Testing the Production Approach to Markup Estimation*

Devesh Raval

Federal Trade Commission

devesh.raval@gmail.com

December 22, 2022

Abstract

Under the production approach to markup estimation, any flexible input should recover the markup. I test this implication using manufacturing datasets from Chile, Colombia, India, Indonesia, the US, and Southern Europe, as well as store-level data from a major US retailer, and overwhelmingly reject that markups estimated using labor and materials have the same distribution. For every dataset, markups estimated using labor are negatively correlated with markups estimated using materials, exhibit greater dispersion, and have opposite time trends. I continue to find stark differences in markups estimated using energy and non-energy raw materials. Non-neutral productivity differences across firms can explain these findings.

*First Version: November 2018. I would like to thank Chris Adams, John Asker, Emek Basker, Ben Bridgman, Allan Collard-Wexler, Emin Dinlersoz, Liran Einav, Amit Gandhi, John Haltiwanger, Berthold Herrendorf, Dan Hosken, Liran Einav, Ethan Kaplan, Rob Kulick, Fernando Luco, Ildiko Magyari, Ryne Marksteiner, Aviv Nevo, Ezra Oberfield, Ariel Pakes, Ted Rosenbaum, Pierre-Daniel Sarte, Dave Schmidt, Marshall Steinbaum, Philippe Sulger, Andrew Sweeting, Nico Trachter, James Traina, Brett Wendling, Kirk White, Nate Wilson, and Mo Xiao, as well as editor Thomas Chaney and three anonyomous referees for their comments on this paper. I also thank Jordi Jaumandreu, Paul Scott, and Dan Ackerberg for discussing this paper at the 2019 IIOC, 2019 NYC IO Day, and 2020 AEA conferences, Ana Fernandes, Joep Konings, and Benni Moll for providing deflators for various datasets in this paper, and Miklos Koren and Ann Law for their help with the final submission. Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the Federal Trade Commission, or its Commissioners.

Measuring the markup of price over cost is central to recent debates on whether market power has been rising for the US and the world economy (Basu, 2019; Berry et al., 2019; De Loecker et al., 2020; Syverson, 2019). Markups are crucial to evaluate the effects of mergers and changes in trade barriers. Rising aggregate markups can also explain macroeconomic phenomena such as the decline in the labor share of income (Grossman and Oberfield, 2022).

The *production approach* to markup estimation (Hall, 1988; Klette, 1999; De Loecker and Warzynski, 2012) has allowed economists to measure aggregate markups by estimating firm level markups across industries. The production approach uses flexible input choice to identify the markup as a variable input's output elasticity divided by its share of revenue.[1]

This approach requires one to know the production function. In practice, economists using the approach have typically assumed that productivity is Hicks neutral. However, when productivity is labor augmenting, more productive firms will have different output elasticities of labor and materials than less productive firms. Ignoring such heterogeneity will lead to systematically different markups estimated using different inputs.

Because *any* flexible input identifies the markup, the markup is overidentified with multiple flexible inputs. I thus compare markups estimated using labor, materials, or, mirroring cost of goods sold in De Loecker et al. (2020), a composite of both.[2] I conduct these comparisons using manufacturing censuses from Chile, Colombia, India, and Indonesia, firm level

[1]Given competitive input markets, a cost minimizing firm sets the additional revenue from a marginal increase in a flexible input equal to the marginal cost of the input multiplied by the markup.

[2]In the literature, De Loecker and Warzynski (2012) and Blonigen and Pierce (2016) use labor, De Loecker et al. (2016) materials, De Loecker and Scott (2017) both, De Loecker and Eeckhout (2018) cost of goods sold, and De Loecker et al. (2020) cost of goods sold (Compustat) and labor (Economic Census).

data from financial statements for the US and Southern Europe, as well as unique data on individual stores from a nationwide US retailer.

Because the production approach requires estimates of the production function, these comparisons jointly test the assumptions of the production approach itself and auxiliary assumptions on production technology. The validity of estimates of markups using the production approach will thus also depend upon the accuracy of auxiliary assumptions such as Hicks neutrality. I follow De Loecker and Warzynski (2012) and estimate production functions using the Ackerberg et al. (2015) control function estimator, which assumes productivity is Hicks neutral.

I strongly reject that different inputs estimate the same markup in all seven datasets. I focus on three major features of the markup distribution. Labor markups are much more dispersed than materials markups. Markup measures using labor and materials are *negatively* correlated in the cross-section. Finally, their time trends are negatively correlated as well.

One explanation for these findings is that frictions such as hiring and firing costs or monopsony power affect the labor, but not materials, static cost minimization condition. However, I continue to find stark differences between markups estimated using energy and non-energy raw materials. In addition, after controlling for local labor markets using the retailer data – which should capture both differences in input prices and monopsony power across stores – labor markups remain highly negatively correlated with materials markups.

My findings of conflicting correlations when estimating markups with different inputs are also robust to several estimation approaches that assume only neutral productivity differences, estimating production functions at the subindustry or product level, and estimating quantity rather than revenue production functions.

Non-neutral technology can explain these findings. Cost minimization implies that the ratio of output elasticities for two inputs must equal the ratio of their costs. With only Hicks neutral productivity, the ratio of output elasticities is a constant for the Cobb-Douglas, and a deterministic function of production function parameters and inputs with no error for the translog, and so cannot match the variation in input costs in the data. In contrast, non-neutral productivity provides an error term for the ratio of output elasticities that saturates the model.

When labor and materials are complements, higher labor augmenting productivity would both lower labor's output elasticity relative to materials' output elasticity and labor costs relative to materials costs. By ignoring such productivity differences when estimating output elasticities, markups based upon alternative inputs would have opposing time trends and negative correlations, as I have found.

This article is most similar to work that examines differences between markup estimates using the production approach. De Loecker et al. (2020), Karabarbounis and Neiman (2018), and Traina (2018) debate how using different inputs from Compustat affects the aggregate trend in US markups, while Bridgman and Herrendorf (2022) examines the same question using the National Accounts. De Loecker and Scott (2017) find similar average markup estimates using the demand approach, as in Berry et al. (1995), to those from the production approach using data on US breweries.

This article is also related to the literature on non-neutral productivity. Raval (2019) and Oberfield and Raval (2021) document growth in labor augmenting productivity and labor augmenting productivity differences for US manufacturing; Doraszelski and Jaumandreu (2018) and Zhang (2019) do the same using Spanish manufacturing and Chinese steel data.

Section 1 lays out the production approach to estimating markups. Section 2 detail the data and control function estimators, while Section 3 tests the production approach. Section 4 argues that labor augmenting technology differences can explain the failure of the tests and discusses new estimators (Demirer, 2020; Doraszelski and Jaumandreu, 2019) accounting for such differences. Section 5 concludes.

1 Production Approach

Hall (1988) introduced the production approach by showing that, under perfect competition, increases in an input should increase output by the input's share of revenue. He identified the markup as the wedge between the two; that is, the ratio of the change in output divided by the change in the input multiplied by the revenue share. Because productivity shocks would also affect output and inputs, Hall (1988) proposed using aggregate instruments for input growth to estimate the markup.[3]

[3] Hall (1988) estimated this model on time series data for several industries, with aggregate instruments

Klette (1999) was the first to extend the Hall (1988) framework to firm-level data, deriving that the firm level markup is the output elasticity of an input divided by its share of revenue for the firm, as in (4) below. However, he argued that this equation was unlikely to hold without error. Instead, Klette (1999) developed a panel data framework to jointly estimate the average markup across firms in the industry using within firm changes in variable inputs, and average returns to scale using within firm changes in fixed inputs.

Unlike Klette (1999), De Loecker and Warzynski (2012) recovered firm level markups by estimating the output elasticities for each firm. The key assumptions in their model were that the firm cost minimizes in each period with respect to a variable input, and that the firm is a price taker in the input market for that input. Below, I follow De Loecker and Warzynski (2012) and derive the estimator for the markup under these assumptions.

