



ANNUAL CONFERENCE
GLOBAL FORUM ON
PRODUCTIVITY
C O S T A R I C A

Aggregate Productivity And Microdata: Theory, Measurement And Applications

Setting the scene for the roundtable “The Importance of Microdata”

Federico Huneus (Duke University & Central Bank of Chile) · GFP Annual Conference, San José, 5 October 2026

How to Aggregate: a Foundational Question in Macroeconomics

1

The origins

A problem going back to the very beginnings of economics (Quesnay, 1758)

2

Derailed in the 1970s

- (i) “Negative” theoretical results
- (ii) Lack of disaggregated data

3

Recent explosion

- (i) Building on new theories from the 80s/90s
- (ii) Explosion of disaggregated data

4

⇒ The new problem

Need flexible aggregation theories that accommodate rich data
Otherwise, impossible to blend rich new data

TWO SIDES OF THE SAME COIN

Aggregation

Propagation

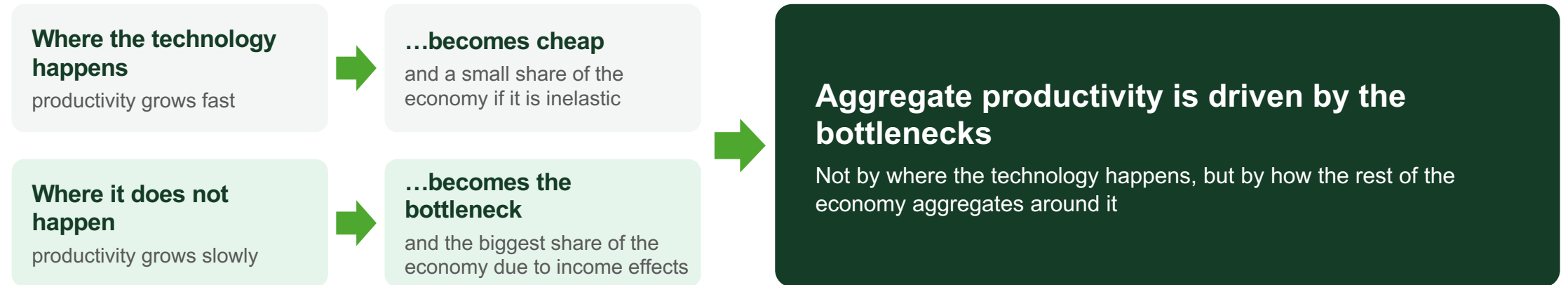
Aggregation: how changes at the firm or sector level add up to aggregate productivity

Propagation: how a change at one firm or sector travels to the rest of the economy

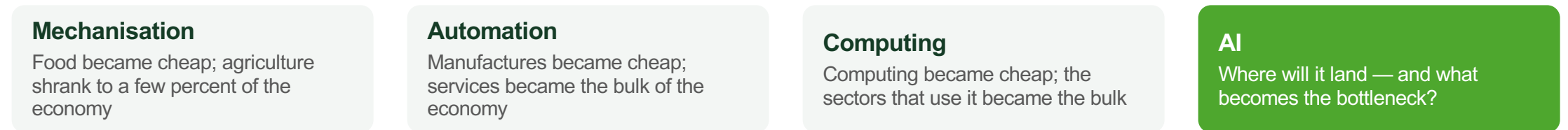
Once you know how to aggregate, you also know how to propagate

Why Aggregation Matters: AI And The Baumol Bottleneck

THE BAUMOL MECHANISM



EVERY MAJOR TECHNOLOGICAL DISRUPTION HAS HAD THE SAME PATTERN



Why aggregation matters: to know what AI does to aggregate productivity, need to know how aggregation (and substitution) occurs — across firms, sectors, places and the network between them...for that, one needs data, theory and measurement

Currently doing a project on this with OpenAI



Three Applications

Three Applications Within Chile's Data Infrastructure

Firm-to-firm electronic invoices

Every domestic transaction: product, price, quantity, location of both parties. ~2.1 billion transactions and 36 million buyer–supplier pairs in a year

Customs: transaction-level trade

Imports and exports by firm, product, partner and foreign counterpart (5.3 million transactions in a year)

Firm tax returns

Monthly VAT form (sales, inputs, investment, trade) and annual balance sheets for the universe of formal firms, 626 industries

