Markup dispersion across buyers*

Ariel Burstein
UCLA

Javier Cravino
University of Michigan

Marco Rojas
Central Bank of Chile

VERY PRELIMINARY, PLEASE DO NOT CIRCULATE.

Abstract

We use data on the universe of firm-to-firm invoices in Chile to document substantial dispersion in the prices that different buyers pay for individual products. The gap between the highest- and the lowest-price at which the median product is sold during a given month is 6 percent, and this gap exceeds 30 percent for 10 percent of the products. If marginal costs of production are independent of the identity of the product's buyer, the differences in prices across buyers reflect differences in buyer-specific markups. We evaluate the welfare gains of eliminating the dispersion in markups across buyers while keeping average product-level markups unchanged, using a quantitative model that takes into account the full network of firms in Chile. Preliminary results indicate that eliminating the cross-buyer dispersion in the markups of manufacturing products increases manufacturing productivity as much as 8 percent.

Keywords: markups, misallocation

*Email: arielb@econ.ucla.edu, jcravino@umich.edu, mrojaso@bcentral.cl .

1 Introduction

How large are the efficiency losses arising from markup dispersion? A growing literature in macroeconomics quantifies these losses using models of variable markups disciplined by firm-level data on market shares.[1] An alternative approach is to estimate markups at the firm level, and introduce them as exogenous wedges in framework a la Hsieh and Klenow (2009). As characterized by Baqaee and Farhi (2020), under both approaches the implied losses depend on the inferred dispersion of markups.

This paper uses transaction-level price data to measure how markups for individual products vary across buyers while imposing minimal assumptions. Our data comes electronic invoices collected by the Chilean Internal Revenue Service (Servicio de Impuestos Internos, or SII). Electronic Invoices have been mandatory for all firm-to-firm transactions in Chile since 2017. Each invoice records, for each product sold, a product code, a product description, and the price and quantity transacted. The invoice also records the date of the transaction along with tax-identifiers for both the seller and buyer. Thus, for each unique-product, the data records transacted prices that potentially vary across invoices and buyers.

We document that product-level prices vary substantially across transactions occurring within a given month. Most of this variation is accounted for by the fact that prices are buyer-specific: the buyers' identity accounts for nearly 90 percent of the variance in prices within products. The gap between the highest- and the lowest-price at which the median product is sold is 5 percent. Among products for which we observe multiple buyers, the gap for the median product reaches 30 percent.

Under the assumption that product-level marginal cost do not depend on the identity of the product's buyer, differences in prices across buyers reflect differences in buyer-specific markups. This variation in markups can lead to an inefficient allocation of resources across firms. We quantify these losses using a standard framework in the spirit of Hsieh and Klenow (2009), where markups are exogenous wedges that are product-buyer specific. We discipline our framework using our data, matching the full network of product to firm sales observed in the invoices.

We use our framework to evaluate a counterfactual in which we eliminate the dispersion in the markups of manufacturing products across intermediate buyers, while keeping average markups at the product-level as well as markups on final sales unchanged. We the counterfactual change in markups directly from the price data under the assumption

[1] See e.g. Edmond et al. (forthcoming), Dhyne et al. (2022) and Bornstein and Peter (2023).

above. We take the network (who buys from whom) as exogenous, though buyers can adjust the intensity with which they use each of their inputs. In our calibrated model, eliminating the dispersion in manufacturing markups across intermediate buyers results in manufacturing productivity by as much as 8 percent.

Literature: [To be added]

2 Data

2.1 Data sources

We use data collected by the Chilean Internal Revenue Service (SII for its acronym in Spanish). Our data comes from two sources:

Electronic Invoices (EI): Chilean Firms must issue an Electronic Invoice (EI) for every transaction with another domestic firm.[2] Each EI contains vast amount of information, which is separated into two sections: Heading and Detail. The first section contains the date of the transaction, tax-identifiers for the both seller and the buyer (RUT for its acronym in Spanish), the municipalities where they are located, their economic sectors, and the terms of the transaction, including who is responsible for the item's delivery, among other variables. The second part has information about the products sold, with separate entries for each different product. Each entry contains the price, quantity, discount (if any), surcharge (if any), a text description, a product code and its type (e.g., internal, SKU).[3] Firms can declare up to five product codes per product.[4]

