Universidad de los Andes Colombia | Facultad de Economía

Distributed microdata analysis for Latin America

Marcela Eslava - Universidad de Los Andes

The problem of distributed microdata analysis

- The distribution of productivity across establishments within subsectors, and across subsectors within sectors, is a key tool for productivity diagnosis.
  - Especially important for developing economies
- Comparative analysis across countries frequently very useful.
- But access to necessary microdata subject to strict protocols at statistical offices. The analysis for different countries cannot be carried by same researcher in same place.
- Distributed microdata analysis (Bartelsman, Haltiwanger and Scarpetta, 2009, 2013): tool to overcome these difficulties

Distributed microdata analysis (DMA)

- DMA proposes unified protocols for data processing that is carried independently by different researchers using different datasets from different facilities.
- Central team:
  - gathers information on the datasets for different countries
  - designs protocols to make data comparable (inclusion thresholds, data cleaning)
  - distributes codes that execute those protocols and generate comparable statistics
- Country teams execute those codes and generates, e.g., by sector datasets with moments of the distribution of productivity
  - First and higher order univariate moments
  - Multivariate moments (e.g. size-productivity covariances)

DMA in Latin America

- Manufacturing surveys, many census-like, in at least Brazil, Chile, Colombia, Ecuador, Mexico, Uruguay.
  - Annual
  - Changes in methodology over time, but relatively long series