A firm produces output with production function $F_{it}(K_{it}, L_{it}, M_{it})$, where K_{it} is capital for firm i and time t, L_{it} is labor, and M_{it} is materials. The firm receives price P_{it} for its output and faces input prices p_{it}^X for input X. A cost minimizing firm sets marginal products equal to factor prices. This implies, for variable input X_{it},

$$P_{it} \frac{\partial F_{it}}{\partial X_{it}} = \frac{P_{it}}{\lambda_{it}} p_{it}^X, \tag{1}$$

where λ_{it} is the firm's marginal cost.[4] The left hand side is the marginal revenue product of

such as military spending, oil prices, and the president's party. Others followed with more disaggregate industry-level data (Domowitz et al., 1988; Basu and Fernald, 1997).

[4]The marginal cost is the Lagrange multiplier on the production function in the cost minimization prob-

increasing input X_{it}. The right hand side is the marginal cost of increasing X_{it} – its price, p_{it}^X – multiplied by the markup $\frac{P_{it}}{\lambda_{it}}$. Thus, the markup is a wedge between the marginal revenue product of an input and the marginal cost of an input.

Converting this expression to elasticity form[5], the output elasticity for input X, β_{it}^X, is equal to the markup μ_{it} multiplied by input X's share of revenue s_{it}^X:

$$\begin{aligned} \frac{\partial F_{it}}{\partial X_{it}} \frac{X_{it}}{F_{it}} &= \frac{P_{it}}{\lambda_{it}} \frac{p_{it}^X X_{it}}{P_{it} F_{it}} && (2) \\ \beta_{it}^X &= \mu_{it} s_{it}^X. && (3) \end{aligned}$$

The markup μ_{it} is then the output elasticity of input X divided by X's share of revenue:

$$\mu_{it} = \frac{\beta_{it}^X}{s_{it}^X}. \tag{4}$$

This expression for markups holds for *all* variable inputs at the firm level. Thus, I can test the production approach by examining whether the markup recovered using one input is the same as the markup recovered using another.

lem.

[5]Formally, multiply each side by $\frac{X_{it}}{F_{it}}$ and divide each side by the price P_{it}.

2 Data and Estimation

I summarize the seven datasets used in this paper in Table I, and provide more details on data construction in Appendix A.

The first six datasets cover manufacturing; I use yearly plant censuses for Chile, Colombia, and India, firm census for Indonesia, data from Compustat for an unbalanced panel of US public firms, and data from ORBIS for a balanced panel of all firms in Italy, Spain, and Portugal ("Southern Europe"). I define industries at an equivalent to 2 digit SIC (a broader 2 digit NAICS for the US) and include only industries with at least 1,000 observations.

I also use unique store level data on revenue and inputs for thousands of retail stores from a major nationwide US retailer ("Retailer") for three years. It provides each store's location, which allows me to assess explanations for differences in markups estimated using different inputs in Section 4.2. Because the retailer's data is from its internal accounting systems, it is much higher quality than manufacturing survey data.

Table I Datasets

Dataset	Unit of Observation	Time Period	No. Establishments	No. Industries Used
Chile	Manufacturing Plant	1979-1996	5,000 / year	16
Colombia	Manufacturing Plant	1978-1991	7,000 / year	21
India	Manufacturing Plant	1998-2014	30,000 / year	23
Indonesia	Manufacturing Firm	1991-2000	14,000 / year	22
US (Compustat)	Manufacturing Firm	1970-2010	500 / year	3
Southern Europe (ORBIS)	Manufacturing Firm	2011-2020	100,000 / year	23
Retailer	Retail Store	3 years	Thousands / year	1

I follow De Loecker and Warzynski (2012) and estimate industry-level revenue production functions using the Ackerberg et al. (2015) control function estimator, which imposes that productivity is Hicks neutral and evolves following a Markov process. Appendix B provides more details on estimation.

I then estimate Cobb-Douglas and translog production functions, with two input specifications for each production function. In one specification, inputs are capital, labor, and materials; in another, inputs are capital and a composite variable input of labor and materials costs. I examine a composite variable input in order to mimic cost of goods sold, which combines labor and materials, in De Loecker et al. (2020).[6] With the resulting output elasticities, I build six markup estimates for each establishment-year, using one of three inputs (labor, materials, or the composite input) and one of the two production functions.

3 Empirical Tests

Under the production approach, any flexible input identifies the markup. I test the production approach by examining how markup dispersion, time series correlations, and cross-sectional correlations vary using different inputs. For all of these tests, and in all the datasets, I strongly reject that different inputs estimate the same markup.

In Web Appendix D, I continue to find sharp differences between inputs examining markup correlations with size, exports, a profit share based markup, and the level of com-

[6] In some cases, datasets record aggregated variable expenditures of the firm but do not differentiate between inputs like labor and materials.

petition.[7]

3.1 Dispersion in Markup Estimates

Under the production approach, the degree of markup dispersion should be the same using different flexible inputs. Instead, I find very different levels of dispersion using different inputs. As an illustrative example, I plot the distribution of each markup across manufacturing plants in the Chilean Food Products industry in 1996 using the translog estimates in Figure 1. The red solid lines are the labor markup, the blue dashed lines the materials markup, and the green dash-dot lines the composite variable input markup. The labor markups are much more dispersed than the materials markups, which are in turn more dispersed than the composite input markups. Because of its greater dispersion, a large fraction of labor markups are below one, which might be considered a lower bound on markups (Flynn et al., 2019).

I find a similar pattern in all the datasets. I measure dispersion by calculating the 90/50 ratio of the markup estimates, which I report in Table II.[8] Just as in Figure 1, labor markups are more disperse than materials markups, which are more disperse than composite input markups, for each dataset and production function except for the US Cobb-Douglas estimates. For example, using the translog estimates, the 90th percentile markup is 103% higher than the median markup for Chile using labor, 39% higher using materials, and 17%

[7] I also examine average markups in Web Appendix E.3.
[8] I report the 75/25 and 90/10 ratios in Web Appendix E.2.

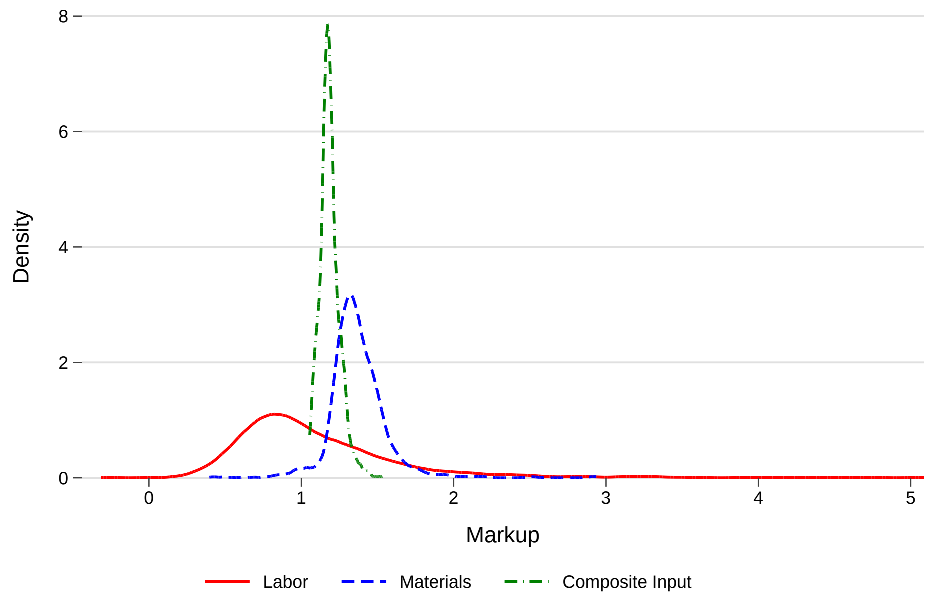


Figure 1 Distribution of Translog Markups for Chilean Food Products, 1996

using the composite input.