A particularly important code is the European Article Number or EAN. This is a standardized barcode symbology that can identify a specific product, in a specific packaging configuration, and from a specific manufacturer and its country of origin. The main version of this code has a length of 13 digits and is referred to as EAN13. In this version the

[2] These documents are a subset of all Tax Electronic Documents used in Chile. In particular, we use "Factura Electrónica" that corresponds to the document type 33. Other related documents are final consumers receipts ("Boleta Electrónica"), export invoices ("Factura de Exportación", document type 110), tax-exempt invoices ("Factura No Afecta o Exenta Electrónica", document type 34), and more.

[3] To submit a valid EI, the SII only requires seller's RUT, buyer's RUT, text description of goods sold, total amount paid (including discounts and surcharges even if not reported), and a date. More information can be found (in Spanish) at https://www.sii.cl/factura_electronica/formato_dte.pdf.

[4] In practice, we only consider the first two.

first three digits corresponds to the country with the exception of the US and Canada that has two codes (e.g., Chile: 780, Argentina: 779) , the next five digits (though it can vary) correspond to the manufacturer, the next five to six digits corresponds to a feature-specific product code, and the last code is a verifying digit.

Form 22 Form 22 (F22) is an annual document that firms must complete in order to calculate the taxes owed. It outlays all the inflows and outflows of the firm. The inflows include sales related to the business activity, and other sources of income such as interests earned. The outflows consider wages, purchases of variable inputs such as materials and inputs, and other expenses such as donations.

2.2 Product-level prices

Table 1: Sample

Product definition: We define a unique product as a triplet of i) a product description, ii) a product code, iii) and a seller tax-id. Product codes are assigned by firms and are optional and varying quality: some firms use 13 digit EAN barcodes, while other firms use internal product codes. We define unique products only for the firms in the manufacturing sector, as including product codes in the electronic invoices is the norm in this sector. Table 1 summarizes the sample using only one month of data (June 2021).

Product-level prices: We denote the transacted price of of product ω in transaction i by $P_{ib}(\omega)$, where b is the buyer in transaction i. We then compute the average log-price for each product as $\overline{\log P(\omega)} \equiv \frac{\sum_b \sum_i Q_{ib}(\omega) \log P_{ib}(\omega)}{\sum_b \sum_i Q_{ib}(\omega)}$ and calculate the deviation between transaction prices from their average, $\log P_{ib}(\omega) - \overline{\log P(\omega)}$. Table 2 shows that around 87 percent of the variation is explained across buyers and not within.

Table 2: Within/between variation

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. var.: $\log P_{ib}(\omega) - \overline{\log P(\omega)}$						
R^2	0.122	0.537	0.789	0.416	0.749	0.870
Buyer FE	Yes	No	No	Yes	No	No
Buyer-Seller FE	No	Yes	No	No	Yes	No
Buyer-Product FE	No	No	Yes	No	No	Yes
Weighted	No	No	No	Yes	Yes	Yes

Finally, we compute the average price paid by buyer b for product ω across transactions $P_{bt}(\omega)$ as the average unit price:

$$P_{bt}(\omega) \equiv \frac{\sum_i P_{ibt}(\omega) Q_{ibt}(\omega)}{\sum_i Q_{ibt}(\omega)}. \tag{1}$$

