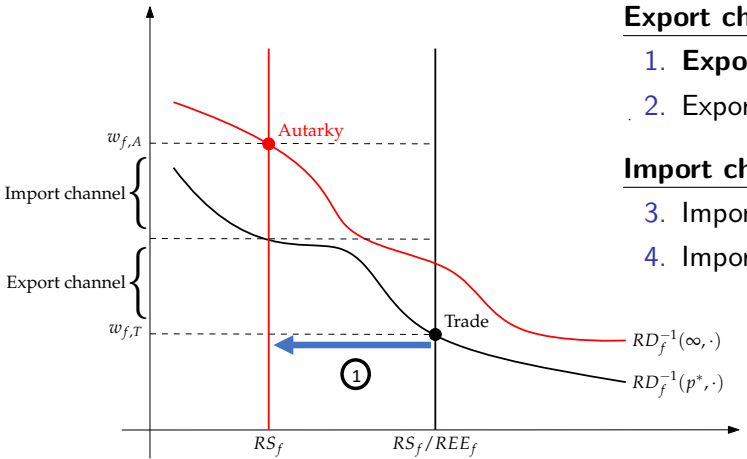
Export chanel:

1. Export exposure ($\boldsymbol{REE_f \to 1}$)
2. Export incidence

Import chanel:

3. Import exposure ($\boldsymbol{p^* \to \infty}$)
4. **Import incidence**

Step 1: Quantifying Export Exposure



Export chanel:

1. **Export exposure ($REE_f \to 1$)**
2. Export incidence

Import chanel:

3. Import exposure ($p^* \to \infty$)
4. Import incidence

How to Measure Export Exposure (EE)?

- Under assumption that multi-product firms use same input proportions in all products, can measure EE with administrative datasets

- This is the big payoff from **Strategy #1**—theory has suggested how to reduce what needs to be estimated (and admin data makes it possible)

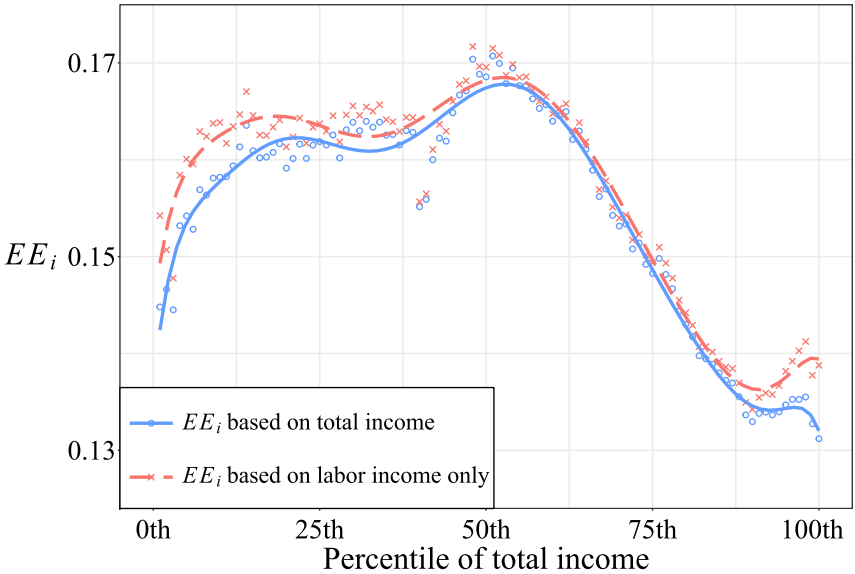
- Definition:

$$EE_f \equiv \frac{FCT_f}{\bar{L}_f} \equiv \frac{[(\text{factors hired in H}) \times (\text{Leontief inverse in H}) \times (\text{exports of H})]_f}{(\text{total earnings})_f}$$

