Multi-Product Markups

John Cairncross
University of Toronto
john.cairncross@mail.utoronto.ca

Peter Morrow
University of Toronto
peter.morrow@utoronto.ca

Scott Orr
UBC Sauder School of Business
scott.orr@sauder.ubc.ca

Swapnika Rachapalli*
UBC Sauder School of Business
s.rachapalli@sauder.ubc.ca

June 27, 2023

Abstract

We examine the identification of firm-product markups in multi-product firms using production-side data. Identifying within-firm markup differences relies on identifying the marginal rates of transformation across goods. Since marginal rates of transformation are generally functions of (i) the degree of joint production and (ii) the magnitude of within-firm productivity differences, we explore whether markup estimates are sensitive to misspecification in these factors. Monte Carlo exercises indicate that misspecification of within-firm productivity differences is particularly likely to generate bias. However, a weighted average of firm-product level markups (the "firm markup") can be identified without information on these factors, and nests the firm markup of De Loecker and Warzynski (2012). The firm markup can be estimated using standard empirical methods. Standard parametric restrictions on the marginal rates of transformation across goods often deliver implausible estimates of plant-product markups. Firm markups, which do not require these restrictions, are more well behaved. We also discuss the welfare properties of the firm markup.

*We thank Matthew Backus, Germán Bet, Bernardo Blum, Matilde Bombardini, Cecile Gaubert, Gunnar Heins, Volodymyr Lugovsky, Devesh Raval, David Rivers, Valerie Smeets, Frederic Warzynski, as well as seminar participants at Aarhus, Berkeley, Columbia GSB, University of British Columbia-Sauder, University of Florida, University of Toronto, the Canadian Economics Association Meetings (2021), EARIE (2022), and EIIT (2022), IIOC (2023) and NAPW (2023) for helpful feedback. Any remaining errors are our own.

1 Introduction

In an extremely influential paper, De Loecker and Warzynski (2012) build on Hall (1986) to develop an approach for estimating firm-level markups. This "production-side" approach is valuable because it does not require a researcher to estimate demand nor impose market conduct assumptions, as is common elsewhere in applied industrial organization (e.g. Bresnahan 1989, Berry et al. 1995). While most exercises using De Loecker and Warzynski (2012) provide estimates of markups at the firm or plant level, textbook treatments of market power emphasize market power at the *product level*. While this distinction makes no difference for single-product firms, multi-product firms account for a majority of economic activity in many countries.[1]

Given the importance of multi-product production in modern economies, this paper aims to answer two key questions. First, what features of the production technology does an econometrician need to identify to estimate firm-product markups in multi-product firms? Second, how might we interpret firm-level markups estimated for a multi-product producer using the De Loecker and Warzynski (2012) methodology? This second question is important, as a recent literature asking whether market power has increased relies crucially on this production side approach to estimating markups (De Loecker and Eeckhout, 2018; De Loecker et al., 2020).

To help us answer both questions, we first consider the problem of identifying firm-product-level markups when output data is available at the firm-product level but input data is only available at the firm-level. We refer to such data as *standard production data*.[2] De Loecker et al. (2016) provide a popular solution to this problem based on two key restrictions. First, they require production to be *non-joint*, meaning that each multi-product producer allocates all of their inputs in a mutually exclusive and exhaustive way across their product-line-specific production functions. Second, they require that there are *no within-firm productivity differences*, such that a firm is equally productive at producing all products.

We consider identification of product-level markups in a setting that allows for joint or non-joint production, as well as within-firm productivity heterogeneity. We characterize each firm's technology through an output distance function (e.g. Caves et al. 1982). This approach allows us to represent non-joint technologies, as well as joint production technologies that incorporate economies of scope, a potentially important rationale for multi-product pro-

[1]For example, Bernard et al. (2011) document that 12% of U.S. firms in 2000 that exported more than five products to more than five destinations accounted for more than 90% of export value, and multi-product firms produce roughly two-thirds of aggregate output in our data for India that we discuss below.

[2]While the data used in this paper is reported at the plant-level, we use the terms firms and plants interchangeably in the paper unless otherwise stated.

duction (Baumol et al. 1982). Because within-firm productivity dispersion can be difficult to define in settings involving joint production, we consider technologies that allow for heterogeneity in *marginal rates of transformation*; a natural metric for productivity differences in multi-product firms.