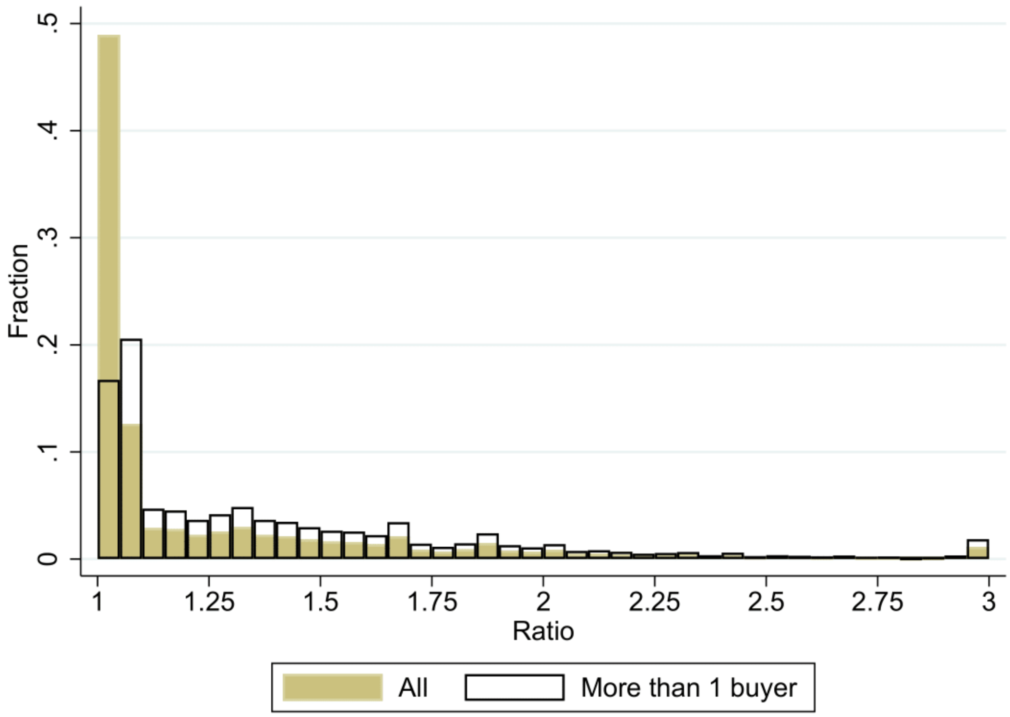
2.3 Differences in prices across buyers

We now compute, the ratio of betwen the highest and lowest price paid for each product, $P_{bt}^{max}(\omega)/P_{bt}^{min}(\omega)$. Table 3 computes the distribution of this statistic. For the (sales weighted) median product in our sample, the gap is 6 percent. Across products with multiple buyers, the gap is 28%, revealing substantial variation in prices across products.

Table 3: Percentiles of the price gap distribution

	All		More than 1 buyer	
	Unweighted	Weighted	Unweighted	Weighted
p25	1.0000	1.0000	1.0000	1.0741
p50	1.0000	1.0604	1.1111	1.2803
p75	1.0246	1.4021	1.3333	1.6593
p90	1.3092	1.9028	1.6840	2.1727
p95	1.5592	2.3344	2.0179	2.7077
p99	2.3669	3.4364	2.8827	3.8066
p99.5	2.7098	4.0772	3.4135	4.8688

Figure 1: Histogram of the price gap distribution



3 Theoretical framework

This Section presents the theoretical framework used to quantify the productivity costs of markup dispersion.

Preliminaries: We consider a closed economy with N sectors indexed by k and l, each populated by a discrete number of firms indexed by f. Firms produce differentiated products, indexed by i and j, which they can sell to other firms or to final consumers. A firm can potentially produce multiple products. Markups are exogenous and can change across products and buyers.

Technologies final producers/preferences: The final good is produced by aggregating the output from the different sectors:

$$C = \left[\sum_{l} \left[\bar{\beta}^{l} \right]^{\frac{1}{\eta_C}} \left[C^{l} \right]^{\frac{\eta_C - 1}{\eta_C}} \right]^{\frac{\eta_C}{\eta_C - 1}},$$

where C^l is a sectorial aggregate given by

$$C^l = \left[\sum_{i \in l} \bar{\gamma}_i^{\frac{1}{\rho_{lc}}} c_i^{\frac{\rho_{lc}-1}{\rho_{lc}}} \right]^{\frac{\rho_{lc}}{\rho_{lc}-1}},$$

where c_i denotes the quantity of product i consumed by the final good producers.