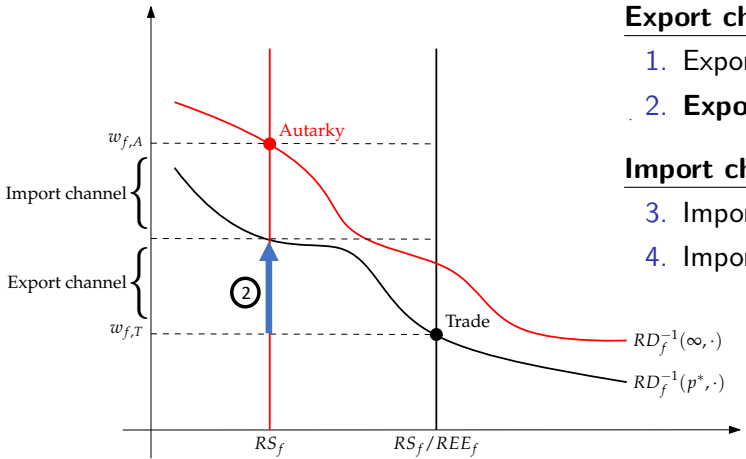
- Then extend to individuals i based on factor ownership shares ϕ_{fi}:

$$EE_i \equiv \sum_f \phi_{fi} EE_f$$

Export Exposure Across the Income Distribution



Step 2: Quantifying Export Incidence



Export chanel:

1. Export exposure ($\boldsymbol{REE_f} \to \mathbf{1}$)
2. **Export incidence**

Import chanel:

3. Import exposure ($\boldsymbol{p^*} \to \infty$)
4. Import incidence

How to Measure Export Incidence?

- This requires knowledge of $RD_f(\boldsymbol{w}, \boldsymbol{p}^*)$ as a function of $\boldsymbol{w}$. Build this up from model of Home economy with (in baseline case)...
- Home individuals i with nested CES preferences over Home sectors k and goods v:

$$u_i = \prod_{k \in \mathcal{K}} (u_{i,k})^{\alpha_k}$$

$$u_{i,k} = \left(\sum_{v \in \mathcal{V}_k} \theta_{vc}^{\frac{1}{\sigma}} q_{i,v}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

- Nested CES production functions for competitive domestic firms v:

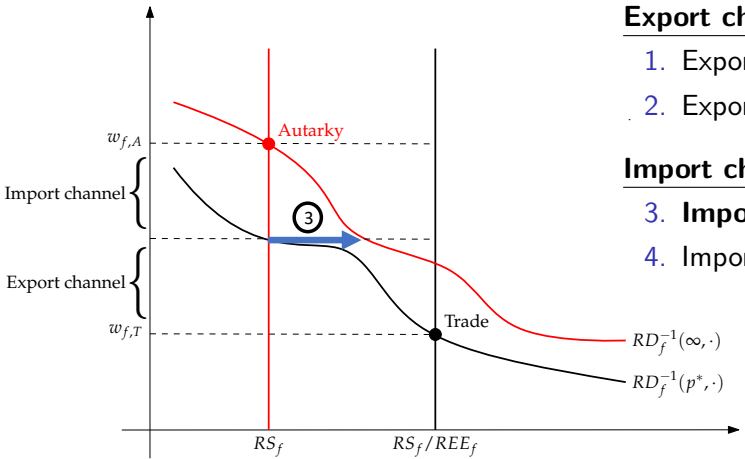
$$y_v = \varphi_v (l_v)^{\beta_v} (m_v)^{1-\beta_v}$$

$$l_v = \left(\sum_{f \in \mathcal{F}} \gamma_{fv}^{\frac{1}{\eta}} l_{fv}^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}}$$

$$m_v = \left[\prod_{r \in \mathcal{V}} (m_{rv})^{\gamma_{rv}} \right]^{\Gamma_v} \left[\prod_{r \in \mathcal{V}^*} (m_{rv})^{\gamma^*_{rv}} \right]^{1-\Gamma_v}$$

- Combine to yield a $RD_f(\boldsymbol{w}, \boldsymbol{p}^* | \sigma, \eta, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\varphi}, \boldsymbol{\Gamma})$, to be estimated.