We show that identification of product-level markups using standard production data requires that the researcher know the extent of joint production, as well as the degree of heterogeneity in marginal rates of transformation across firms ("MRT heterogeneity"). We use Monte Carlo simulations to show that abstracting from joint production and (especially) MRT heterogeneity can lead to noticeable differences in estimated markups. Unfortunately, past research in this area has shown that identification of MRT heterogeneity may require that researchers know how firms compete with one another, which already disciplines product-level markups within the firm.[3]

This motivates the second part of our paper which derives an alternate *firm-level markup* that *does not* require the researcher to know the magnitude of MRT heterogeneity or joint production. This object is a weighted average of firm-product markups within the firm, with weights equivalent to product-level cost shares, where costs are allocated to products by considering how much costs would rise (locally) when increasing one output, relative to increasing all outputs by the same proportion. This object is welfare relevant, in that it can be interpreted as a firm-level output wedge which scales down the production of each product within the firm uniformly, decreasing overall welfare relative to what a social planner would choose. Therefore, our firm-markup is a symptom of across-firm misallocation as in Edmond et al. (2015, 2022).

Empirically, the firm-markup is recovered by the De Loecker and Warzynski (2012) formula under reasonable restrictions on the technology. This is important for two reasons. First, our definition of the firm-markup, which is based on a general representation of a firm's technology, provides a coherent lens to interpret the De Loecker and Warzynski (2012) firm markup formula. Second, by building a model of multi-product producers to rationalize the De Loecker and Warzynski (2012) markup when quantity data is available, we address recent criticisms by Bond et al. (2021) that if production technologies are estimated with revenue, then the approach is unlikely to recover the firm's markup. In particular, we show that the procedure developed in De Loecker et al. (2016), which first selects the subsample of single-product firms to estimate a product-level production function, can be used to identify the firm-markup even if there is joint production and within-firm productivity heterogeneity.[4]

[3]See Valmari (2016), Gong and Sickles (2021), Orr (2022). These papers only considered non-joint production settings where MRT heterogeneity ends up being analogous to identifying within-firm productivity dispersion.

[4]We also show that their selection correction can be extended to settings with joint-production as well.

This is important as this approach allows researchers to use quantity data to estimate the relevant technology parameters, while avoiding the complications generated by multi-product firms in estimation. We also show that a cost share estimator (Syverson 2004, Foster et al. 2008, Backus 2020, Raval 2023) can be used to recover the firm markup. Importantly, this approach can be used even when quantity data is not available, and the estimator can be applied to all firms in the sample, including multi-product producers.[5] Consistent with past work in this area (e.g. De Loecker and Syverson 2021), these two approaches rely on non-overlapping restrictions to the specification of a firm's technology and environment.

We close by using Indian Annual Survey of Industries (ASI) data provided by the Indian Ministry of Statistics to compare product-level markups, estimated under various assumptions on MRT heterogeneity and joint production, to plant-level markups. We document that restricting MRT heterogeneity tends to generate a large mass of estimated markups where $\frac{P}{MC} = 0$. We call this a *zeros puzzle*. We show that this puzzle can only partly be resolved by relying on a productivity ladders model following Mayer et al. (2014) that allows for within-firm productivity differences. We also show that the puzzle can be almost completely resolved by relying on revenue shares to allocate inputs across product lines. However, since revenue share input allocation rules impose constant within-firm markups as in Orr (2022), these two results suggest that further work quantifying the degree of MRT heterogeneity is needed to fully pin down reasonable estimates of product-level markups in multi-product plants.

We then compare these results to our plant markups. We find that plant markups are generally quite well behaved, and do not suffer from the zeros puzzle. We also find that relying on cost share estimator, or an estimator based on single product plants with a selection correction, yield remarkably similar estimates of the distribution of plant-markups.