Employer–employee records

Wages and headcount for ~703,000 firms and 8.2 million workers

Firm-to-consumer e-receipts

Purchases of 13.5 million identified consumers from the largest retailers (6 billion transactions)

Ownership, income tax and transfers

1.8 million owners of 1.4 million firms; income and public transfers for 8 million individuals

WHAT MAKES IT WORK

One identifier. Every dataset is linked by the firm or person tax ID

Universe, not sample. All formal firms, no size threshold; 85% of value added in the analysis samples

One home. Anonymised and hosted at the Central Bank of Chile under strict confidentiality agreements with the tax authority (SII), customs and others

Anchored to the aggregates. Reconciled with national and regional accounts and input–output tables, so micro and macro add up

Long panel. Firm data since 2005; transaction-level invoices since 2014

The Geographic And Sectoral Impact of Productivity

Huneeus & Del Valle (2026)

QUESTION

When productivity changes are local (space and sectors), where does the effect travel and through which mechanisms? — and how large is it for the country?

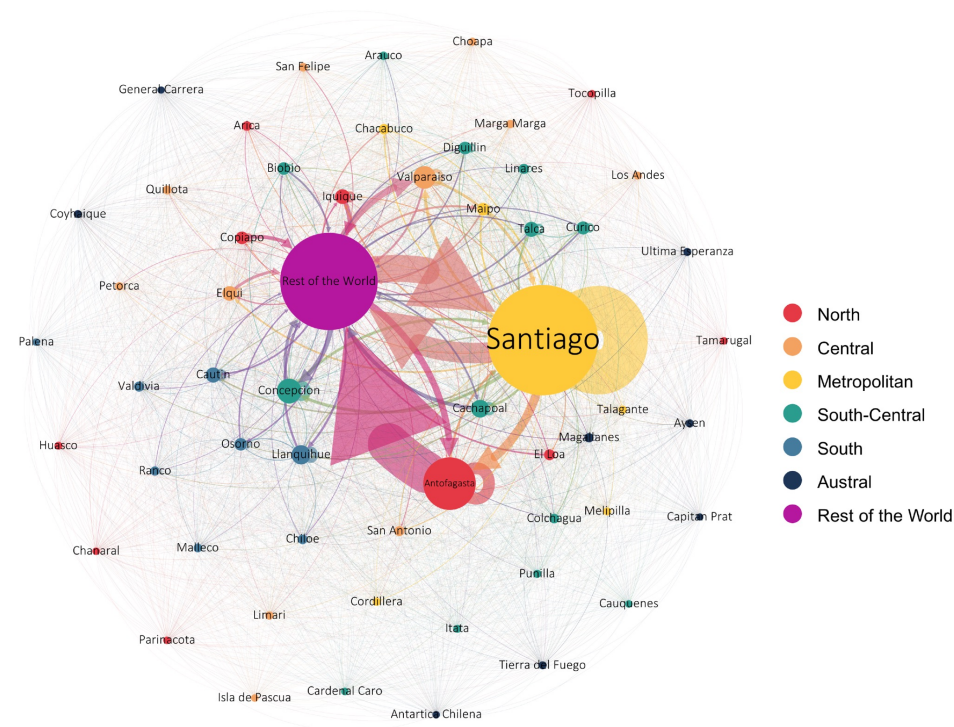
MICRODATA

- Firm-to-firm invoices with the location of both parties: 124 million transactions, 59 million firm pairs, 677,000 firms (2018–2023)
- Linked to firm tax returns, balance sheets and employer–employee records
- Reconciled with regional GDP and the national input–output table

WHAT WE BUILT: THE CED, EMBEDDED IN A MODEL

Cuentas Económicas Desagregadas (CED) — the Disaggregated Economic Accounts of Chile: the first within-country geographic input–output matrix, **56 provinces × 46 sectors = 2,576 local markets**. A free public good with CORFO and the IDB, updated annually (dataterritorios.corfo.cl)