Technologies intermediate producers: The production function for product i is given by:

$$y_i = z_i b_i,$$

where b_i is a bundle of inputs used in the production of product i. Firms use the same bundle for each products that it produces. The bundle used by firm f is:

$$b_f = \left[\left[1 - \bar{\alpha}_f\right]^{\frac{1}{\sigma_{k(f)}}} l_f^{\frac{\sigma_{k(f)}-1}{\sigma_{k(f)}}} + \bar{\alpha}_f^{\frac{1}{\sigma_{k(f)}}} m_f^{\frac{\sigma_{k(f)}-1}{\sigma_{k(f)}}} \right]^{\frac{\sigma_{k(f)}}{\sigma_{k(f)}-1}},$$

where $k(f)$ denotes the sector of firm f. Here l_f is the labor input and m_f is a bundle of intermediate inputs:

$$m_f = \left[\sum_l \left[\bar{v}_f^l\right]^{\frac{1}{\eta_{k(f)}}} \left[m_f^l\right]^{\frac{\eta_{k(f)}-1}{\eta_{k(f)}}} \right]^{\frac{\eta_{k(f)}}{\eta_{k(f)}-1}},$$

with

$$m_f^l = \left[\sum_{i' \in l} \bar{\theta}_{i'f}^{\frac{1}{\rho_{lk(f)}}} x_{i'f}^{\frac{\rho_{lk(f)}-1}{\rho_{lk(f)}}} \right]^{\frac{\rho_{lk(f)}}{\rho_{lk(f)}-1}}.$$

Here $x_{i'f}$ denotes the quantity of product i' used by firm f. The firm uses it's input bundle across all its products:

$$b_f = \sum_{i \in f} b_i = \sum_{i \in f} \frac{y_i}{z_i}. \tag{2}$$

Market clearing: Goods market clearing implies:

$$y_i = c_i + \sum_f x_{if} \qquad \forall i. \tag{3}$$

Labor market clearing implies

$$\sum_f l_f = \bar{L}. \tag{4}$$

Prices and demands: Let p_{if} denote the price of product i when sold to firm f, and p_{ic} the price when sold to final consumers. These prices are given by:

$$p_{if} = \frac{\mu_{if} mc_{f'(i)}}{z_i}, \quad p_{ic} = \frac{\mu_{ic} mc_{f'(i)}}{z_i}, \tag{5}$$

where $f(i)$ denotes the firm that produces product i. Here μ_{if} and μ_{ic} are exogenous markups, and $mc_{f'(i)}$ is the cost of the input bundle for the firm that produces product i, given by

$$mc_f = \left[\left[1 - \bar{\alpha}_f\right] w^{1-\sigma_{k(f)}} + \bar{\alpha}_f q_f^{1-\sigma_{k(f)}} \right]^{\frac{1}{1-\sigma_{k(f)}}}. \tag{6}$$

Here w is the wage, and q_f is the cost of the intermediate input bundle used by firm f:

$$q_f = \left[\sum_l \bar{v}_f^l \left[q_f^l\right]^{1-\eta_{k(f)}} \right]^{\frac{1}{1-\eta_{k(f)}}}, \tag{7}$$

with

$$q_f^l = \left[\sum_{i' \in l} \bar{\theta}_{i'f} p_{i'f}^{1-\rho_{lk(f)}} \right]^{\frac{1}{1-\rho_{lk(f)}}}. \tag{8}$$

The price of the final goods is:

$$P = \left[\sum_l \bar{\beta}^l \left[P^l\right]^{1-\eta_c} \right]^{\frac{1}{1-\eta_c}}, \tag{9}$$

With

$$P^l = \left[\sum_{i \in l} \bar{\gamma}_i p_{ic}^{1-\rho_{lc}} \right]^{\frac{1}{1-\rho_{lc}}}. \tag{10}$$

Demands are given by:

$$x_{if} = \bar{\theta}_{if} \left[\frac{p_{if}}{q_f^{k(i)}} \right]^{-\rho_{k(i)l(f)}} m_f^{k(i)}; \quad c_i = \bar{\gamma}_i \left[\frac{p_{ic}}{P^{k(i)}} \right]^{-\rho_{k(i)c}} C^{k(i)}. \tag{11}$$

with

$$\begin{aligned} m_f^{k(i)} &= \bar{v}_j^{k(i)} \left[\frac{q_f^{k(i)}}{q_f} \right]^{-\eta_{k(j)}} m_f, \\ m_f &= \bar{\alpha}_f \left[\frac{q_f}{mc_f} \right]^{-\sigma_{k(f)}} b_f, \end{aligned} \tag{12}$$

and

$$C^k = \bar{\beta}^k \left[\frac{P^k}{P} \right]^{-\eta_{kc}} C. \tag{13}$$

Labor demand is given by:

$$l_f = [1 - \bar{\alpha}_f] \left[\frac{w}{mc_f} \right]^{-\sigma_{k(f)}} b_f. \tag{14}$$