Contribution to the Literature

First and foremost, our paper contributes to the vast and growing literature on estimating markups using production data. The paper's core contributions are to two separate strands of this literature. First, we build on past work by De Loecker et al. (2016), Grieco et al. (2016), and Dhyne et al. (2022) on how to approach estimation of markups using production-side data when firms produce many products. On the theoretical front, we propose and study a class of technologies that nest both single- and multi-product production, which allows us

[5]This also allows our approach to deal with critiques raised by Flynn et al. (2019), Doraszelski and Jaumandreu (2019), and Bond et al. (2021) concerning the internal consistency of proxy-variable approach to production function estimation in the presence of market power, and also allows us to construct estimators that can deal with the issues documented in Raval (2022, 2023).

to explore what features of a firm's technology can be identified by a (selected) sample of single-product firms. This allows us to extend the identification results in De Loecker et al. (2016) to settings with joint production. Our identification results for the firm-markup also extends insights in De Loecker and Warzynski (2012) on the interpretation of firm markups. In particular, they note that when firms sell products to many markets, their firm-level markup can be interpreted as an input expenditure share weighted average of market specific markups. We provide a generalization of this result to joint production settings, where input allocations may not be well defined. This generalization relies on the insight that one can obtain product-level cost shares by considering the ratio of local cost increases for increasing a given output, relative to the local cost increase of increasing all outputs by the same proportion. We further show that this firm-level object is welfare relevant, and can be used to quantify the degree of across-firm misallocation, taking *within-firm* misallocation as given. Finally, our Monte Carlo exercises, as well as out empirical results, illustrate that accounting for MRT heterogeneity in multi-product firms may be of first-order importance for identifying product-level markups.

We also contribute to a second strand of the literature on production-side markups, focused on clarifying the data requirements and assumptions necessary to properly identify markups. Bond et al. (2021) point out that standard approaches based on estimating revenue production functions, rather than quantity production functions, are unlikely to identify markups. In response, Kasahara and Sugita (2020), Kirov and Traina (2021) have proposed alternative methods that will still identify the markup in settings where only revenue is observed, although they restrict attention to single-product firms. De Ridder et al. (2022) also consider a single-product firm environment, and argue that even though revenue production functions generate bias, the recovered markup may still be strongly correlated with true markups. Our paper avoids the Bond et al. (2021) critique by embracing the fact that production-side markups will be correctly estimated as long as one can estimate the relevant *quantity* elasticities. This data is becoming increasingly available (e.g. Foster et al. 2008, Baldwin and Gu 2009, De Loecker et al. 2016, Blum et al. 2018, De Roux et al. 2021, Orr 2022, Dhyne et al. 2023), and the techniques developed in this paper provide a useful guide for how to obtain these relevant elasticities in practice. We do this by addressing a specific data challenge generated by quantity data — the inability to allocate inputs to product lines — and provide multiple ways to recover firm markups in this setting.

This paper contributes to the literature on specifying multi-product production technologies. We provide what we believe to be a novel articulation, building on insights in Samuelson (1966) and Hall (1973), of how different specifications of a firm's production possibility set can capture joint or non-joint production. We also provide a representation of a firm's technology through it's output distance function that is dual to the CES cost function for multi-product firms studied by Baumol et al. (1982), and estimated in other applied studies (e.g. Johnes 1997). We derive this representation of a firm's technology through an input allocation problem, where firms can use their inputs to produce private, rival intermediates, as well as public, non-rival intermediates, that contribute to the production of all products simultaneously.[6] This particular specification of a firm's technology ends up having a similar structure to the class of production models considered in Eslava et al. (2023) and De Roux et al. (2021), except that we rely on an entirely production-side focused derivation of the technology, with no reference to consumer preferences. We believe that this particular specification of a firm's technology will be useful to further applied work, as it provides a convenient way to nest both joint and non-joint production, as well as parameterize the degree of scope economies within a firm.[7]

The paper is structured as follows. Section 2 describes how we use output distance functions to parameterize a firm's technology. Section 3 derives expressions for plant-product markups using output distance functions, highlighting the key features of a production technology that needs to be known by a researcher to identify these markups. Section 4 presents our firm-level object and offers welfare based interpretations. Section 5 provides techniques for identifying the firm-level markup. Section 6 discusses our data and econometric implementation. Section 7 presents results. Section 8 concludes.

2 Production Technologies

We first introduce notation and what is known to the econometrician. Firms are indexed $i = 1, .., N$, products $j = 1, ..., J$, and inputs $s = 1, ..., S$. We assume that the econometrician observes what we call *standard production data* that includes a $1 \times J$ vector of output quantities $\mathbf{Y}_i \equiv (Y_i^1, Y_i^2, ...Y_i^J) \in \mathbb{R}^J_{\geq 0}$ and revenues $\mathbf{R}_i \equiv (R_i^1, R_i^2, ...R_i^J) \in \mathbb{R}^J_{\geq 0}$, as well as a $1 \times S$ vector of input quantities $\mathbf{X}_i \equiv (X_{1i}, X_{2i}, ...X_{Si}) \in \mathbb{R}^S_{\geq 0}$ and expenditures $\mathbf{E}_i \equiv (E_{1i}, E_{2i}, ...E_{Si}) \in \mathbb{R}^S_{\geq 0}$. Superscripts j refer to different outputs, and subscripts (s, i) index inputs s and firms i. We assume that standard production data only contain information on firm-level aggregate inputs and not how inputs are allocated across product lines j.