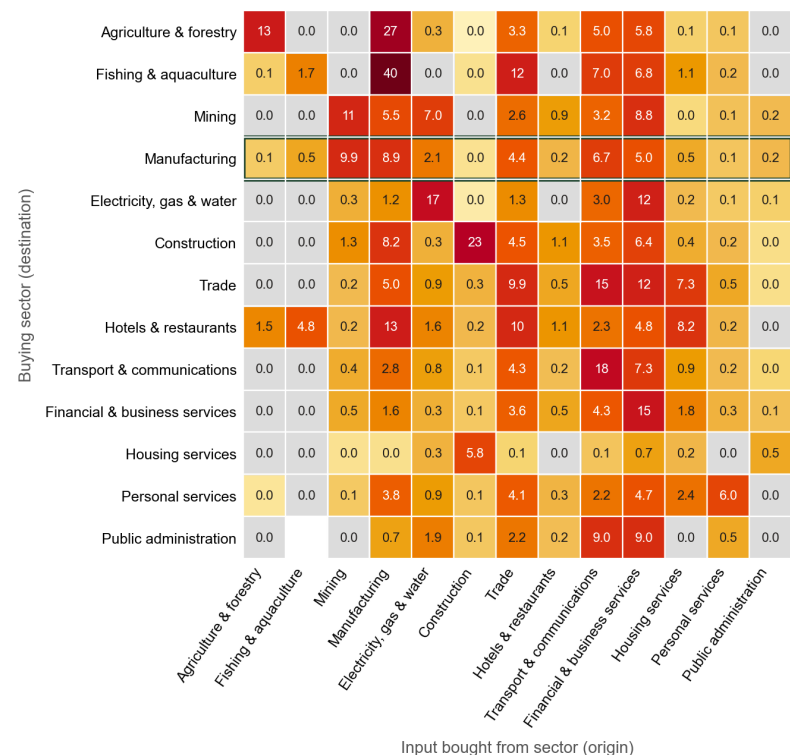
We embed the data into a state-of-the-art quantitative spatial trade model — production networks, labour mobility, firm selection, domestic and international trade, knowledge diffusion — to quantify how local productivity changes propagate



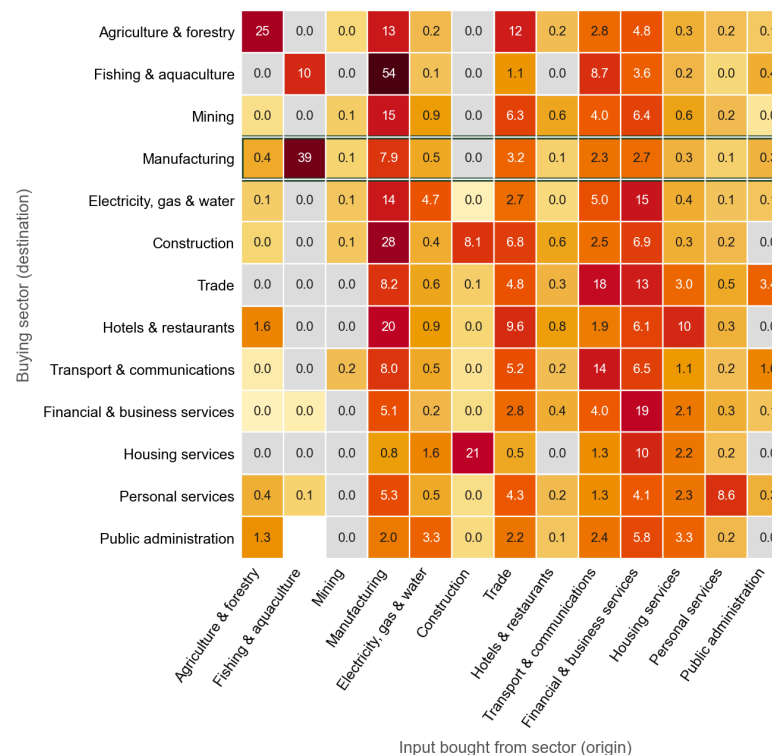
Province-to-province trade flows in the Chilean production network. Node size = output; Santiago, Antofagasta and the rest of the world are the hubs

Geography Matters: Input–Output Matrices Differ Across Space

Antofagasta (north: mining)



Llanquihue (south: salmon, forestry)



Same sector, different inputs

Manufacturing in **Antofagasta** buys ~10% of its gross output from **mining**

Manufacturing in **Llanquihue** buys 39% from **fishing & aquaculture**

A national input–output table gives both the same coefficients. Local and national coefficients correlate with an R^2 of only 0.41

Neither “where” nor “which sector” is enough — propagation depends on the interaction

Local input–output coefficients: purchases from each origin sector as % of the buying sector’s gross output (manufacturing row outlined). Colour scale is logarithmic. Source: Disaggregated Economic Accounts of Chile (CED)