For the retailer, there is hardly any dispersion in materials markups – the 90th percentile markup is only 3% higher than the median and 6% higher than the 10th percentile – but substantial dispersion in the labor markup. For the labor markup, the 90th percentile is 30% higher than the median markup and 76% higher than the 10th percentile under the translog estimates.

3.2 Time Trends

Under the production approach, the time path in markups should be the same using different flexible inputs. To test this, I estimate the following specification:

Table II 90/50 Ratio of Markup Estimates

	Labor		Materials		Composite Input	
Dataset	Cobb-Douglas	Translog	Cobb-Douglas	Translog	Cobb-Douglas	Translog
Chile	2.67	2.03	1.53	1.39	1.17	1.17
	(0.013)	(0.008)	(0.003)	(0.004)	(0.001)	(0.001)
Colombia	2.88	1.82	1.82	1.43	1.16	1.17
	(0.016)	(0.005)	(0.008)	(0.004)	(0.001)	(0.001)
India	4.04	2.95	1.38	1.29	1.14	1.14
	(0.013)	(0.007)	(0.001)	(0.001)	(0.000)	(0.000)
Indonesia	4.06	3.12	1.66	1.46	1.15	1.16
	(0.025)	(0.019)	(0.004)	(0.003)	(0.001)	(0.001)
US	2.30	3.44	2.61	2.25	1.26	2.13
	(0.017)	(0.040)	(0.026)	(0.022)	(0.003)	(0.009)
S Europe	2.44	1.79	1.83	1.27	1.11	1.10
	(0.002)	(0.001)	(0.002)	(0.001)	(0.000)	(0.000)
Retailer	1.23	1.30	1.02	1.03	1.02	1.02
	(0.002)	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)

Note: Estimates use all establishments and years. Standard errors are based on 20 bootstrap simulations. For India, these estimates ignore the sample weights.

$$\log(\mu_{it}^X) = \alpha + \gamma_t + \delta_n + \epsilon_{it} \tag{5}$$

where μ_{it}^X is the markup using input X for establishment i in year t, and γ_t and δ_n are year and industry fixed effects. I then plot the year effects using the translog estimates in Figure 2, with the first year normalized to zero. The red solid lines are the labor markup, the blue dashed lines the materials markup, and the green dash-dot lines the composite input markups.[9]

For all of the datasets, I find *opposing* patterns over time using labor compared to mate-

[9] I include the Cobb-Douglas trends in Figure E.1 in Appendix E.1. I always find significantly different markup trends using different inputs.

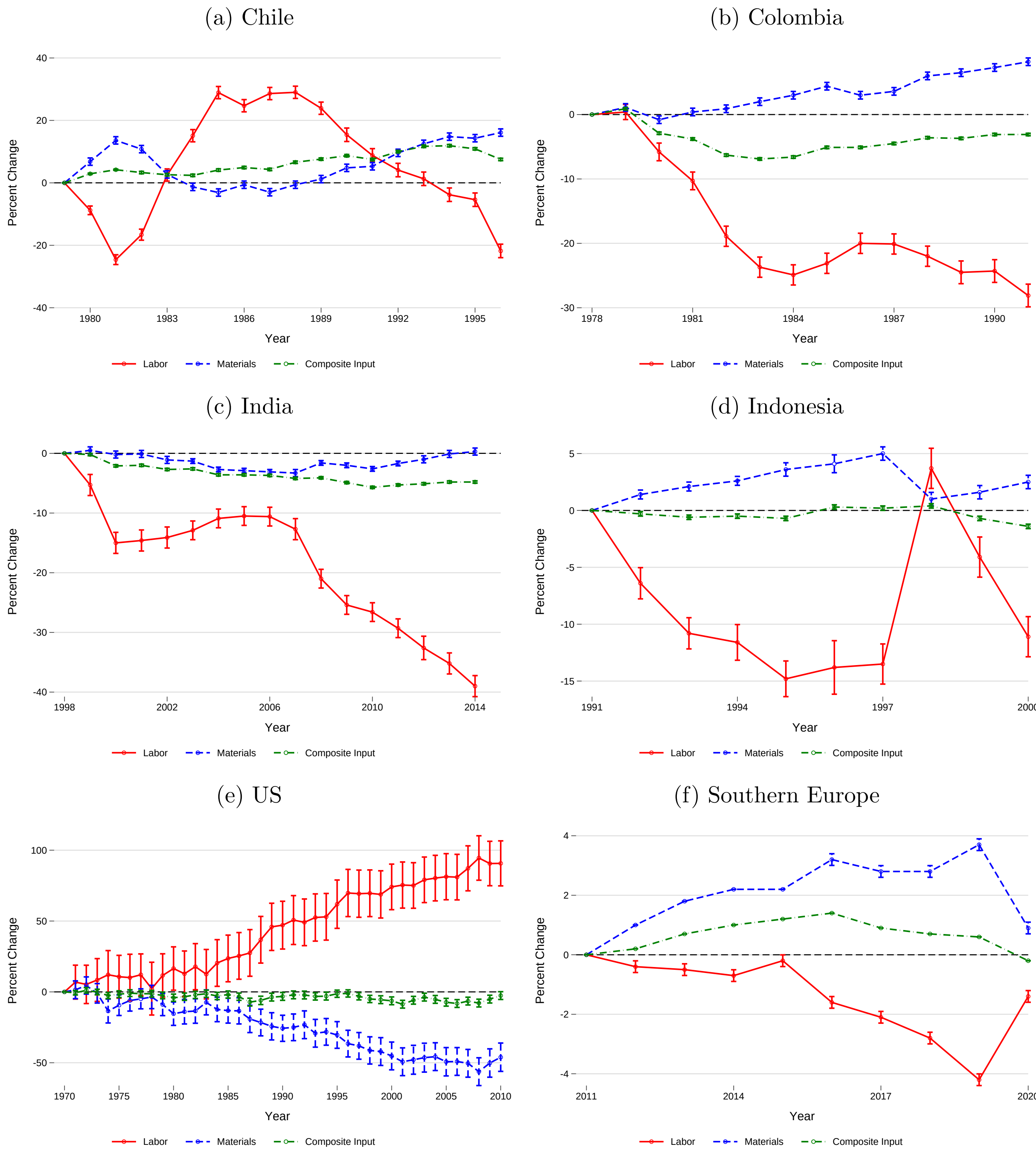
rials to measure the markup. The time trend for composite input markups lie between the two, but much closer to materials, and exhibit less extreme movements.

For example, for Colombia, the average labor markup falls by 28% over the sample, while the average materials markup rises by 8% and the composite input markup declines by 3%. For India, the average labor markup is 39% lower at the end of the sample, while the materials markup exhibits little change and the composite input markup declines by 5%. For Indonesia, the Asian financial crisis strikes in 1998. The average labor markup rises by 17% in 1998, while the average materials markup declines by 4% and the composite input markup remains unchanges. Except for Southern Europe, estimated markups exhibit large changes over the sample period.

A major advantage of the production approach to markup estimation has been the ability to aggregate across producers to estimate the aggregate markup, which is useful to answer many macroeconomic questions, such as reasons for the decline in the labor share. While De Loecker et al. (2020) aggregate by weighting markups by sales, Edmond et al. (2018) argue that the cost weighted markup is the right benchmark for the aggregate markup for welfare calculations of the cost of market power.