We define a firm's technology as its *production possibility set*, $\mathbb{P}_i$, where the firm is only

[6]This problem has a similar structure to the models of scale and scope considered in Ding (2022) and Argente et al. (2020); our key distinction is that we load spillovers across product lines through non-rival inputs that directly affect production, rather than knowledge or ideas which affect production through changes in productivity.

[7]See concurrent work by Khmelnitskaya et al. (2023) on identifying scope economies in the beer industry using the class of technologies proposed in this paper.

capable of producing with output and input vectors $(\mathbf{Y}_i, \mathbf{X}_i) \in \mathbb{P}_i$.[8] A firm's *production possibilities frontier* summarizes the *maximal* quantities of output that may be obtained for a given vector of inputs. As in Caves et al. (1982), we characterize a firm's production possibility frontier using an *output distance function* that maps all possible non-negative output and input vectors $(\mathbf{Y}_i, \mathbf{X}_i)$ to a non-negative scalar δ:

Definition 1 *A firm's* ***output distance function****,* $D_i(\mathbf{Y}_i, \mathbf{X}_i) : \mathbb{R}^{J+S}_{\geq 0, \neq \mathbf{0}} \to \mathbb{R}_+$*, solves:*

$$D_i(\mathbf{Y}_i, \mathbf{X}_i) \equiv \min_{\delta} \delta \quad s.t. \left(\frac{\mathbf{Y}_i}{\delta}, \mathbf{X}_i\right) \in \mathbb{P}_i. \tag{1}$$

The solution to this problem tells us by what (minimum) factor a firm must scale an output vector $\mathbf{Y}_i$ such that it can produce $\frac{\mathbf{Y}_i}{\delta}$ with $\mathbf{X}_i$. If $D_i(\mathbf{Y}_i, \mathbf{X}_i) > 1$, the firm cannot produce $\mathbf{Y}_i$ with $\mathbf{X}_i$; if $D_i(\mathbf{Y}_i, \mathbf{X}_i) < 1$, the firm can produce a vector of outputs which is strictly larger than $\mathbf{Y}_i$ with $\mathbf{X}_i$. This leads to the following definition:

Definition 2 *A firm's* ***production possibility frontier****,* $\mathbb{P}_i^F \subset \mathbb{P}_i$*, is the set of all* $(\mathbf{Y}_i, \mathbf{X}_i)$ *satisfying*

$$D_i(\mathbf{Y}_i, \mathbf{X}_i) = 1. \tag{2}$$

For the rest of the paper, we make the following assumption on the shape of the output distance function, $D_i(\mathbf{Y}_i, \mathbf{X}_i)$:

Assumption 1 $D_i(\mathbf{Y}_i, \mathbf{X}_i)$ *is continuous, twice-differentiable, and quasi-convex.*

We now define several properties of a firm's technology that are relevant for results that follow: *heterogeneity in marginal rates of transformation*, *joint production*, and *nesting single-product production*. We then use these properties to characterize whether there is joint or non-joint production based on the shape of the output distance function.

2.1 Heterogeneous Marginal Rates of Transformation

Large firms tend to produce multiple products with the firm's *core* product often defined as the product that is produced with the highest level of productivity or that which generates the highest revenue (Eckel and Neary 2010, Mayer et al. 2014, Orr 2022). It is reasonable to examine technologies that allow the core product to vary across firms. We consider broad classes of technologies —including joint production technologies where the notion of product-line specific productivity can be difficult to define. To this end, we capture the idea

[8] We assume $\mathbb{P}_i$ always includes $(\mathbf{0}, \mathbf{X}_i)$ so that a firm can always produce nothing, and that production possibility sets satisfy a no free lunch property, such that for $\mathbf{Y}_i > 0$, $(\mathbf{Y}_i, \mathbf{0}) \notin \mathbb{P}_i$.

of core products varying across firms by considering technologies where the *marginal rates of transformation* (MRT) for various products can vary across firms.