Equilibrium: An equilibrium for this economy is given by a set of product-firm prices and product-consumer prices and quantities $\{p_{if}, p_{ic}\}_{\forall if}$ and $\{x_{if}, c_i\}_{\forall if}$, product level quantities $\{y_i\}_{\forall i}$, firm level costs and inputs $\left\{mc_f, q_f, q_f^l\right\}_{\forall f}$, and $\left\{l_f, m_f, m_f^l, b_f\right\}_{\forall f}$, and final prices and bundles $\{P^l\}_{\forall l}$ and P, and $\{C^l\}$, C and a wage $\{C^l\}$, C, such that $w = 1$ and the pricing equations (5), (6), (7), (8), (9), (10), the demands (11), (12), (13), (14), and market clearing (2), (3), (4) are satisfied.

3.1 Response to changes in markups

We now evaluate the changes in the equilibrium prices and quantities in response to exogenous changes in markups. We use the notation $\hat{X} \equiv \frac{X^c}{X}$ to denote the ratio of a variable in the counterfactual relative to the observed equilibrium. The change in the equilibrium is characterized by:

Changes in prices:

$$\hat{p}_{if} = \hat{\mu}_{if}\widehat{mc}_{f'(i)}, \quad \hat{p}_{ic} = \hat{\mu}_{ic}\widehat{mc}_{f'(i)}. \tag{15}$$

with

$$\begin{aligned} \widehat{mc}_f &= \left[1 - \alpha_f + \alpha_f \hat{q}_f^{1-\sigma_{k(f)}}\right]^{\frac{1}{1-\sigma_{k(f)}}} \\ \alpha_f &\equiv \frac{q_f m_f}{w l_f + q_f m_f}, \end{aligned} \tag{16}$$

and

$$\begin{aligned} \hat{q}_f &= \left[\sum_l \nu_f^l \left[\hat{q}_f^l\right]^{1-\eta_{k(f)}}\right]^{\frac{1}{1-\eta_{k(f)}}} \\ \nu_f^l &\equiv \frac{q_f^l m_f^l}{\sum_l q_f^l m_f^l}, \end{aligned} \tag{17}$$

and

$$\begin{aligned} \hat{q}_f^l &= \left[\sum_{i' \in l} \theta_{i'f} \hat{p}_{i'f}^{1-\rho_{k(f)l(i')}}\right]^{\frac{1}{1-\rho_{k(f)l(i')}}} \\ \theta_{i'f} &\equiv \frac{p_{i'f} x_{i'f}}{\sum_{j \in l(i')} p_{jf} x_{jf}} = \frac{p_{i'f} x_{i'f}}{q_f^{l(i')} m_f^{l(i')}}. \end{aligned} \tag{18}$$

Prices for the final goods are:

$$\begin{aligned} \hat{P} &= \left[\sum_l \beta^l \left[\hat{P}^l\right]^{1-\eta_c}\right]^{\frac{1}{1-\eta_c}} \\ \beta^l &\equiv \frac{P^l C^l}{PY}. \end{aligned} \tag{19}$$

and

$$\hat{P}^l = \left[\sum_{j\in l} \gamma_j \hat{p}_{jc}^{1-\rho_{l(j)c}}\right]^{\frac{1}{1-\rho_{l(j)c}}} \tag{20}$$
$$\gamma_j \equiv \frac{p_{jc}c_j}{P^{l(j)}C^{l(j)}}.$$

Changes in quantities

$$\hat{x}_{if} = \left[\frac{\hat{p}_{if}}{\hat{q}_f^{k(i)}}\right]^{-\rho_{k(i)l(f)}} \hat{m}_f^{k(i)}; \quad \hat{c}_i = \left[\frac{\hat{p}_{ic}}{\hat{P}^{k(i)}}\right]^{-\rho_{k(i)c}} \hat{C}^{k(i)}, \tag{21}$$

with

$$\hat{m}_f^{k(i)} = \left[\frac{q_f^{k(i)}}{\hat{q}_f}\right]^{-\eta_{k(j)}} \left[\frac{\hat{q}_f}{\widehat{mc}_f}\right]^{-\sigma_{k(f)}} \hat{b}_f,$$
$$\hat{C}^k = \left[\frac{\hat{P}^k}{\hat{P}}\right]^{-\eta_{kc}} \hat{C}. \tag{22}$$