Step 3: Quantifying Import Exposure



Export chanel:

1. Export exposure ($\boldsymbol{REE_f \to 1}$)
2. Export incidence

Import chanel:

3. **Import exposure ($\boldsymbol{p^* \to \infty}$)**
4. Import incidence

How to Measure Import Exposure (IE)?

- Shaped by size of demand shifter: $\sum_v \frac{\partial \ln RD_f}{\partial \ln p_v^*}$
- In above model, this takes the form (locally):

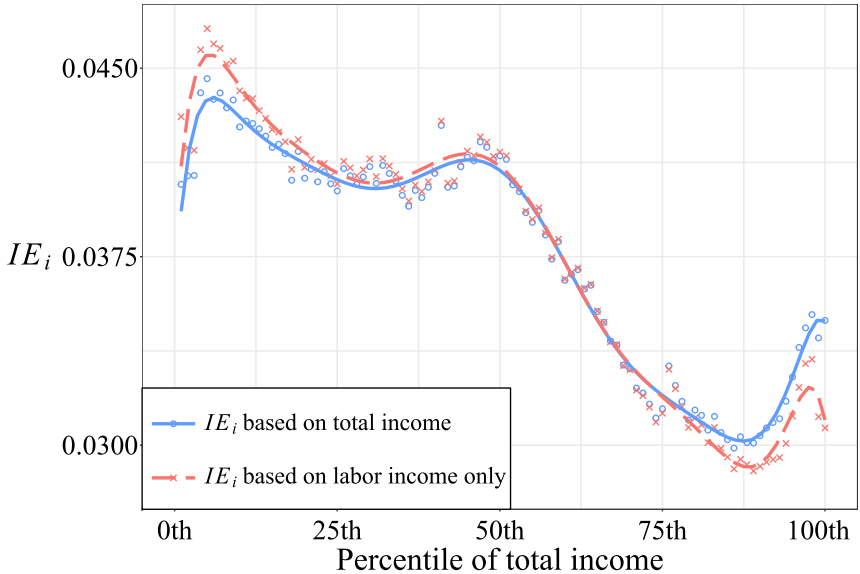
$$\sum_v \frac{\partial \ln RD_f}{\partial \ln p_v^*} = (\sigma - 1)(IE_f - IE_0)$$

$$\text{with:} \quad IE_f \equiv -\sum_{k \in \mathcal{K}} \sum_{m \in \mathcal{V}_k} s_{fm} \left(\bar{x}_m^* - \bar{\bar{x}}_k^*\right)$$

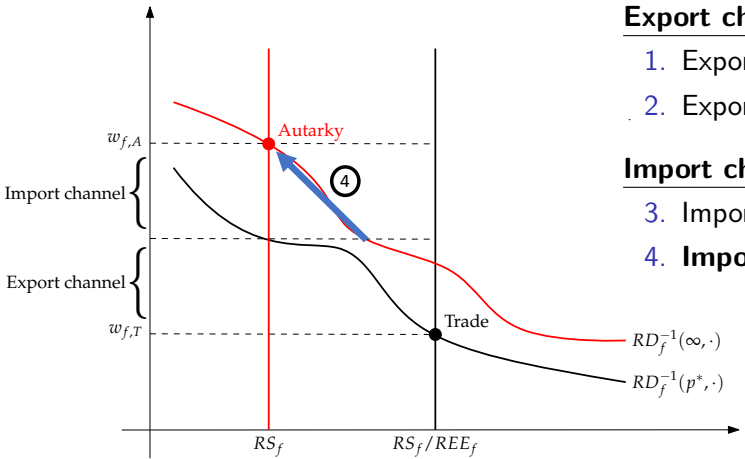
- Using:
  - direct+indirect $s_{fm} \equiv$ share of factor f in firm m's final sales
  - direct+indirect $\bar{x}_m^* \equiv$ share of firm m's costs spent on imports
  - $\bar{\bar{x}}_k^* \equiv$ weighted equivalent of $\bar{x}_m^*$ for industry k
- And again extend to individuals i based on factor ownership shares ϕ_{fi}:

$$IE_i \equiv \sum_f \phi_{fi} IE_f$$

Import Exposure (Locally) Across the Income Distribution



Step 4: Quantifying Import Incidence



Export chanel:

1. Export exposure ($\boldsymbol{REE_f \to 1}$)
2. Export incidence

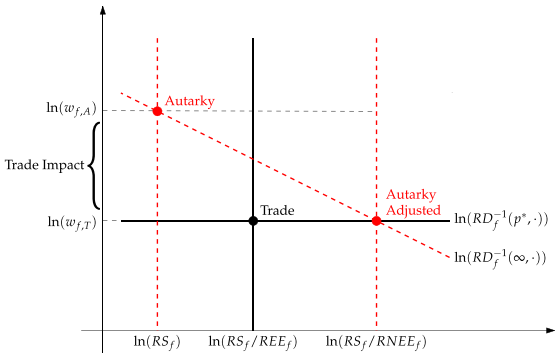
Import chanel:

3. Import exposure ($\boldsymbol{p^* \to \infty}$)
4. **Import incidence**

How to Measure Import Incidence?

- This requires knowledge of $RD_f(\boldsymbol{w}, \infty)$ as function of $\boldsymbol{w}$.
- Compute it with knowledge of estimated $RD_f(\boldsymbol{w}, \boldsymbol{p}^* | \sigma, \eta, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\varphi}, \boldsymbol{\Gamma})$, as with export incidence but evaluated at $\boldsymbol{p}^* = \infty$

Digression: Comparison to Deardorff-Staiger (1988) Approach



- If all imported goods are produced at home (i.e. traditional H-O model, inside FPE set) then $RD_f(\boldsymbol{p}^*,\cdot)$ is perfectly elastic
- Deardorff (2000): if all preferences and technology are common-elasticity CES (η_{agg}) then $(\Delta \ln w)_{\text{trade}} = \frac{\ln RNEE}{\eta_{\text{agg}}}$
- But much debate ... e.g. Wood (1995); and Krugman (2000) vs. Leamer (2000)

Estimating elasticities (η, σ)

- Now have to take a stand on what a "factor" is. Baseline assumption will be:
  - Labor: region $\times$ education groups (26 provinces $\times$ 3 ed. levels)
  - Capital: firm profits in oil and non-oil sectors

- η: standard within-firm CES factor demand estimation
- σ: standard across-firm CES consumer demand estimation

- IVs: shift-share structure based on composition of firms' direct+indirect imports and exports (i.e. factor-level analogs of Aghion et al, 2018 and Amiti et al, 2016):

$$E^{IV}_{f,t} \equiv \sum_{v \in \mathcal{HS}(6)} EE_{fv,t_0} \times (\text{Export Demand Shifter})_{v,t}$$

$$I^{IV}_{f,t} \equiv \sum_{v \in \mathcal{HS}(6)} IE_{fv,t_0} \times (\text{Import Price Shifter})_{v,t}$$

Estimates of elasticity of substitution across factors (η)

	OLS	2SLS	Alternative 2SLS Specifications			
	(1)	(2)	(3)	(4)	(5)	(6)
Estimate of η	1.34 (0.19)	2.09 (0.34)	2.12 (0.47)	2.02 (0.73)	2.05 (0.33)	2.30 (0.49)
First-stage F statistic	-	10.0	19.2	3.0	10.5	12.6
IV construction:	-	Export and import IVs	Export IV only	Import IV only	$(\text{Shock})_{vt}$ from large countries only	$(\text{Shock})_{vt}$ de-meaned by small countries

Notes: Sample of incorporated firms with positive payments for more than one factor and more than one employee. All specifications use a balanced panel of 627,399 factor-firm-year observations from 2009-2015, include firm-year and factor fixed effects, and include controls for year fixed effects interacted with factor exposure at t_0 to exports and imports. Observations weighted by initial factor-firm payments (winsorized at the 95th percentile). Standard errors in parentheses are clustered by factor (of which there are 75).