Marginal rates of transformation can be obtained from a firm's output distance function by totally differentiating a firm's output distance function at $D_i(\mathbf{Y}_i, \mathbf{X}_i) = 1$ for an arbitrarily small change in the output of good j and k, holding the output of all other goods fixed:

$$\frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial Y_i^j} dY_i^j + \frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial Y_i^k} dY_i^k = 0. \tag{3}$$

The marginal rate of transformation, which we represent as $\Delta_i^{kj}(\mathbf{Y}_i, \mathbf{X}_i)$, is defined as the marginal change in good k needed to increase good j by one unit:

$$\Delta_i^{kj}(\mathbf{Y}_i, \mathbf{X}_i) \equiv \frac{dY_i^k}{dY_i^j} = -\frac{\frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial Y_i^j}}{\frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial Y_i^k}}. \tag{4}$$

We define a model as having heterogeneous marginal rates of transformation — or *MRT heterogeneity* for short — if $\Delta_i^{kj}(\mathbf{Y}_i, \mathbf{X}_i)$ is indexed by i:

Definition 3 *A production model allows for* ***MRT heterogeneity****, if for any* $(\mathbf{Y}_i, \mathbf{X}_i)$, $\Delta_i^{kj}(\mathbf{Y}_i, \mathbf{X}_i)$ *can take on any value in* $\mathbb{R}_+$ *and can differ across* i.

The simplest version of a production model allowing for MRT heterogeneity is a firm-level Ricardian model where labor L_i is the only factor of production, there is no joint production, and there are constant returns to scale. In this case, the output distance function takes the form $D_i(\mathbf{Y}_i, \mathbf{X}_i) = \frac{\frac{Y_i^1}{A_i^1} + \frac{Y_i^2}{A_i^2}}{L_i}$, and $\Delta_i^{12} = -\frac{A_i^2}{A_i^1}$, where A_i^j are product-line Hicks specific TFP shifters. Here (A_i^1, A_i^2) acts a vector of unobserved heterogeneity at the firm-level.

2.2 Joint Production

Joint production can occur when there are public (i.e. shared) inputs that affect the output of all product lines simultaneously. Examples include shared managerial inputs, machines that produce multiple outputs at once, or production processes involving by-products, such as how the refining of sugar generates molasses.

Non-joint production is a setting where production takes place through a series of product-line specific production functions, $Y_i^j = F_i^j(\mathbf{X}_i^j)$, where $\mathbf{X}_i^j$ is a vector of input quantities allocated to product j, and $\sum_j \mathbf{X}_i^j = \mathbf{X}_i$. Since the inputs are allocated across product lines in a mutually exclusive and exhaustive manner in non-joint production settings, we can define the input share of input s into production line j as $S_{si}^j \equiv \frac{X_{si}^j}{\sum_k X_{si}^k}$, with $\mathbf{S}_i^j$ representing the $1 \times S$ vector of input shares S_{si}^j.

We define non-joint and joint production following Hall (1973) as:

Definition 4 *A technology is **non-joint** if there exists a set of functions $\{F_i^j(.)\}_j$ such that for all $(\mathbf{Y}_i, \mathbf{X}_i) \in \mathbb{P}_i^F$ there exist input shares $\{\mathbf{S}_i^j\}_j$ such that $Y_i^j = F_i^j(\mathbf{S}_i^j \circ \mathbf{X}_i)$ and $\sum_j \mathbf{S}_i^j = \mathbf{1}$. Otherwise the technology involves **joint production**.*

In general, whether a particular technology involves joint or non-joint production will depend on the shape of a firm's production possibility frontier (Samuelson 1966, Hall 1973). Section 2.4 formalizes this by characterizing how the shape of a firm's output distance function reveals whether production is joint or non-joint. We first describe the remaining features a firm's technology that we rely on for our results.