I compare changes in the unweighted aggregate markup to a cost aggregated markup in Figure 3. The unweighted average labor markup increases by 91% from 1970 to 2010, compared to a decline of 46% with the materials markup and 3% for the composite input markup. With cost weights, the markup increase is larger; the aggregate labor markup rises

Figure 2 Markup Time Trends using Translog Estimates

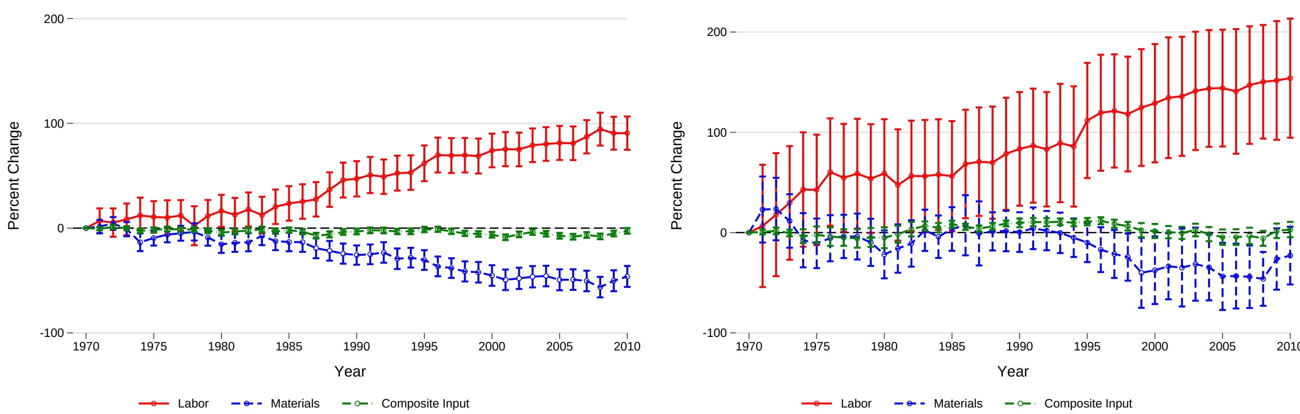


Note: Estimates based on (5), and include 95% Confidence Intervals (vertical bars) based on clustering at the establishment level. All estimates relative to the first year, which is set to zero.

by 154%, while the materials markup declines by 23% and composite input markup rises by 3%. In Web Appendix E.4, I show that using different inputs to estimate markups continues to lead to very different trends after aggregating with cost weights, or with sales weights, for all of the datasets.

Figure 3 Markup Time Trends using Translog Estimates: US

(a) Unweighted (b) Cost-Weighted



Note: Estimates based on (5), and include 95% Confidence Intervals (vertical bars) based on clustering at the establishment level. All estimates relative to the first year, which is set to zero. The left figure weights all firms equally, whereas the right figure weights each firm by its share of cost.

3.3 Correlations of Markup Estimates

Under the production approach, markup estimates using different inputs for the same establishment should be highly correlated with each other. Instead, I find negative correlations between labor and materials markups. For example, in Figure 4, I plot the materials markup on the x-axis against the labor markup on the y-axis for all plants in the Chilean Food Products industry in 1996 using the translog estimates. Each a point is a different manufacturing plant with the best linear fit as a solid black line. There is a slight negative relationship between the labor markup and materials markup.

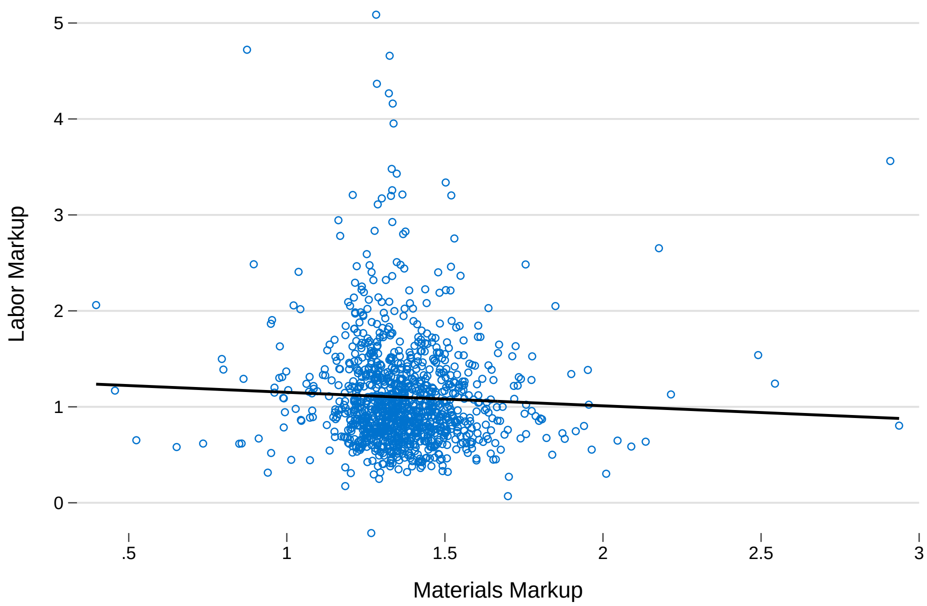


Figure 4 Correlation of Markups for Chilean Food Products, 1996

Note: Each point is the translog markup for a manufacturing plant in Chilean Food Products in 1996; the x-axis is the materials markup and the y-axis is the labor markup. Solid black line is the the best linear fit.

I examine the correlation between markup estimates for all the datasets by estimating the following regression:

$$\log(\mu_{it}^L) = \alpha + \beta \log(\mu_{it}^M) + \gamma_t + \delta_n + \epsilon_{it} \tag{6}$$

where μ_{it}^L and μ_{it}^M are the markups using labor and materials for establishment i in year t.

I also include controls γ_t and δ_n, year and industry fixed effects, so estimated correlations do not reflect the yearly trends discussed in the previous section. In this specification, β represents the elasticity of the markup using labor with respect to the markup using materials.

I report these correlations between markup measures in Table III. The labor and materials markups are *negatively* correlated with each other, the opposite of the relationship implied by the production approach. Under the translog estimates, an establishment with a 1% higher materials markup has, on average, a 0.16% lower labor markup for Chile, 0.28% lower for Colombia, 0.53% lower for India, 0.48% lower for Indonesia, 0.50% for the US, 0.32% for Southern Europe, and 10.08% lower for the Retailer. In general, the magnitude of the negative correlation is even higher using the Cobb-Douglas estimates.[10]

3.4 Energy

One potential explanation for these findings is that labor is not a flexible input. The literature suggests that violations of the static first order conditions are likely to be more severe for labor (Dobbelaere and Mairesse, 2013), either due to hiring and firing costs when adjusting labor (Petrin and Sivadasan, 2013), bargaining with unions, or labor monopsony power.[11]

[10]The large magnitude of the elasticities for the retailer is due to the measurement error correction to the input share of revenue as in (15), because the estimated measurement error in sales is negatively correlated with the materials share of revenue. If I ignore this correction, the elasticity between the labor and materials markup is -1 for the Cobb-Douglas case and -2.3 for the translog case.

[11]Union bargaining under a "right to manage" model, in which bargaining is over the wage but the firm can freely choose the number of workers, does not violate my baseline approach. See Nickell and Andrews (1983) and Dobbelaere and Mairesse (2013).

Table III Relationship between Markup Estimates

Dataset	Cobb-Douglas	Translog
Chile	-0.66	-0.16
	(0.017)	(0.014)
Colombia	-0.99	-0.28
	(0.015)	(0.021)
India	-1.73	-0.53
	(0.012)	(0.009)
Indonesia	-0.97	-0.48
	(0.018)	(0.021)
US	-0.20	-0.50
	(0.032)	(0.029)
S Europe	-0.80	-0.32
	(0.005)	(0.008)
Retailer	-7.51	-10.08
	(0.143)	(0.102)