Estimates of elasticity of substitution across firms (σ)

	OLS	2SLS	Alternative 2SLS Specifications			
	(1)	(2)	(3)	(4)	(5)	(6)
Estimate of σ	1.04 (0.04)	2.03 (0.58)	1.71 (0.20)	1.99 (0.58)	1.95 (0.57)	2.31 (0.74)
First-stage F statistic	-	13.2	1.3	19.4	14.5	7.9
IV construction:	-	Export and import IVs	Export IV only	Import IV only	$(\text{Shock})_{vt}$ from large countries only	$(\text{Shock})_{vt}$ de-meaned by small countries

Notes: Sample of incorporated firms with positive final sales and more than one employee. All specifications use a balanced panel of 181,804 firm-year observations from 2009-2015, include firm and sector-year fixed effects, and include controls for year fixed effects interacted with firm cost shares spent on primary factors and imports. Observations are weighted by initial firm final sales (weights winsorized at the 95 percentile). Standard errors in parentheses are clustered by firm (of which there are 25,972).

Should You Believe Any of This?

- The model of Ecuador's factor price determination proposed and estimated here is clearly a preposterous abstraction
  - Market structure(s)? Functional forms? Static? Factor definitions/supply?
- So why would you believe the counterfactual exercise that follows?
- Recall notation introduced yesterday:

$$\text{True model: } y_{n,t} = g_n^*(\tau_t, \epsilon_t^*) \qquad \text{Researcher's model: } y_{n,t} = g_n(\tau_t, \epsilon_t)$$

$$\Delta x_n^* \equiv g_n^*(\tau_{t+1}, \epsilon_{t+1}^*) - g_n^*(\tau_t, \epsilon_{t+1}^*) \qquad \Delta x_n \equiv g_n(\tau_{t+1}, \epsilon_{t+1}) - g_n(\tau_t, \epsilon_{t+1})$$

- Why should you believe that $\Delta x_n^* \approx \Delta x_n$ for all relevant outcomes n?
- Can we deploy **Strategy #2 ("Guess and verify")** to help?

A Simple Version of "Guess and Verify" (More Tomorrow!)

- Begin with simple identity:

$$\Delta y_n = g_n^*(\tau_{t+1}, \epsilon_{t+1}^*) - g_n^*(\tau_t, \epsilon_t^*) = \Delta x_n^* + \Delta \eta_n^*$$

  where $\Delta \eta_n^* \equiv g_n^*(\tau_t, \epsilon_{t+1}^*) - g_n^*(\tau_t, \epsilon_t^*)$ denotes the causal impact of the other shocks

- And suppose we have some "instrument" z that satisfies $z \perp\!\!\!\perp \epsilon_{t+1}^* | (\epsilon_t^*, \tau_t)$

- Then can use *IV-based test statistic*

$$\hat{\beta}_z \equiv \frac{1}{N} \sum\nolimits_n z_n (\Delta y_n - \Delta x_n)$$

  and then note that $\Delta x_n^* = \Delta x_n$ implies $E_t[\hat{\beta}_z] = 0$.

- Equivalent to IV regression of Δy_n on Δx_n using z_n as IV (and test coeff=1)

Results of IV-Based Test

TABLE II

GOODNESS OF FIT TESTS

	Δ Log of observed factor price				
	(1)	(2)	(3)	(4)	(5)
Δ Log of predicted factor price	1.10	1.61	1.26	1.04	0.89
	(0.15)	(0.62)	(0.62)	(0.16)	(0.20)
p-value (H_0: $\beta_{\text{fit}} = 1$)	[.53]	[.33]	[.68]	[.79]	[.58]
First-stage F-statistic	2,103.9	205.0	189.6	304.7	125.9

Notes. All specifications use a balanced panel of 525 factor-year observations from 2009–2015 and are estimated with year and factor fixed effects. Columns (2)–(5) add, cumulatively, controls for interactions between year indicators and: column (2) EE_{f,t_0} and IE_{f,t_0}; column (3) capital factor indicators; column (4) province indicators; and column (5) education-level indicators. Observations are weighted by initial factor payments (winsorized at the 95th percentile). Standard errors in parentheses are clustered by factor (of which there are 75).