  - 10 employees or more the most frequent inclusion criteria
  - Sales, employees, capital stocks (book values)…
- Ecuador, Colombia also have surveys of services
  - Inclusion criteria more restrictive and coverage of only some subsectors
- Some have open access microdata but only for “anonymized” data and some variables, others fully confidential do not have

Distributed microdata analysis

- The distribution of productivity across establishments within subsectors, and across subsectors within sectors, is a key tool for productivity diagnosis.
- Comparative analysis across countries frequently very useful.
- But characterizing the productivity distribution requires access to microdata that is frequently confidential and can only be accessed under strict protocols at statistical offices. The analysis for different countries cannot be carried by same researcher in same place.
- Distributed microdata analysis (Bartelsman, Haltiwanger and Scarpetta, 2009, 2013): tool to overcome these difficulties

Two examples of MDA

- CAF's Report on Economic Development, 2018: Institutions for Productivity
  - Manufacturing: Colombia, Chile, Mexico, Uruguay
  - Services: Colombia
  - The "anatomy" of productivity
  - Proxy causes of productivity gap vs. the YS
  - Eslava, 2018

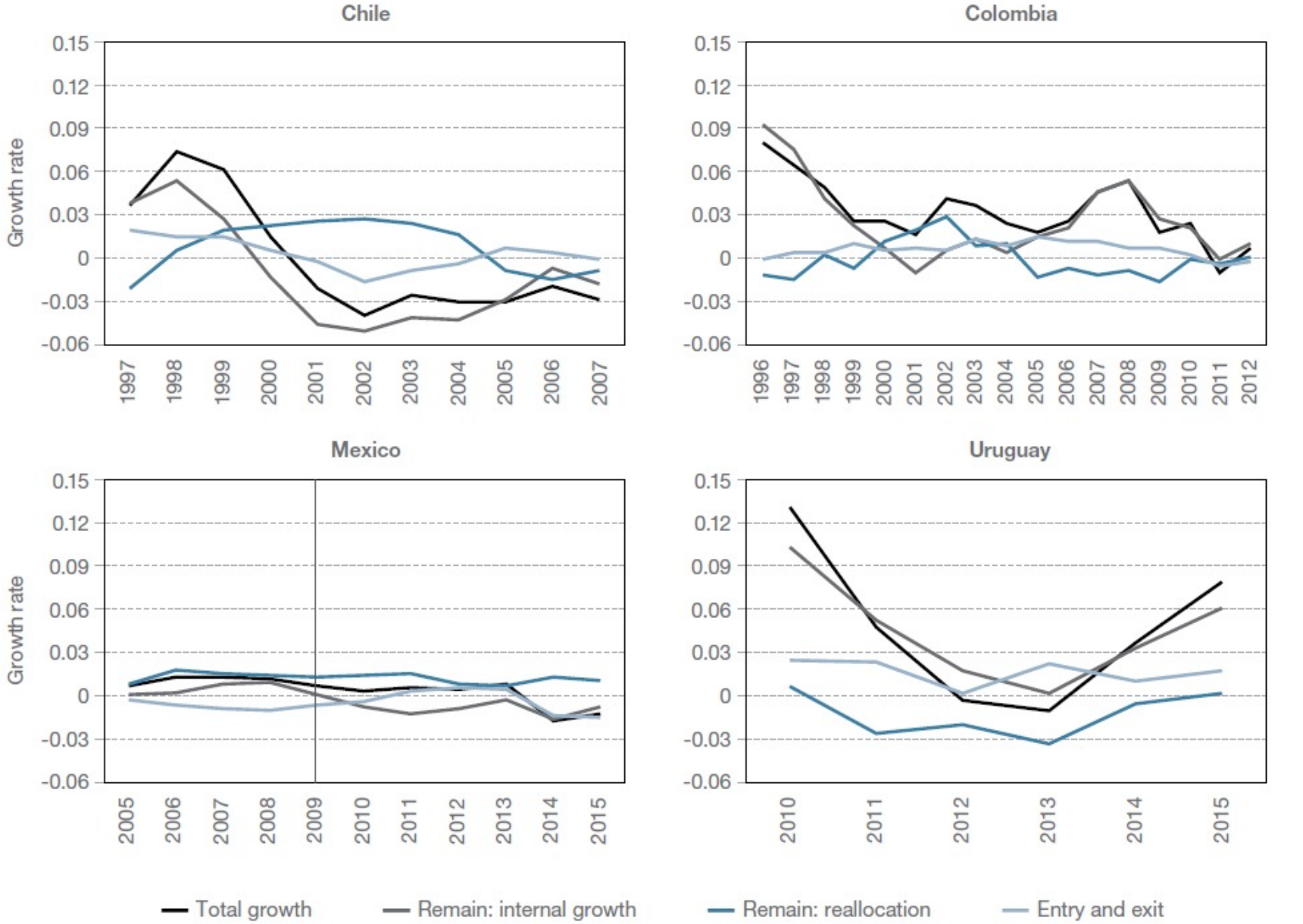
- Background paper for UNDP's annual report: "Market concentration, market fragmentation and inequality in Latin America"
  - Manufacturing: Colombia, Chile, Ecuador, Mexico, Uruguay
  - Markups, market concentration, earnings inequality
  - Eslava, Meléndez, Urdaneta, 2021