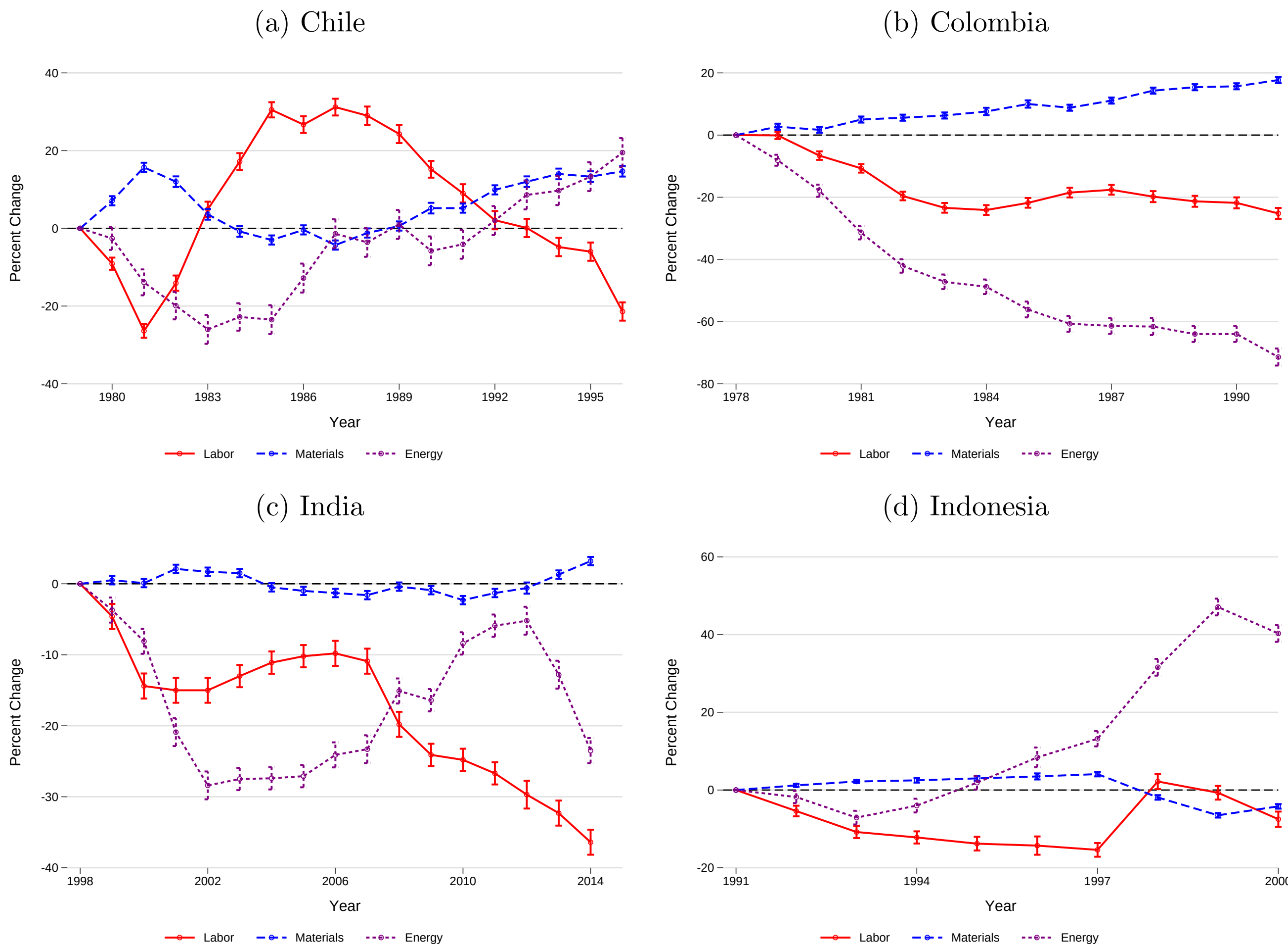
Note: Estimates based on (6) where the labor markup is the dependent variable and materials markup the independent variable. Standard errors are clustered at the establishment level.

Thus, for the four manufacturing censuses, I separate materials into raw materials and energy, where energy includes both electricity and fuel expenditure, and examine markups for each separately. Both raw materials and energy should be robust to labor-specific violations of the static cost minimization conditions. I then estimate production functions with capital, labor, raw materials, and energy as separate flexible inputs.

I examine time trends separating raw materials and energy estimated using (5). I depict the translog estimates in Figure 5, and the Cobb-Douglas figures in Figure E.2. In all four datasets, the raw materials markup has a different time trend than the energy markup.

I report correlations between markup estimates using (6) in Table IV; for example, "Labor on Energy" indicates that the (logged) labor markup is the dependent variable and energy

Figure 5 Markup Time Trends using Translog Estimates, with Energy



Note: Estimates based on (5), and include 95% Confidence Intervals (vertical bars) based on clustering at the establishment level. All estimates relative to the first year, which is set to zero.

markup the independent variable. Neither the labor or raw materials markup is highly correlated with the energy markup. The raw materials markup is uncorrelated with the energy markup using the translog estimates. The labor markup is negatively correlated with the energy markup under the translog estimates, with a 0% increase in the energy markup leads, on average, to a 0.02% to 0.1% decline in the labor markup. These findings are inconsistent with purely labor-specific violations of the cost minimization conditions.

Table IV Relationship between Markup Estimates: Energy and Raw Materials Separated

	Labor on Raw Materials		Labor on Energy		Raw Materials on Energy	
Dataset	Cobb-Douglas	Translog	Cobb-Douglas	Translog	Cobb-Douglas	Translog
Chile	-0.60	-0.05	0.21	-0.08	-0.13	-0.01
	(0.017)	(0.013)	(0.008)	(0.006)	(0.003)	(0.002)
Colombia	-0.71	-0.05	0.16	-0.05	-0.26	0.00
	(0.014)	(0.011)	(0.006)	(0.005)	(0.006)	(0.003)
India	-1.38	-0.32	0.28	-0.12	-0.11	0.00
	(0.019)	(0.008)	(0.003)	(0.003)	(0.001)	(0.001)
Indonesia	-0.75	-0.18	0.16	-0.10	-0.14	0.01
	(0.023)	(0.019)	(0.005)	(0.006)	(0.002)	(0.002)

Note: Estimates based on (6) for markups from two flexible inputs, so Labor on Raw Materials indicates a regression where the labor markup is the dependent variable and raw materials markup the independent variable. Standard errors are clustered at the establishment level.

3.5 Robustness

In Web Appendix C, I show that the large, substantive differences between markups estimated with different inputs demonstrated in this section are robust to several additional specifications. First, these patterns are robust to using several alternative production function estimators assuming neutral productivity, including a dynamic panel approach (Blundell and Bond, 2000), an alternative control function approach (Flynn et al., 2019), and a industry-year level cost share approach. Second, these patterns continue to hold estimating production functions at the subindustry or product level. Third, I find similar patterns estimating quantity rather than revenue production functions using a set of Indian homogenous products. Finally, the data patterns are not consistent with measurement error explanations.

4 Non-Neutral Productivity and Markups

Why are markups estimated using labor so different than those estimated using materials? Dividing the labor markup by the materials markup, we have[12]:

$$\frac{\hat{\mu}_{it}^{L}}{\hat{\mu}_{it}^{M}} = \frac{\beta_{it}^{L}}{\beta_{it}^{M}}\frac{s_{it}^{M}}{s_{it}^{L}} = \frac{\beta_{it}^{L}}{\beta_{it}^{M}}(\frac{w_{it}L_{it}}{p_{it}^{m}M_{it}})^{-1}, \tag{7}$$

so the ratio of markups is equal to the ratio of output elasticities divided by the labor cost to materials cost ratio. To estimate the same markup using different inputs, the labor cost to materials cost ratio has to equal the ratio of output elasticities. For a Cobb-Douglas production function, the ratio of output elasticities is a constant ($\frac{\beta_l}{\beta_m}$), so any differences in the labor to materials cost ratio across plants violate this equality and will show up as differences in estimated markups. A translog production function allows the output elasticities to vary based upon inputs, but they remain a deterministic function of production

[12]I would like to thank an anonymous referee for suggesting this framing of the problem.

parameters and inputs with no error term. Thus, the translog estimated output elasticities cannot capture the full degree of heterogeneity in input shares across plants.[13]

So far, I have assumed Hicks neutral productivity, so productivity differences do not affect output elasticities. Below, I show that non-neutral productivity differences can fully capture the heterogeneity in input shares across plants by providing an error term in the ratio of output elasticities.