Labor demand is:

$$\hat{l}_f = \left[\frac{1}{\widehat{mc}_f}\right]^{-\sigma_{k(f)}} \hat{y}_f. \tag{23}$$

The change in the firms' bundle satisfies

$$\hat{b}_f = \sum_i \hat{y}_i \frac{b_i}{b_f}. \tag{24}$$

Goods market clearing implies:

$$\hat{y}_i = \frac{c_i}{y_i}\hat{c}_i + \sum_f \frac{x_{if}}{y_i}\hat{x}_{if}. \qquad \forall i \tag{25}$$

Labor market clearing implies

$$1 = \sum_f \hat{l}_f \frac{l_f}{L}. \tag{26}$$

4 Quantitative results

This section presents a calibrated version of the model and provides the algorithm for evaluating the response to changes in markups.

We focus on an economy with two sectors, which we denote Manufacturing (firms that report product-level ids), and Non-Manufacturing (firms that don't report no product-level ids). We distinguish two type of firms in Non-Manufacturing: those that sell to other firms (so that they issue Electronic Invoices) and those that only sell to final consumers (and hence only show up in the Electronic Invoices as buyers). We denote the number of manufacturing firms by NF_M and the number of non-manufacturing firms as $NF_{TN} = NF_{XN} + NF_{FN}$, where NF_{XN} and NF_{FN} respectively denote the number of non-manufacturing firms with and without intermediate sales. The total number of firms that sell intermediate goods is $NF_{XT} \equiv NF_M + NF_{XN}$. The total number of firms and the number of potential buyers of intermediate goods is $NF_T = NF_M + NF_{TN}$. Finally, we assume non-manufacturing firms are single product firms, while Manufacturing firms are potentially multi-product. The number of manufacturing products is $NP_M \geq NF_M$. The total number of intermediate products is $NP_{XT} = NP_M + NF_{XN}$, and the total number of products is $NP_T = NP_{XT} + NF_{FN}$.

4.1 Calibration

We focus on an economy with two sectors, which we denote Manufacturing (firms that report product-level ids, indexed by M), and Non-Manufacturing (firms that don't report no product-level ids, indexed by O).

We assume that elasticities of substitution are the same across uses and take their values form the literature. Thus, we denote the within sector elasticities as $\rho^{kl} = \rho^{k'c} = \rho$, the elasticities across sectors as $\eta_c = \eta_k = \eta_{k'} = \eta$, and the elasticities across factors as $\sigma_k = \sigma$.

For the shares and shocks, we will use electronic invoices, which contain all the sales between firms in the Chilean Economy. From the invoices we obtain the sales for each link in the network $s_{if} \equiv p_{if} x_{if}$ and prices for each link in the network in which the seller is a manufacturing firm p_{if}. We use the shares to compute the shares of each firm in total intermediate sales and purchases: $s_f^m = \sum_{f'} \sum_{i \in f'} s_{if}$ and $s_f^x = \sum_{i \in f} \sum_{f'} s_{if'}$. To obtain $\hat{\mu}_{if} \equiv \frac{\mu_{if}}{\mathcal{M}_i}$, first define $\mathcal{M}_i \equiv \frac{\sum_f p_{if} x_{if}}{\sum_f mc_i x_{if}} = \left[\sum_f \mu_{if}^{-1} \frac{s_{if}}{\sum_f s_{if}}\right]^{-1}$. Then we can compute $\hat{\mu}_{if} = \sum_{f'} \frac{\mu_{if}}{\mu_{if'}} \frac{s_{if'}}{\sum_{f'} s_{if'}} = \sum_{f'} \frac{p_{if}}{p_{if'}} \frac{s_{if'}}{\sum_{f'} s_{if'}}$ for all manufacturing firms. In the counterfactual, non-manufacturing markups and markups to final consumers are kept constant, so we set $\hat{\mu}_{if} = 1$ for non-manufacturing products, and $\hat{\mu}_{ic} = 1$ for all products.