What we Learned: Place × Sector Drives Propagation

73%

of the variation in how local productivity shocks move national GDP comes from the location × sector interaction (geography alone: 12%; sector alone: 16%)

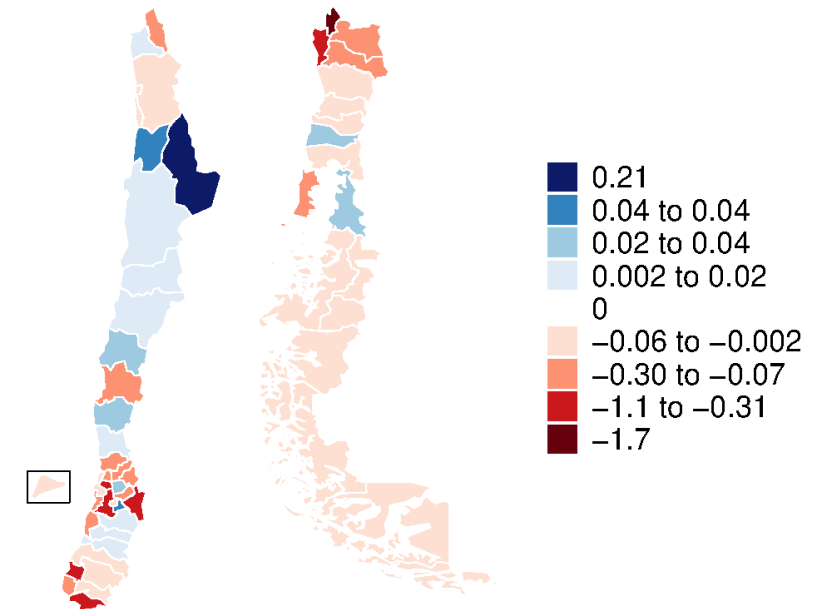
2.0×

average GDP response to a local productivity shock relative to the shock's own size (the standard benchmark is 1×)

3.4 vs 1.5

GDP elasticity to the Huachipato steel-plant closure with the geographic vs. the national input–output table — aggregates miss over half of the propagation

Huachipato steel plant closure (2024): GDP effect by province, %



National GDP -0.14%, employment -0.17%. Biobío -1.3%; 41% of the national loss occurs outside the plant's province. Blue = provinces that gain

Policy relevance

- Ex-ante evaluation of place-based and industrial policy; ex-post evaluation of plant closures, disasters and commodity shocks — with winners and losers by province and sector
- The invoice data already existed. The marginal cost was linking, harmonising and reconciling — now the CED, a public statistical product anyone can use

The Anatomy of Aggregate Productivity

Huneeus, Koike-Mori & Martner (2025)

QUESTION

Chile's productivity stagnated after 2010. Is it technology, or the allocation of resources across firms — and which firms, sectors and regions are responsible?

MICRODATA

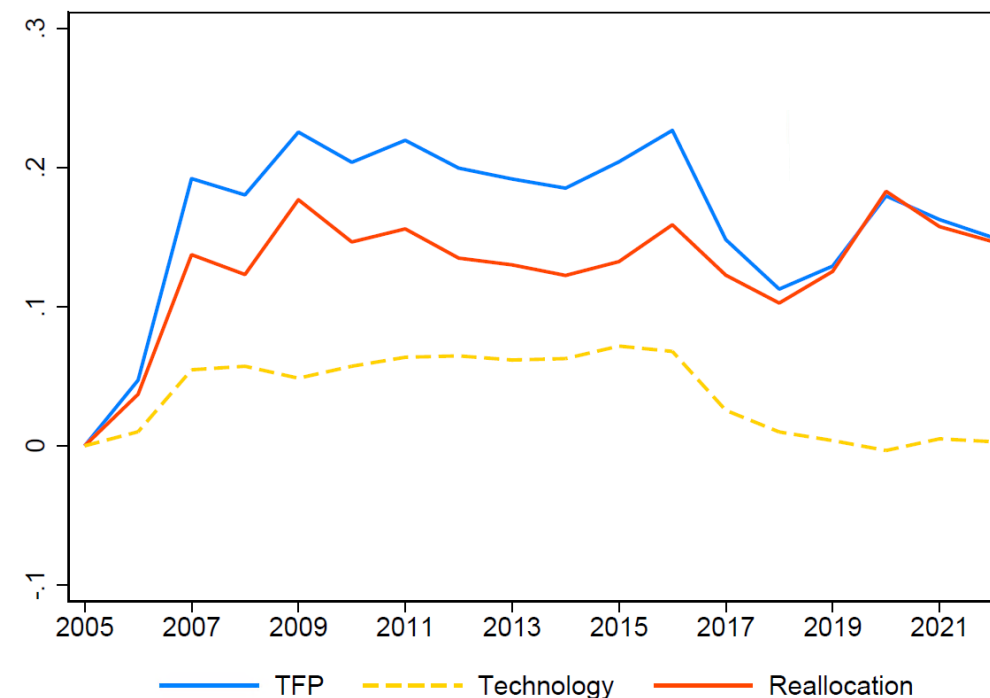
- Universe of formal firms 2005–2022: VAT returns, balance sheets, employer–employee records and firm-to-firm transactions
- 327,000 firms covering 85% of value added; production functions estimated for 626 industries
- Decomposition by 11 sectors × 16 regions × firm size × exporter status