4.1 Theory

In order to allow productivity to be non-neutral, I assume a CES production function with elasticity of substitution σ, neutral productivity A_{it}, labor augmenting productivity B_{it}, and distribution parameters α_l and α_m:

$$F_{it} = A_{it}((1-\alpha_l-\alpha_m)K_{it}^{\frac{\sigma-1}{\sigma}} + \alpha_l(B_{it}L_{it})^{\frac{\sigma-1}{\sigma}} + \alpha_m M_{it}^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}}. \tag{8}$$

Input revenue shares are equal to the input's output elasticity divided by the markup μ_{it}:

$$\frac{w_{it}L_{it}}{P_{it}F_{it}} = \frac{1}{\mu_{it}}(\frac{w_{it}}{\lambda_{it}A_{it}})^{1-\sigma}(\alpha_l)^{\sigma}(B_{it})^{\sigma-1} \tag{9}$$

$$\frac{p_{it}^m M_{it}}{P_{it}F_{it}} = \frac{1}{\mu_{it}}(\frac{p_{it}^m}{\lambda_{it}A_{it}})^{1-\sigma}(\alpha_m)^{\sigma} \tag{10}$$

[13]The ratio of output elasticities for the translog is:

$$\frac{\beta_l + 2\beta_{ll}l_{it} + \beta_{kl}k_{it} + \beta_{ml}m_{it}}{\beta_m + 2\beta_{mm}m_{it} + \beta_{km}k_{it} + \beta_{lm}l_{it}}.$$

where λ_{it} is the marginal cost, w_{it} the wage, and p^m_{it} the price of materials. An increase in neutral productivity A_{it} does not affect input shares of revenue, as the marginal cost λ_{it} falls to exactly compensate.

Labor augmenting productivity, in contrast, does affect input shares of revenue. An increase in B_{it} is akin to more labor. Thus, after an increase in B_{it}, a firm will increase materials M_{it} to exactly match the increase in effective labor $B_{it}L_{it}$. However, the increase in B_{it} also reduces the cost of an efficient unit of labor, which is $\frac{w_{it}}{B_{it}}$. The plant will then substitute towards relatively cheaper labor, with the ratio of effective labor to materials $\frac{B_{it}L_{it}}{M_{it}}$ changing by σ given the change in the ratio of prices $(w_{it}/B_{it})/p^m_{it}$. Hence the labor cost to materials cost ratio $w_{it}L/p^m_{it}M_{it}$ decreases 1 by a direct effect and increases σ by a substitution effect when B_{it} increases.

When inputs are gross complements, as estimated in Doraszelski and Jaumandreu (2018) and Raval (2019), σ is less than one and so the direct effect is stronger than the substitution effect. A plant with higher labor augmenting productivity will then have a lower labor share, higher materials share, and lower labor cost to materials cost ratio.

Thus, changes in labor augmenting productivity B_{it} move the output elasticities of labor and materials in different directions. In the case when σ is less than one, improvements in B_{it} decrease labor's output elasticity, but increase materials's output elasticity as the marginal cost of production λ falls. If production function estimates ignore labor augmenting productivity differences, a plant with a higher B_{it} would have a lower labor share and higher materials share, and so a higher labor markup and lower materials markup. Estimated markups estimated using different inputs would be negatively correlated.

In Web Appendix F, I develop a Monte Carlo in which firms have different labor augmenting productivities and set different markups. In the Monte Carlo, control function estimators assuming neutral productivity imply negatively correlated labor and materials markups. In addition, neither labor or materials markups are highly correlated with the true markup.

The revenue share of the composite input of labor and materials is equal to the sum of the right hand side of (9) and (10). Thus, the output elasticity of the combined input also declines as labor augmenting productivity B_{it} increases, albeit less than the labor output elasticity.[14] If not accounted for, labor augmenting productivity differences will also bias the composite input markup. For example, improvements in labor augmenting productivity over time, all other things equal, would decrease the output elasticity of the composite input, and so increase aggregate markups.

4.2 Alternative Explanations

Given (9) and (10), the labor cost to materials cost ratio is:

$$\frac{w_{it}L_{it}}{p^m_{it}M_{it}} = (\frac{w_{it}}{p^m_{it}})^{1-\sigma}(\frac{\alpha_l}{\alpha_m})^{\sigma}(B_{it})^{\sigma-1},$$

[14]The composite input elasticity is $\frac{(w_{it}/B_{it})^{1-\sigma}\alpha_l^{\sigma}+(p^m_{it})^{1-\sigma}\alpha_m^{\sigma}}{(\lambda_{it}A_{it})^{1-\sigma}}$.

so non-neutral productivity differences B_{it} can clearly soak up any differences in the labor cost to materials cost ratio across plants. However, unobserved differences in input prices $\frac{w_{it}}{p^m_{it}}$ could also lead to differences in factor costs (Grieco et al., 2016), as could monopsonistic behavior by firms that would show up as "wedges" in input first order conditions. Finally, measurement errors in inputs would also lead to differences in factor cost ratios across plants.

In general, it is difficult to assess whether differences in factor cost ratios across plants are due to productivity differences as opposed to these other factors, as we typically do not observe the plant's input prices or its degree of monopsony power. Manufacturing survey data also contain a substantial amount of measurement error (White et al., 2016).

With the retailer, however, I can control for these alternative explanations. First, the retailer's data is based on the internal records of the firm and so should have very little measurement error compared to self-reported survey data. Second, I have detailed data on the store location that allows me to control for local labor markets; the same chain's input prices and monopsony power should not vary within a local labor market. I control for local labor markets through two types of location-year fixed effects. First, I use the MSA of the store, defined as either the Metropolitan Statistical Area or Micropolitan Statistical Area of the retail store's location.[15] Second, I use the store's district as defined by the internal structure of the retailer; each district has about 10 to 20 stores.

I then use the translog estimates and re-estimate (6) after controlling for either MSA-year

[15]For retail stores not located in a Metropolitan Statistical Area or Micropolitan Statistical Area, the fixed effect is for all non-MSA locations in the same state.

or district-year fixed effects. A retail store with a 100% higher materials markup has, on average, a 1008% lower labor markup in the previous, baseline estimates, compared to 992% lower with MSA-year fixed effects and 1006% lower with district-year fixed effects. Because the negative correlation between labor and materials markups remains almost unchanged after controlling for local labor markets, differences in input prices or monoposony power are unlikely to explain the patterns that I find for the retail store.

On the other hand, there are several reasons why stores in the same retail chain may vary in productivity. First, each store is run by its own store manager, and differences in managerial ability are well known to affect productivity (Bloom and Van Reenen, 2007; Fenizia, 2022), such as by affecting staff turnover (Hoffman and Tadelis, 2021). Second, each retail store is a multi-product producer, with each section of the store using its own production function and potentially different types of labor; stores vary in the importance of components of the store.[16] Finally, this retail chain operates some 24-hour stores, which will have a different production technology than stores open during more standard hours.

4.3 Estimation

How can we estimate the production function given non-neutral productivity differences? One possibility is to directly estimate first order conditions such as (9) and (10) using variation in factor prices across firms and then recover the elasticity of substitution and

[16] This type of heterogeneity is common in retail. For example, Walmart Superstores have a much larger grocery component than standard Walmart stores.

productivities. For example, Raval (2019) exploits cross-sectional variation in wages across US locations together with several instruments for such wages – local amenities and labor demand instruments from Bartik (1991) and Beaudry et al. (2012) – for identification.

However, the researcher often does not have exogenous variation in input prices, or wants to assume a different production function than the CES. Thus, three new approaches modify existing methods of production function estimation to allow non-neutral productivity.

First, Doraszelski and Jaumandreu (2019) provide a new dynamic panel estimator for markups. The estimator assumes a translog production function in which the combined output elasticity of materials and labor is constant, quantity produced is observable, and neutral productivity follows an autoregressive process. Log quantity produced is then a linear function of its lag and differences in inputs and input shares, and model parameters can be estimated via GMM. Output elasticities are measured with error (similar to Klette (1999)), so individual markups are recovered up to this error.