Anatomy of productivity

The HFS decompos.:

Within component dominates time series

Graph 2.9 Decomposition of output per worker growth in the average manufacturing subsector



The Olley-Pakes decomposition

Table 2.2 Components of manufacturing productivity, 2003-2007

	Allocative efficiency among subsectors	Allocative efficiency among establishments
Output per worker		
Chile	0.17	0.34
Colombia	0.02	0.42
Mexico	0.12	0.26
United States[a/b/]	-0.03	0.51

Explaining the gap

	Output per worker		
	1950	1980	2010
	Latin America relative to the USA (PPP)		
Agriculture	0.37	0.29	0.21
Mining	0.37	0.53	0.50
Manufacturing	0.61	0.62	0.34
Electricity, gas, water	0.45	0.38	0.36
Construction	0.21	0.27	0.37
Commerce services	1.29	0.89	0.29
Transport services	0.66	0.52	0.39
Financial services	0.55	0.46	0.19
Government services	0.35	0.43	0.40
Personal services	0.33	0.33	0.28

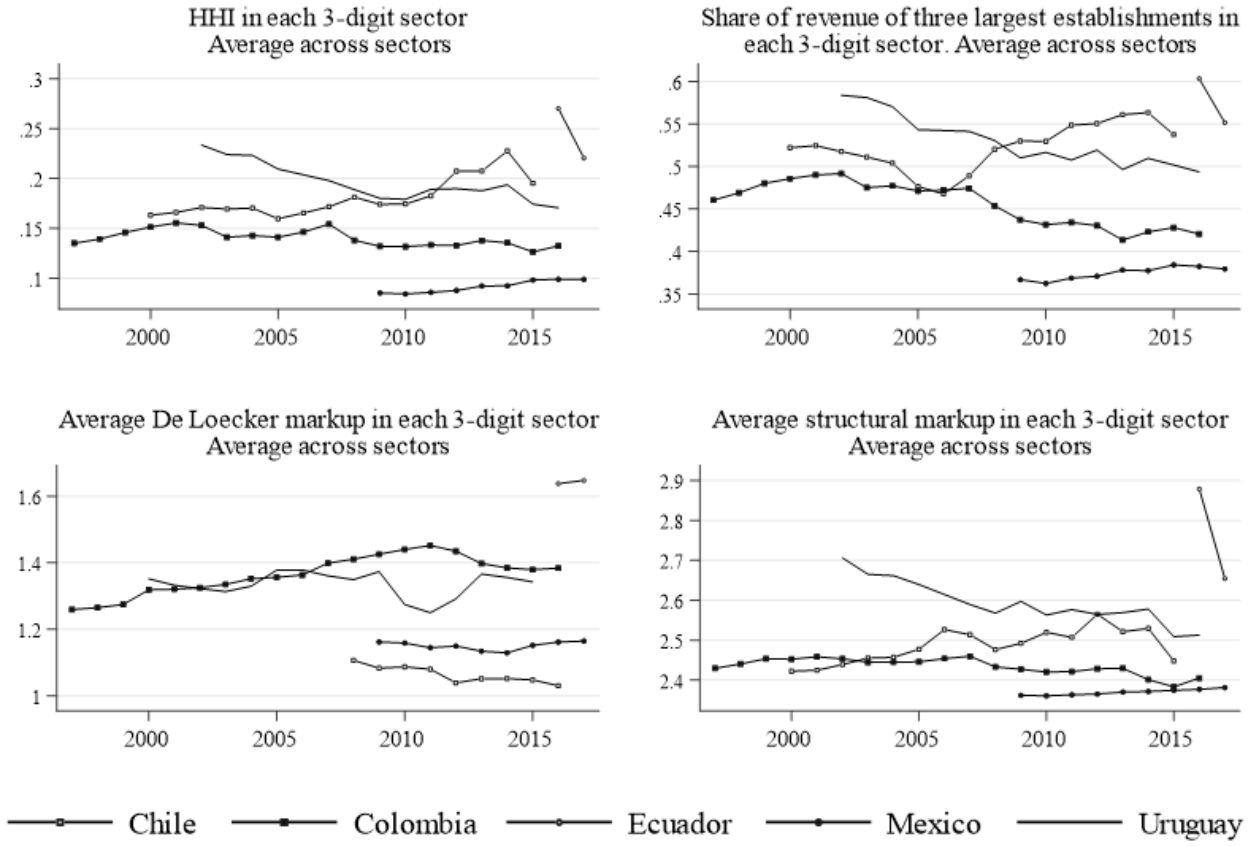
$$\frac{Total_{LA}}{Total_{US}} = 0.342 = \frac{Internal_{LA}}{Internal_{US}} \frac{1.41}{1.66} \frac{1.11}{0.97}.$$

$$\frac{Internal_{LA}}{Internal_{US}} = 0.354$$

Source: Produced by the author using data from the GGDC 10-Sector Database (Timmer, de Vries, & de Vries, 2015).

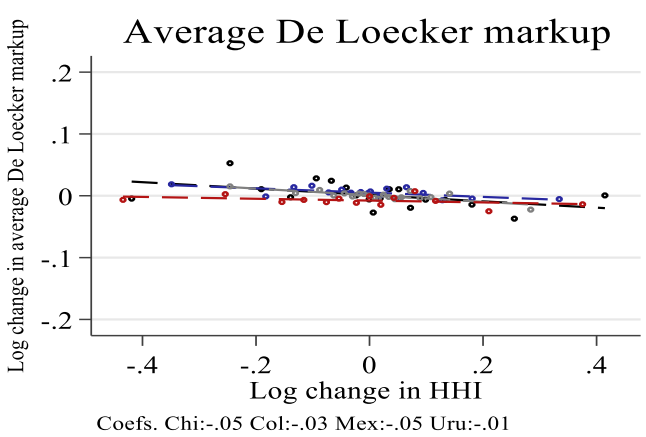
Concentration, market power and inequality (ongoing)

No clear increase in concentration or average markup in Latin America in recent years