APPROACH AND KEY THEORETICAL CONTRIBUTION

We characterise how to decompose aggregate productivity into arbitrary partitions of the economy — sectors, space, firm size, exporter status — using sufficient statistics measurable in standard data

Aggregate TFP = **technology** (firms get better) + **reallocation** (resources move toward or away from the most productive firms)

Cumulative aggregate TFP growth, Chile 2005–2022 (log points)



TFP = technology + reallocation. The technology term (dashed) is back to zero by 2022

What we Learned: a Lost Decade of Reallocation

+15%

cumulative aggregate TFP growth by 2022 — below the +22.5% already reached by 2009. The 2010s were a lost decade

≈ 98%

of 2005–2022 TFP growth is explained by reallocation across firms; technology contributes about zero

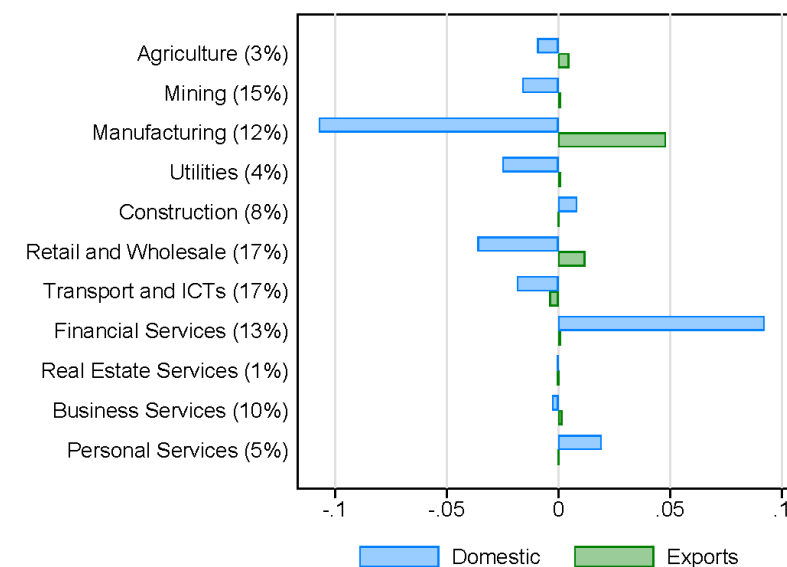
+21 vs +15

TFP growth measured with sector-level data vs. firm-level data (log points): aggregation overstates growth and reshuffles who matters

Policy relevance

- The stagnation is a problem of where resources go among large incumbent firms — not an innovation deficit. Different diagnosis, different policy
- A sector's share of GDP is a poor guide to its relevance for productivity: retail and transport both have 17% of value added, retail matters three times more

Reallocation contribution to TFP, 2010–2022, by sector and export status (log points)



Domestic manufacturing (−0.11) and retail (−0.04) drive the stagnation; financial services (+0.09) and manufacturing exports (+0.05) offset part of it. Sector shares of value added in parentheses

The Incidence of Distortions

Atkin, Bernadac, Donaldson, Garg & Huneus (2025)