Distribution (along Pre-Shock Y) of Gains from Trade in Ecuador

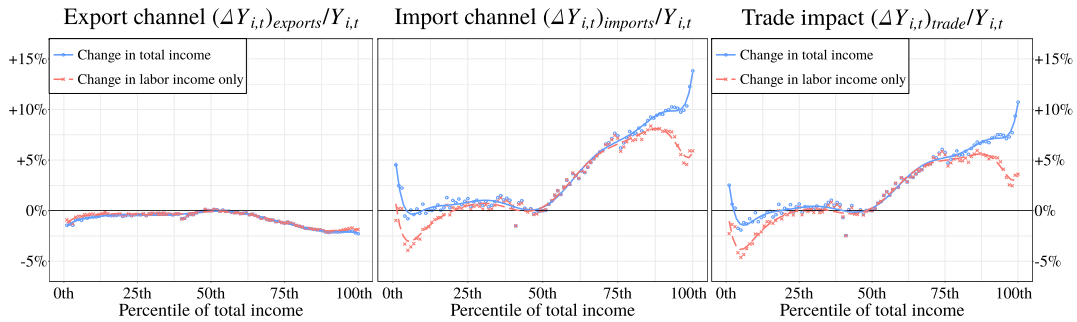
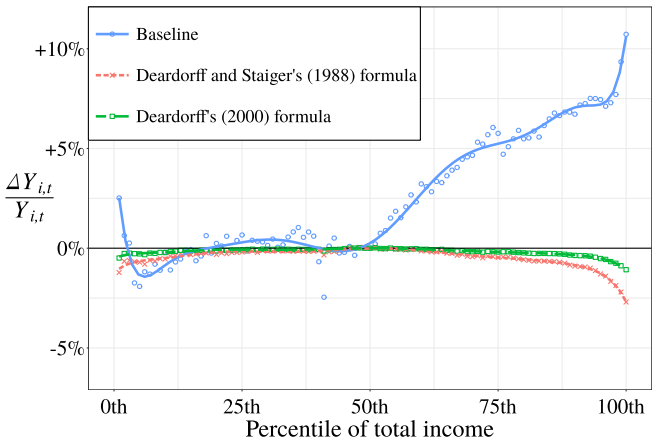


FIGURE IV

Trade and Earnings Inequality, Baseline

Blue circles correspond to the total (including both labor and capital) income change for each individual, averaged within each percentile and normalized to zero at the median percentile, between 2012 and the counterfactual autarkic equilibrium. Positive numbers therefore reflect larger gains from trade than at the median. Red crosses do the same but for labor income only. Lines indicate fitted 10th-order polynomials. Trade impact is the sum of the export and import channels. All changes are expressed as percentages.

Comparison to Deardorff-Staiger (1988) Approach

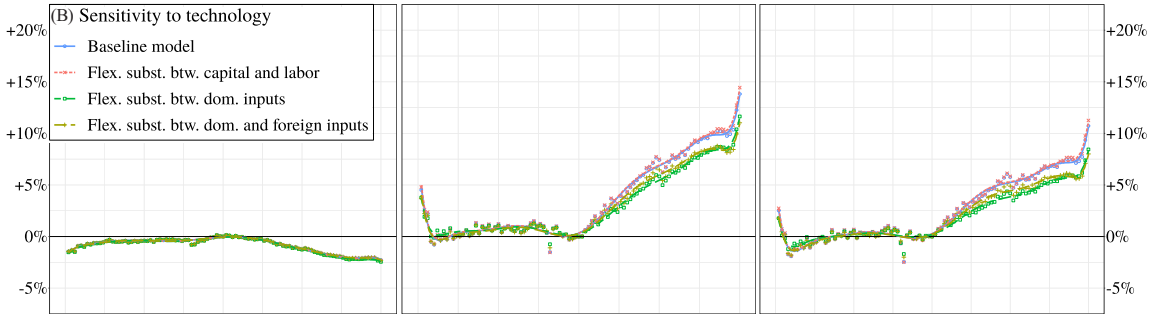


- Trefler (1995): for most countries and factors, measures of *net* factor content of trade are very small
- In our model, gross factor exports and imports play distinct roles (hence small size of net FCT is not necessarily relevant)