Second, Demirer (2020) develops a non-parametric control function approach to estimate output elasticities given non-neutral productivity. This approach requires only homothetic separability between labor and materials, together with input choice timing assumptions. As in (7), the ratio of input costs identifies the ratio of output elasticities. Demirer (2020) uses the materials-labor ratio equation to form a control function for labor augmenting productivity, and then the materials demand equation to form a control function for neutral productivity.

Lastly, Raval (2022) generalizes the cost share approach to markup estimation to account for labor augmenting productivity differences. This flexible cost share estimator assumes that cost minimization conditions hold for all inputs on average and returns to scale are constant. Because labor augmenting productivity is proportional to the ratio of labor to materials costs, Raval (2022) groups plants into bins with similar labor augmenting productivities based upon this ratio and estimates output elasticities from cost shares within each group.

The attractiveness of these approaches will depend upon which assumptions the econometrician is comfortable with placing on the data. For example, Doraszelski and Jaumandreu (2019) and Demirer (2020) estimate an output equation and so require data on quantity produced, whereas Raval (2022) does not require quantity data as he estimates output elasticities only using cost data. Doraszelski and Jaumandreu (2019) and Raval (2022) assume all plants have the same variable returns to scale and total returns to scale, respectively. And Raval (2022) implicitly places strong assumptions on the adjustment process of capital.

5 Conclusion

A key advantage of the production approach to estimating markups is that it allows one to estimate markups across widely differing industries, and thus estimate the aggregate markup. However, the production approach, as currently implemented, delivers very different markups using alternative flexible inputs. Labor markups are negatively correlated with materials markups, have opposing time trends, and are much more disperse. Non-neutral technological differences across plants can explain these findings, as ignoring such differences will lead to negatively correlated labor and materials markups.

The development of the parallel demand approach to markup estimation provides guidance on how to measure markups going forward. The demand approach focuses on modeling the heterogeneity in preferences across consumers; for example, Berry et al. (1995) estimate random coefficients that allow consumers to vary in their sensitivity to price. In order to use the production approach, economists will have to allow more heterogeneity in technology.

Data Availability Statement: The data underlying this article are available on Zenodo at `https://doi.org/10.5281/zenodo.7473821`.

References

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica 83*(6), 2411–2451.

Bartik, T. J. (1991). *Who Benefits from State and Local Economic Development Policies?* W. E. Upjohn Institute for Employment Research.

Basu, S. (2019, August). Are price-cost markups rising in the united states? a discussion of the evidence. *Journal of Economic Perspectives 33*(3), 3–22.

Basu, S. and J. G. Fernald (1997). Returns to scale in us production: Estimates and implications. *Journal of political economy 105*(2), 249–283.

Beaudry, P., D. A. Green, and B. Sand (2012). Does industrial composition matter for wages? a test of search and bargaining theory. *Econometrica 80*(3), 1063–1104.

Berry, S., M. Gaynor, and F. Scott Morton (2019, August). Do increasing markups matter? lessons from empirical industrial organization. *Journal of Economic Perspectives 33*(3), 44–68.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–890.

Blonigen, B. A. and J. R. Pierce (2016). Evidence for the effects of mergers on market power and efficiency. Technical report, National Bureau of Economic Research.

Bloom, N. and J. Van Reenen (2007). Measuring and explaining management practices across firms and countries. *The quarterly journal of Economics 122*(4), 1351–1408.

Blundell, R. and S. Bond (2000). Gmm estimation with persistent panel data: An application to production functions. *Econometric Reviews 19*(3), 321–340.

Bridgman, B. and B. Herrendorf (2022). Labor share, markups, and input-output linkages: Evidence from the national accounts. *Mimeo*.

De Loecker, J. and J. Eeckhout (2018). Global market power. Technical report, National Bureau of Economic Research.

De Loecker, J., J. Eeckhout, and G. Unger (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics 135*(2), 561–644.

De Loecker, J., P. K. Goldberg, A. K. Khandelwal, and N. Pavcnik (2016). Prices, markups, and trade reform. *Econometrica 84*(2), 445–510.

De Loecker, J. and P. T. Scott (2017). Estimating market power: Evidence from the us brewing industry. Technical report, US Census Bureau, Center for Economic Studies.

De Loecker, J. and F. Warzynski (2012). Markups and firm-level export status. *American Economic Review 102*(6), 2437–71.

Demirer, M. (2020). Production function estimation with factor-augmenting technology: An application to markups. Technical report, Mimeo.

Diewert, W. E. and D. A. Lawrence (2000). Progress in measuring the price and quantity of capital. *Econometrics 2*, 273–326.

Dobbelaere, S. and J. Mairesse (2013). Panel data estimates of the production function and product and labor market imperfections. *Journal of Applied Econometrics 28*(1), 1–46.

Domowitz, I., R. G. Hubbard, and B. C. Petersen (1988). Market structure and cyclical fluctuations in us manufacturing. *The Review of Economics and Statistics*, 55–66.

Doraszelski, U. and J. Jaumandreu (2018). Measuring the bias of technological change. *Journal of Political Economy 126*(3), 1027–1084.

Doraszelski, U. and J. Jaumandreu (2019). Using cost minimization to estimate markups.

Edmond, C., V. Midrigan, and D. Y. Xu (2018). How costly are markups? Technical report, National Bureau of Economic Research.

Fenizia, A. (2022). Managers and productivity in the public sector. *Econometrica 90*(3), 1063–1084.

Flynn, Z., A. Gandhi, and J. Traina (2019). Measuring markups with production data.

Gandhi, A., S. Navarro, and D. A. Rivers (2020). On the identification of gross output production functions. *Journal of Political Economy 128*(8), 2973–3016.

Grieco, P. L., S. Li, and H. Zhang (2016). Production function estimation with unobserved input price dispersion. *International Economic Review 57*(2), 665–690.

Grossman, G. M. and E. Oberfield (2022). The elusive explanation for the declining labor share. *Annual Review of Economics 14*(1), null.

Hall, R. E. (1988). The relation between price and marginal cost in us industry. *Journal of Political Economy 96*(5), 921–947.

Harper, M. J., E. R. Berndt, and D. O. Wood (1989). Rates of return and capital aggregation using alternative rental prices. In D. Jorgenson and R. London (Eds.), *Technology and Capital Formation*. Cambridge, MA: MIT Press.

Hoffman, M. and S. Tadelis (2021). People management skills, employee attrition, and manager rewards: An empirical analysis. *Journal of Political Economy 129*(1), 243–285.

Karabarbounis, L. and B. Neiman (2018). Accounting for factorless income. Technical report, National Bureau of Economic Research.

Keller, W. and S. R. Yeaple (2009). Multinational enterprises, international trade, and productivity growth: firm-level evidence from the united states. *The review of economics and statistics 91*(4), 821–831.

Klette, T. J. (1999). Market power, scale economies and productivity: estimates from a panel of establishment data. *The Journal of Industrial Economics 47*(4), 451–476.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies 70*(2), 317–341.

Nickell, S. and M. Andrews (1983). Unions, real wages, and employment in britain 1951-1979. *Oxford Economic Papers 35*(0), 183–206.

Oberfield, E. and D. Raval (2021). Micro data and macro technology. *Econometrica 89*(2), 703–732.

Petrin, A. and J. Sivadasan (2013). Estimating lost output from allocative inefficiency, with an application to chile and firing costs. *Review of Economics and Statistics 95*(1), 286–301.

Raval, D. (2019). The micro elasticity of substitution and non-neutral technology. *RAND Journal of Economics 50*(1), 147–167.

Raval, D. (2022). A flexible cost share approach to markup estimation.

Syverson, C. (2019, August). Macroeconomics and market power: Context, implications, and open questions. *Journal of Economic Perspectives 33*(3), 23–43.

Traina, J. (2018). Is aggregate market power increasing? production trends using financial statements.

White, T. K., J. P. Reiter, and A. Petrin (2016). Imputation in us manufacturing data and its implications for productivity dispersion. *Review of Economics and Statistics* (0).