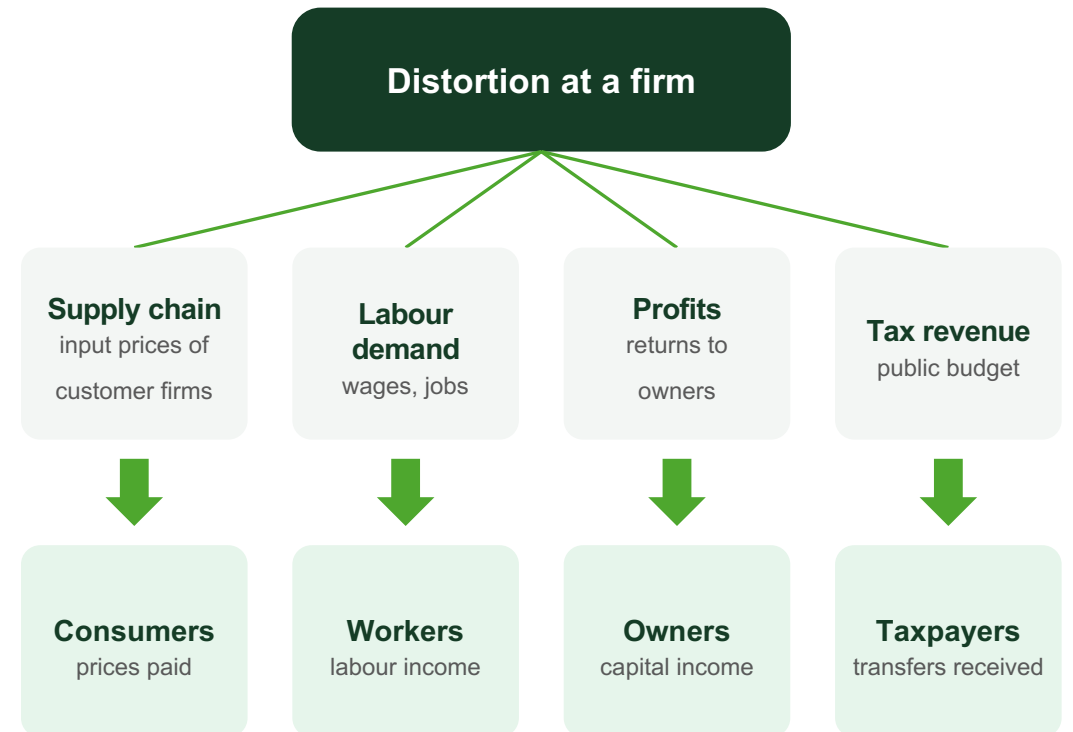
QUESTION

Distortions — market power, taxes, credit frictions — lower aggregate productivity. But who actually pays: workers, consumers, firm owners, or taxpayers?

MICRODATA

- Ten administrative datasets + four household surveys, linked by tax ID (2022)
- Firm-to-firm invoices (2.1 bn transactions), firm-to-consumer e-receipts (13.5 m consumers), employer–employee (8.2 m workers), ownership registry (1.8 m owners), income tax and transfers (8.0 m individuals)
- Result: ~13 million individuals linked to ~3.5 million firm-products

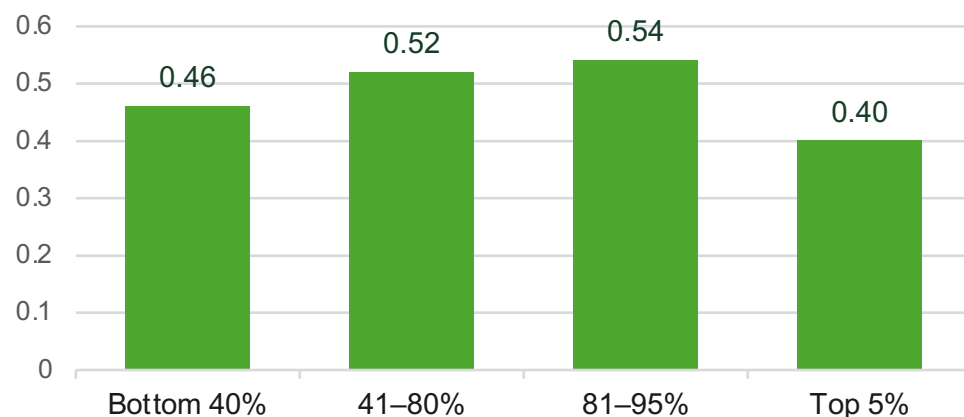
Following a distortion from the firm to the household



Each arrow is a different dataset linked at the firm or person level. Indirect links through the production network are included