Sensitivity Analysis (Alternative Technologies)

	Technology				Preferences				Factors	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Parameter	η_L	η	μ	ε	σ_1	σ_2	σ_3	σ_4	η	η
Elasticity of substitution between	Labor types	Labor vs. capital	Domestic inputs	Domestic vs. foreign inputs	Tradables sector firms	Retail & Wholesale sector firms	RE & Construc. sector firms	Other Services sector firms	College vs. non-college labor	HS vs. non-HS labor
Estimating equation	(C.4)	(C.5)	(C.8)	(C.10)	(C.12)	(C.12)	(C.12)	(C.12)	(24)	(24)
Estimate	3.15	1.29	1.61	1.08	2.09	1.32	2.03	1.78	1.96	2.06
	(0.69)	(0.95)	(0.54)	(0.27)	(0.97)	(0.57)	(2.18)	(0.69)	(0.39)	(0.33)
First-stage F-statistic	4.6	128.6	7.9	103.5	5.6	12.9	0.9	3.3	14.0	9.9
Number of observations	462,487	44,751	1,527,462	17,878	25,886	83,377	30,800	39,312	485,070	447,299
Number of clusters	73	6,393	33,631	2,554	3,698	11,911	4,400	5,616	51	51
SE cluster level	factor	firm	seller	firm	firm	firm	firm	firm	factor	factor

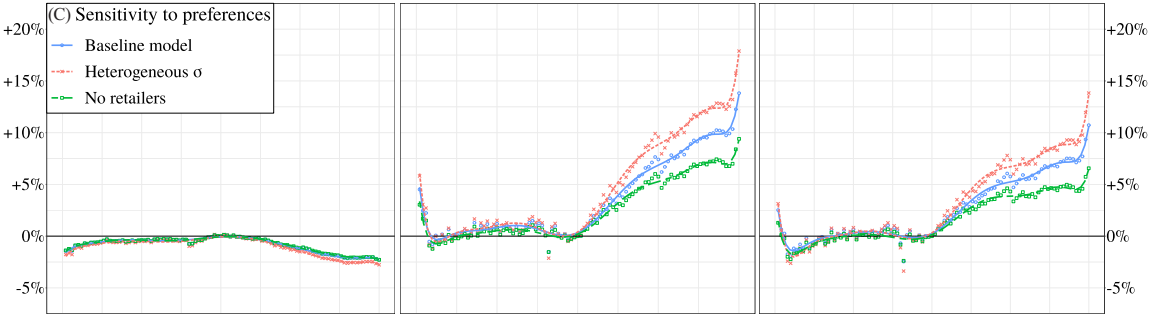
Sensitivity Analysis (Alternative Technologies)



Sensitivity Analysis (Alternative Preferences + Retailing)

	Technology				Preferences				Factors	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Parameter	η_L	η	μ	ε	σ_1	σ_2	σ_3	σ_4	η	η
Elasticity of substitution between	Labor types	Labor vs. capital	Domestic inputs	Domestic vs. foreign inputs	Tradables sector firms	Retail & Wholesale sector firms	RE & Construc. sector firms	Other Services sector firms	College vs. non-college labor	HS vs. non-HS labor
Estimating equation	(C.4)	(C.5)	(C.8)	(C.10)	(C.12)	(C.12)	(C.12)	(C.12)	(24)	(24)
Estimate	3.15	1.29	1.61	1.08	2.09	1.32	2.03	1.78	1.96	2.06
	(0.69)	(0.95)	(0.54)	(0.27)	(0.97)	(0.57)	(2.18)	(0.69)	(0.39)	(0.33)
First-stage F-statistic	4.6	128.6	7.9	103.5	5.6	12.9	0.9	3.3	14.0	9.9
Number of observations	462,487	44,751	1,527,462	17,878	25,886	83,377	30,800	39,312	485,070	447,299
Number of clusters	73	6,393	33,631	2,554	3,698	11,911	4,400	5,616	51	51
SE cluster level	factor	firm	seller	firm	firm	firm	firm	firm	factor	factor