Next, we either Form F-22 to obtain the ratio of revenues to sales for each firm $\frac{\alpha_f}{\mathcal{M}_f}$. With this data we can obtain the share of firm sales that go to final consumers: $\omega_f \equiv 1 - \frac{\alpha_f}{\mathcal{M}_f} \frac{s_f^x}{s_f^m}$. We also use Form F-22 to compute the ratio of intermediate inputs in total costs for each firm α_f. Equipped with these parameters we can obtain all the remaining shares:

- $s_f^c = \frac{\mathcal{M}_f}{\alpha_f} s_f^m \omega_f \left[\frac{\mathcal{M}^T}{\alpha^T} - 1\right]$ share of each firm in total consumption.
- $\beta_M \equiv \frac{\sum_{i \in M} p_{ic} c_i}{\sum_i p_{ic} c_i} = \sum_{f \in M} s_f^c$ share of manufactures in final good.
- $\gamma_f \equiv \frac{\sum_{i \in f} p_{ic} c_i}{\sum_{i \in k(i)} \sum_{i \in f} p_{ic} c_i} = \frac{s_f^c}{\sum_{f \in k} s_f^c}$ share of firm in it's sectorial consumption bundle.
- $\nu_{fM} \equiv \frac{\sum_{j \in M} s_{jf}}{\sum_j s_{jf}}$ share of sector l in intermediate input purchases of firm f.
- $\theta_{jf} \equiv \frac{s_{jf}}{\sum_{j \in k(j)} s_{jf}}$ share of good j in intermediate input purchases from sector k of firm f.
- $s_f^l = s_f^m \frac{1-\alpha_f}{1-\alpha^T} \frac{\alpha^T}{\alpha_f}$ and the share of each firm in total employment.
- $\frac{x_{if}}{\sum_f x_{if}} = \frac{s_{if}}{\sum_f s_{if}} \left[\frac{\mu_{if}}{\mathcal{M}_i}\right]^{-1}$

To calibrate $\frac{c_i}{y_i}$, we assume that the share of sales that go to final consumers is the same across all products of the same firm, $\omega_i \equiv \frac{p_{ic} c_i}{p_{ic} c_i + \sum_{f'} p_{if'} x_{if'}} = \omega_{f(i)}$. We also assume that $\mu_{ic} = \mathcal{M}_i$, that is, that the markup to final consumers is the same as the average markup to intermediate firms. Under these assumptions $\frac{c_i}{y_i} = \omega_{f(i)}$. Finally, to calibrate $\frac{b_i}{b_f}$ we also need to assume that the average markup is the same for all the products of the same firm, $\mathcal{M}_i = \mathcal{M}_{f(i)}$. We then obtain $\frac{b_i}{b_f} = \frac{\sum_f s_{if}}{\sum_{i' \in f} \sum_f s_{i'f}}$.

4.2 Results

[To be added, results waiting to be disclosed].

5 Conclusion

[To be added]

References

Baqaee, David Rezza and Emmanuel Farhi, "Productivity and Misallocation in General Equilibrium," *The Quarterly Journal of Economics*, 2020, *135* (1), 105–163.

Bornstein, Gideon and Alessandra Peter, "Nonlinear Pricing and Misallocation," Working Papers, Mimeo 2023.

Dhyne, Emmanuel, Ayumu Ken Kikkawa, and Glenn Magerman, "Imperfect Competition in Firm-to-Firm Trade [The Effects of Joining Multinational Supply Chains: New Evidence from Firm-to-Firm Linkages]," *Journal of the European Economic Association*, 2022, *20* (5), 1933–1970.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, "How Costly Are Markups?," *Journal of Political Economy*, forthcoming, *0* (ja), null.

Hsieh, Chang-Tai and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, November 2009, *124* (4), 1403–1448.