What we Learned: Cutting Distortions Helps Middle Class More

Real income gain from a 1% cut in all distortions, by income group (%)

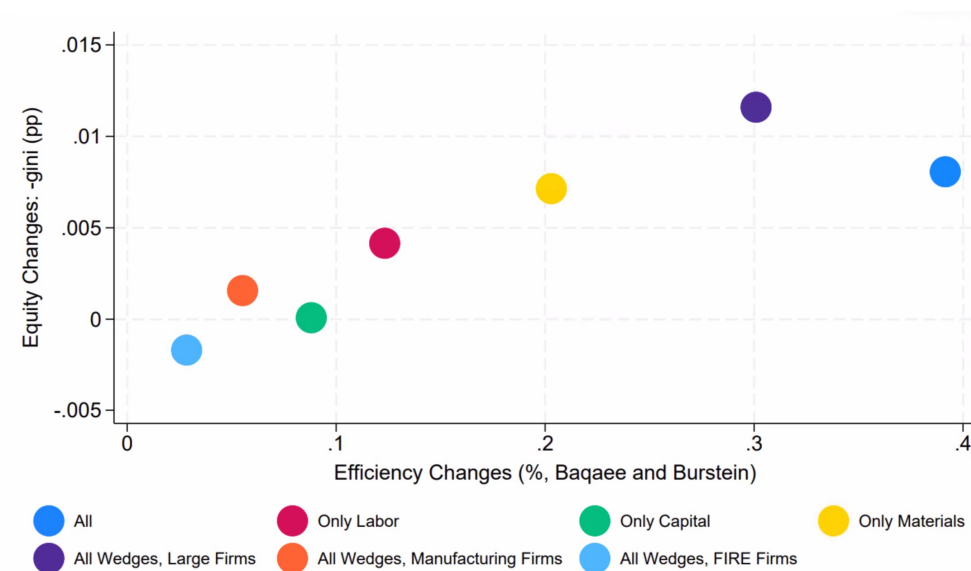


Everyone gains. The top 5% gain least because they own the high-distortion firms and lose profits; the bottom 40% lose some transfers as distortion revenue falls

Policy relevance

Biggest distributional channel is income side, through profit ownership (benefiting the rich) and tax ownership (benefitting the poor)

No equity–efficiency trade-off: seven reforms compared



Each dot is a counterfactual reduction of distortions. Reducing distortions only on large firms gives the largest fall in inequality (–Gini) with high efficiency gains; targeting only financial firms is slightly regressive. Only visible with firm-level distortions linked to ownership, employment and consumption

Priorities And Challenges For Managing Large Microdata

Three Priorities For a Scalable Microdata Infrastructure

1

A dedicated data team

- Obtains data from other institutions and builds trust: agreements, confidentiality by design
- Cleans, documents and links the data; quality improves with every vintage
- Returns to scale in efficient data management: data ready when the question arrives

2

Security and privacy

- Data is the new gold: the data rush brings opportunities — and risks
- Trust in the infrastructure is the binding constraint, more so with AI-era cybersecurity threats
- Invest in solid protocols and safeguards for data security and privacy

3

A research team at the frontier

- Brings frontier methods and best practices (e.g., theory instead of compute for large matrices)
- Attracts coauthors and visitors; reveals data gaps that feed back into cleaning
- Research becomes public goods: the CED started as a paper

Promoting International Cooperation on Data Infrastructures

Examples to build on

- **OECD distributed microdata (MultiProd, DynEmp):** harmonised code runs inside each country; only aggregates travel
- **INEXDA:** central banks sharing standards and practices for granular data
- **Trusted research environments:** UK Five Safes, France CASD, US FSRDC network

A CONCRETE NEXT STEP

Disaggregated national accounts across the globe: CED-style accounts in each country, connected across countries through customs data — firm-level trade is already recorded on both sides of every border

Questions for the panel

- 1 Should secure research environments be certified against a common standard — governance, privacy by design, cybersecurity?
- 2 Can cross-border firm-level analysis (trade, GVCs) be done without moving the data?
- 3 How do we connect country-level disaggregated accounts into a global network — who hosts the customs link?
- 4 Who funds and maintains the infrastructure over time?



ANNUAL CONFERENCE
GLOBAL FORUM ON
PRODUCTIVITY
C O S T A R I C A

Thank You

federico.huneus@duke.edu — www.fedehuneus.com