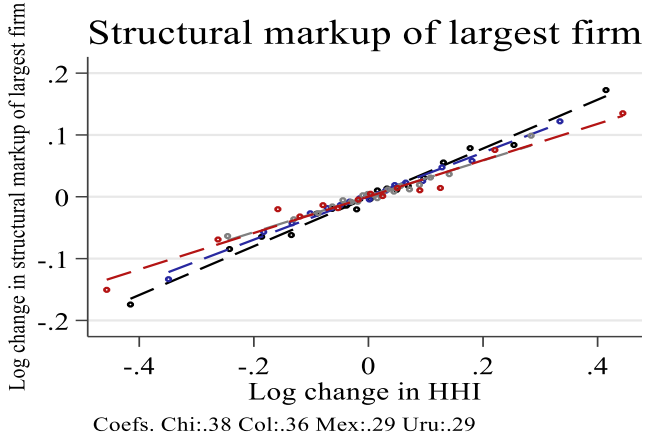


But in sectors where that increase has occurred…

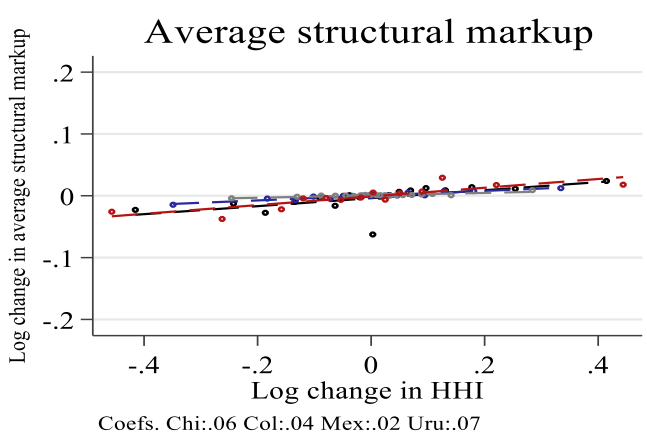
- It has been associated with higher markup of largest firm and decrease in labor share in that firm
- Without affecting the average firm



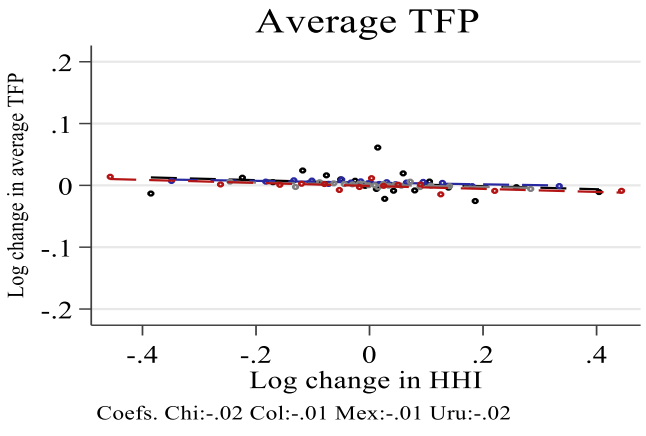
Coefs. Chi:-.05 Col:-.03 Mex:-.05 Uru:-.01



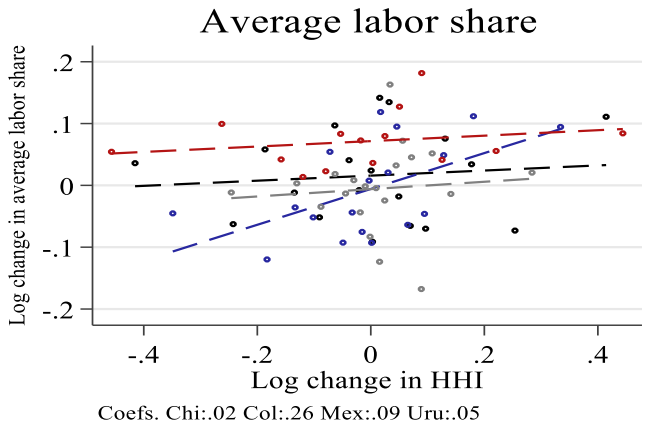
Coefs. Chi:.38 Col:.36 Mex:.29 Uru:.29



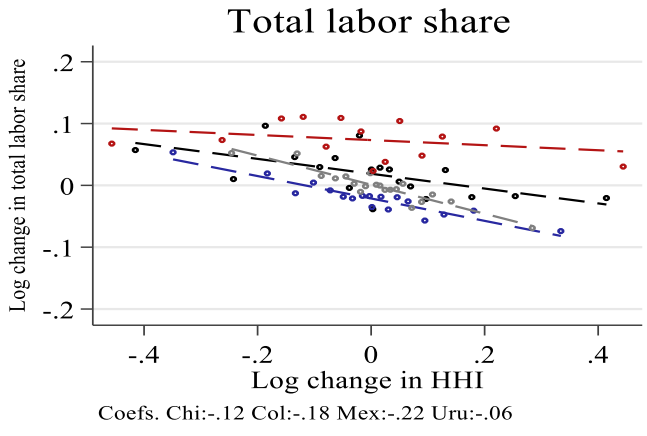
Coefs. Chi:.06 Col:.04 Mex:.02 Uru:.07



Coefs. Chi:-.02 Col:-.01 Mex:-.01 Uru:-.02



Coefs. Chi:.02 Col:.26 Mex:.09 Uru:.05



Coefs. Chi:-.12 Col:-.18 Mex:-.22 Uru:-.06

Chile Colombia Mexico Uruguay

As a conclusion…

- DMA possible in Latin America
  - In some cases with great richness (e.g. establishment prices in Colombia and Chile)
- Key insights may emerge
- But challenges in access are great and uneven across countries
  - Depends on interest by researchers
  - Difficult to engage statistical offices and do quality checks and follow ups
  - For instance, it has been imposible to include Brazil’s