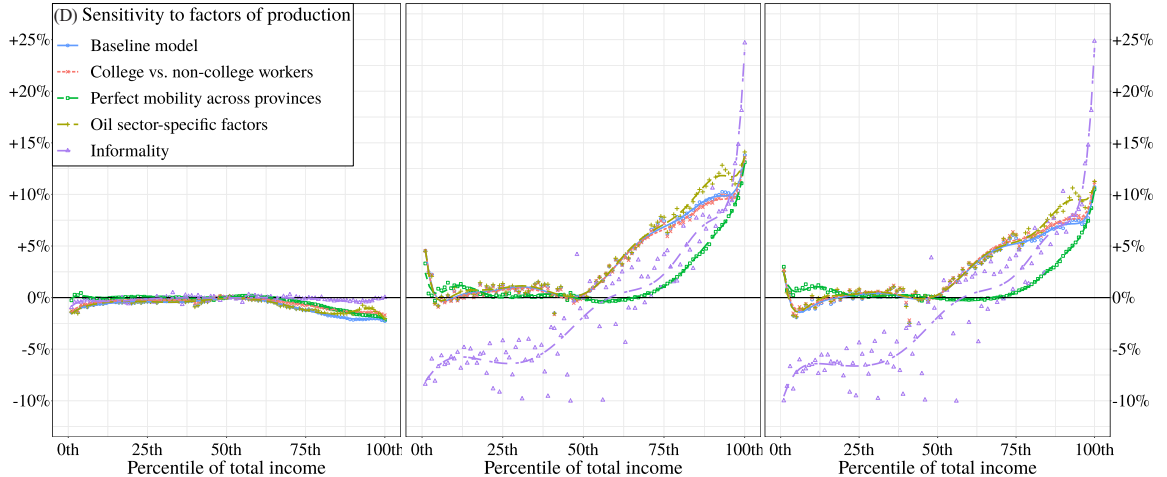
Sensitivity Analysis (Alternative Preferences + Retailing)



Sensitivity Analysis (Alternative Factors)

	Technology				Preferences				Factors	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Parameter	η_L	η	μ	ε	σ_1	σ_2	σ_3	σ_4	η	η
Elasticity of substitution between	Labor types	Labor vs. capital	Domestic inputs	Domestic vs. foreign inputs	Tradables sector firms	Retail & Wholesale sector firms	RE & Construc. sector firms	Other Services sector firms	College vs. non-college labor	HS vs. non-HS labor
Estimating equation	(C.4)	(C.5)	(C.8)	(C.10)	(C.12)	(C.12)	(C.12)	(C.12)	(24)	(24)
Estimate	3.15	1.29	1.61	1.08	2.09	1.32	2.03	1.78	1.96	2.06
	(0.69)	(0.95)	(0.54)	(0.27)	(0.97)	(0.57)	(2.18)	(0.69)	(0.39)	(0.33)
First-stage F-statistic	4.6	128.6	7.9	103.5	5.6	12.9	0.9	3.3	14.0	9.9
Number of observations	462,487	44,751	1,527,462	17,878	25,886	83,377	30,800	39,312	485,070	447,299
Number of clusters	73	6,393	33,631	2,554	3,698	11,911	4,400	5,616	51	51
SE cluster level	factor	firm	seller	firm	firm	firm	firm	firm	factor	factor

Sensitivity Analysis (Alternative Factors)



Concluding Remarks

- For many important questions, structural estimation is necessary. But audience skepticism is severe!
- How can researchers make structural estimation more credible?
- **Today's lecture:**
  - Strategy #1 (Reduce what needs to be estimated) + administrative data helped on export-exposure side of problem
  - Strategy #2 (Guess and verify) applied loosely
- **Tomorrow's lecture:**
  - Change the question: did Trump win his (2018) trade war?
  - Strengthen Strategy #2: connecting how we test models to the question being asked

Thank You!