Zhang, H. (2019). Non-neutral technology, firm heterogeneity, and labor demand. *Journal of Development Economics*.

A Data

For each dataset used in this paper, I construct capital, labor, materials, and sales at the establishment-year level. I provide further details on data construction in Web Appendix G.

An establishment is a manufacturing plant for the Chilean, Colombian, and Indian data, a firm for the Indonesian, US, and Southern European data, and a retail store for the retailer. India is a census for plants above 100 employees and a sample for smaller plants; I thus use the provided sampling weights.

I use capital, materials, and output deflators in order to construct consistent measures of inputs and outputs over time, and drop any observations with zero or negative capital, labor, materials, sales, or labor costs. I also drop observations in the bottom 1% and top 1% of labor's share of revenue, materials's share of revenue, and the composite variable input share of revenue for each industry to remove outliers.

For labor, I use the number of workers for Chile, Colombia, Indonesia, US, and Southern Europe, the number of manufacturing worker-days for India, and the total number of hours worked for the retailer. Labor costs are the total of salaries and worker benefits.

For materials, I include expenses for raw materials, electricity, and fuels for the manufacturing censuses. For Southern Europe, materials is total materials costs, while for the retailer, materials is the sum of the cost of goods sold across different parts of the store. The composite variable input is the sum of materials and labor costs.[17]

For the US, I do not have separate data fields on labor costs and materials costs. Instead, I follow Keller and Yeaple (2009) and Demirer (2020) and estimate labor costs as total employees multiplied by an industry-level average wage, and materials costs as cost of goods sold and selling, general, and administrative expenses minus depreciation and labor costs.

For capital, where possible, I construct a perpetual inventory measure of capital for each type of capital and rental rates of capital based on an average real interest rate over time plus depreciation for that type of capital. My measure of capital is then the sum of capital stocks times their rental rates, plus any rental payments for capital.[18]

For the manufacturing datasets, I estimate production functions at the industry level, defined at a similar level to two digit US SIC (i.e., Chilean Food Products).[19] Given the smaller size of the US data, I follow De Loecker et al. (2020) and estimate production functions at the 2 digit NAICS

[17]I deflate this input using the output deflator to match De Loecker et al. (2020)'s treatment of cost of goods sold.

[18]This provides an approximation to a Divisia index for capital given different types of capital. See Diewert and Lawrence (2000) and Harper et al. (1989) for details on capital rental rates and aggregation. For the US, I do not have data on separate types of capital, and use rental rates from Oberfield and Raval (2021). For Southern Europe, I only have book values of capital. For the retailer, I use BLS rental rates for retail trade. See Web Appendix G for more details on capital construction.

[19]For Chile, Colombia, and Indonesia this is at the three digit ISIC (Rev.2) level, for India at the two digit NIC 08 level, and for Southern Europe at the 2 digit NACE level. Estimating production functions at this level of aggregation is consistent with the production function literature, such as Levinsohn and Petrin (2003) or Gandhi et al. (2020).

level, so there are three industries. I only include industries with at least 1,000 observations over the entire dataset. For the retailer, I estimate a single production function across all retail outlets.

B Estimation

Given (4), estimating the markup requires the input share of revenue and the output elasticity of that input. The input share of revenue, defined as costs for input X divided by total firm revenue, is observed. However, the production function has to be estimated to recover output elasticities. I describe below how De Loecker and Warzynski (2012), and subsequent papers using the production approach such as De Loecker et al. (2020), address this estimation challenge using a control function approach that assumes productivity is Hicks neutral.

B.1 Production Functions

All lower case variables are in logged form, so k_{it} is capital, l_{it} labor, and m_{it} materials. For the Cobb Douglas production function with labor and materials, the (logged) production function excluding the Hicks neutral productivity term is:

$$f_{it} = \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it}$$

and so the output elasticity for input X is simply β_X. For the translog production function, the production function is:

$$f_{it} = \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it} + \beta_{kk} k_{it}^2 + \beta_{ll} l_{it}^2 + \beta_{mm} m_{it}^2 + \beta_{kl} k_{it} l_{it} + \beta_{km} k_{it} m_{it} + \beta_{lm} l_{it} m_{it}$$

and so the output elasticity for each input will depend upon the level of all inputs. For example, the firm's output elasticity for materials would be $\beta_m + 2\beta_{mm} m_{it} + \beta_{km} k_{it} + \beta_{lm} l_{it}$.

For both the Cobb-Douglas and translog production functions, the production function coefficients are not time-varying. However, for the translog, output elasticities can vary over time due to changes in factors.

B.2 Control Function Estimation

I follow De Loecker and Warzynski (2012) and use the Ackerberg et al. (2015) (ACF) estimator for my baseline estimates. The ACF estimator imposes substantial additional assumptions on productivity, including that productivity is Hicks neutral and evolves following a Markov process. In addition, it requires a set of timing assumptions where at least one input is decided at the time the firm learns its productivity shock. I discuss problems with this estimator, and alternative estimation approaches using neutral productivity, in Web Appendix C.1.

The control function approach assumes that observed revenue includes additive measurement error ϵ_{it}. Thus, given log productivity ω_{it}, measured log revenue y_{it} is:

$$y_{it} = f(k_{it}, l_{it}, m_{it}) + \omega_{it} + \epsilon_{it}. \tag{11}$$

Let materials be the flexible input decided at the time the firm learns its productivity shock. Materials is then a function of the observed inputs and productivity $m_{it} = g(k_{it}, l_{it}, \omega_{it})$. It can then be inverted for productivity, so $\omega_{it} = g^{-1}(k_{it}, l_{it}, m_{it})$.[20]

The first stage of the ACF estimator controls for a flexible form of the inputs to recover the additive measurement error ϵ_{it}. Formally, measured log revenue y_{it} is:

$$y_{it} = f(k_{it}, l_{it}, m_{it}) + g^{-1}(k_{it}, l_{it}, m_{it}) + \epsilon_{it} = h(k_{it}, l_{it}, m_{it}) + \epsilon_{it} \tag{12}$$

Since both the production function and productivity are functions of the inputs, they cannot be separated in the first stage. Instead, the nonparametric function h includes both productivity ω_{it} and the production function f. The measurement error in sales ϵ_{it} is a residual in the first stage equation after controlling for h.[21]

The second major assumption of the ACF approach is that productivity follows a first order Markov process. In my implementation, I further assume an AR(1) process. Formally,

$$\omega_{it} = \rho\omega_{it-1} + \nu_{it} \tag{13}$$

with AR(1) coefficient ρ and productivity innovation ν_{it}. In that case, given knowledge of the production function coefficients β, one can recover the innovation in productivity ν_{it} as:

$$\nu_{it}(\beta) = \omega_{it} - \rho\omega_{it-1} \tag{14}$$

The innovation in productivity is a function of production coefficients β because $\omega_{it} = y_{it} - \epsilon_{it} - f_{it}(\beta)$, and ϵ_{it} was recovered in the first stage.

Because the innovation in productivity is, by construction, independent of inputs chosen before time t, moments of the innovations multiplied by inputs chosen before the productivity innovation, such as $E(\nu_{it} l_{it-1})$ or $E(\nu_{it} k_{it})$, identify the production function coefficients.

For the Cobb-Douglas production function, I use capital and the first lag of materials and labor as instruments. For the translog, I use capital and the first lag of materials and labor, as well as their interactions, as instruments.[22]

Finally, I follow De Loecker and Warzynski (2012) and correct the value of sales in the input share of revenue for the measurement error estimated in the first stage. Thus, for input X, the estimate of the markup is:

$$\hat{\mu_{it}} = \frac{\hat{\beta_i^X}}{s_{it}^X exp(\hat{\epsilon}_{it})}. \tag{15}$$

[20]The g() function can include other determinants of materials as well, such as materials prices.

[21]In practice, I use a third order polynomial in inputs for the function h, and also control for year effects.

[22]For the specification with the composite variable input instead of labor and materials separately, I use the lag of the composite input and its interactions as instruments, symmetrically to the case above.