APPENDIX

A Data Appendix

A.1 Data Cleaning

Now we outline the data cleaning process we perform to the raw data coming from EI. The steps are the following:

i. Drop entries that have prices or quantities that are zero or negative.

ii. Drop sales above 10^{11} (around 100 million USD).

iii. Drop if seller and buyer are the same firm.

iv. Drop if buyer is not a firm (as opposed to a person)[1]: this is defined by the Statistics Division at the Central Bank of Chile.

v. Drop if a firm (buyer or seller) has a valid F29:

   (a) The firm has an F29

   (b) Final consumer sales and domestic intermediate sales are positive

   (c) Final consumer share is between 0 and 1 (for seller cannot be 1): if the buyer is not a seller, then we assume all sales are final

vi. Drop if a seller is not a buyer in EI

A.2 Sample

Our initial sample contains the universe of EI between June 2021 to December 2021 for all sectors where prices and quantities are above zero and the buyer and seller are not the same firm (Criteria 0). We focus in the Manufacturing sector (Criteria 1), where we work on three subsamples therein for the main analysis:

i. Any code (Criteria 2): entries that have *any* code for their products

ii. EAN (Criteria 3): entries that have an EAN code

iii. EAN 13 (Criteria 4): entries that have an EAN code of length 13

Table A0 shows the number of entries, their associated total sales, number of sellers, buyers and product codes. It also includes the share of entries with a discount and how much this is on average. Each sample covers 76%, 6% and 3% of the manufacturing sample.

[1]Firms that mostly sell to firms might not have final consumer type receipt ("Goleta's in Chile), so when they sell to people, they use an EI where the buying RUT is that of a person.

Table A0: Summary of samples

Table A0 shows the sectoral distribution of sales for the entire manufacturing sample, and for the other three samples we work with. The Food sector is the largest in sales, but it goes from representing around 30% of total sales for all manufacturing to almost half for the EAN and EAN13 samples. This is partly a consequence of the Fuel refining and Basic metal industries disappearing for the those two samples. These facts hold for reported and implied prices.3

Table A0: Sector distribution by sample

In addition, we can study input-output tables for the whole economy and for each of the subsamples we work with. Table A0 shows the sales (in billion Chilean pesos) between 12 sectors in our data. Rows are sellers and columns are buyers.

Table A0: Input-Output table for the whole economy

A.3 Outlier transactions

In order to get rid off outliers at the transaction level, due to a typo in either price, quantity, code or all the above, we compute the average unit price excluding the transaction i itself as follows,

$$P_{-i,bt}(\omega) = \frac{\sum_{j\neq i} P_{jbt}(\omega) Q_{jbt}(\omega)}{\sum_{j\neq i} Q_{jbt}(\omega)}.$$

Then, we use the difference $\log P_{ibt}(\omega) - \log P_{-i,bt}(\omega)$ and drop all transactions that are not within [-1,1]. This is equivalent to a transaction being lower than 38 percent or higher than 272 percent of their average unit price excluding those transactions. We can do this for reported and implied prices. In addition, we can consider the relevant "market" pooling all periods, which means obtaining differences with respect to $P^{Imp}_{-i,bt}(\omega)$, $P_{-i,b}(\omega)$, and $P^{Imp}_{-i,b}(\omega)$ as well.[2]

After we have dropped outliers for each sample above, we can measure their implied price inflation and assess whether they compare to the official Manufacturing Producer Price Index (ManPPI). First, we compute an average price for each ω,

[2]This means computing,

$$P_{-i,b}(\omega) = \frac{\sum_t \sum_{j\neq i} P_{jbt}(\omega) Q_{jbt}(\omega)}{\sum_t \sum_{j\neq i} Q_{jbt}(\omega)}.$$

$$p_t(\omega) \equiv \sum_b \sum_i \log P_{ibt}(\omega)\, s_{ibt}(\omega)\,,$$

with $s_{ibt}(\omega) = \frac{P_{ibt}(\omega) Q_{ibt}(\omega)}{\sum_b \sum_i P_{ibt}(\omega) Q_{ibt}(\omega)}$. Then, this is used to calculate a product-specific inflation $\pi_t(\omega) = p_t(\omega) - p_{t-1}(\omega)$. These are aggregated to form a sectorial inflation, $\pi_{st} = \sum_{\omega \in s} \pi_t(\omega)\, s_{t_0}(\omega)$, where $s_{t_0}(\omega)$ is sales of product ω divided by all sales in sector s. Finally, we compute $\pi_t = \sum_s w_s \pi_{st}$ using the official manufacturing PPI weights, w_s.

Figure A1 plots the official monthly inflation of Manufacturing PPI calculated by the National Institute of Statistics in Chile, and our measure using our samples.

Figure A1: Manufacturing PPI, official and estimated