2.3 Nesting Single-Product Firms

We focus on production possibility frontiers that allow firms to choose to only produce a single product line j because we are interested in what features of a firm's technology can be estimated from single-product firms as in De Loecker et al. (2016). To start, we define the concept of nesting single-product production as follows:

Definition 5 *A technology **nests single-product production** if for any $\mathbf{X}_i$ and j, the set $\mathbb{P}_i^F$ contains $\left(\mathbf{Y}_i^j, \mathbf{X}_i\right)$ where $\mathbf{Y}_i^j$ is a $1 \times J$ vector of outputs where $Y_i^j \geq 0$ for j, and $Y_i^k = 0$ $\forall k \neq j$.*

This definition formalizes the idea that a multi-product firm using $\mathbf{X}_i$ can choose to produce only a single product. This requires that $D_i\left(\mathbf{Y}_i^j, \mathbf{X}_i\right)$ be well defined for all $\mathbf{Y}_i^j$. This rules out translog distance functions of the form considered in Caves et al. (1982).[9] A sufficient condition for a technology to nest single-product production is that $D_i\left(\mathbf{Y}_i^j, \mathbf{X}_i\right) = \frac{Y_i^j}{F_i^j(\mathbf{X}_i)}$. In this case, the production possibility frontier is characterized by a standard production function:

$$D_i\left(\mathbf{Y}_i^j, \mathbf{X}_i\right) = \frac{Y_i^j}{F_i^j\left(\mathbf{X}_i\right)} = 1 \implies Y_i^j = F_i^j\left(\mathbf{X}_i\right).$$

2.4 A Theorem on Joint vs Non-Joint Production Under Seperability

When specifying technologies using production possibility frontiers, it is common for researchers to rely on *separable* functional forms that possess well defined output and input aggregators (Mundlak 1963, Grieco and McDevitt 2017, Grieco et al. 2018, Maican and Orth

[9]Such functional forms imply that firms must produce positive quantities of every product, which rules out shutting down particular product lines.

2021, Dhyne et al. 2022). While separable transformation functions have useful properties for the purpose of estimation, Hall (1973) emphasizes that separable functional forms can easily imply joint production. In this subsection, we build on his result to show exactly when separability implies joint, or non-joint, production. For this purpose, we rely on the following definition of separability:

Definition 6 *An output distance function is **separable** if it can be written as:*

$$D_i\left(\mathbf{Y}_i, \mathbf{X}_i\right) = \frac{G_i(\mathbf{Y}_i)}{F_i(\mathbf{X}_i)}. \tag{5}$$

Theorem 1 (below) builds on Hall (1973) to establish a simple functional form test on a firm's output distance function to establish whether a firm's technology involves joint or non-joint production. For this to hold, we require two further restrictions on the space of feasible distance functions:

Assumption 2 $D_i\left(\mathbf{Y}_i, \mathbf{X}_i\right)$ *is separable, with* $F_i\left(\mathbf{X}_i\right)$ *continuous, differentiable, strictly increasing in all arguments, quasi-concave, and homogeneous of degree* $\phi_i > 0$.

Assumption 3 $D_i\left(\mathbf{Y}_i^j, \mathbf{X}_i\right) = \frac{Y_i^j}{A_i^j F_i(\mathbf{X}_i)}$ *for all* $\mathbf{Y}_i^j > 0$.

Assumption 2 imposes separability of the output distance function and establishes standard regularity conditions on $F_i\left(\mathbf{X}_i\right)$. The most important of these assumptions is that $F_i\left(\mathbf{X}_i\right)$ be homogeneous degree $\phi_i > 0$. Assumption 3 simply requires that firms can choose to produce a single product using a standard production function, while still respecting the separability assumption.[10] These two assumptions make characterizing non-joint production straightforward.

Theorem 1 *Suppose Assumptions 2 and 3 hold. Then a firm's technology is non-joint if and only if:*

$$D_i(\mathbf{Y}_i, \mathbf{X}_i) = \frac{\left(\sum_j \left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\phi_i}}\right)^{\phi_i}}{F_i(\mathbf{X}_i)}. \tag{6}$$

Proof. See Appendix A. □

An important implication of Theorem 1 is that if the firm's technology is separable and nests single product production, then the firm's technology involves joint production if the output aggregator, $G_i\left(\mathbf{Y}_i\right) \neq \left(\sum_j \left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\phi_i}}\right)^{\phi_i}$. As a result, assuming non-joint production implies specific restrictions on the shape of firm's output distance function.

[10]Notice that $D_i\left(\mathbf{Y}_i^j, \mathbf{X}_i\right) = \frac{Y_i^j}{A_i^j F_i^j(\mathbf{X}_i)}$ would violate separability, since the input aggregator would have different functional forms for $D_i\left(\mathbf{Y}_i^j, \mathbf{X}_i\right)$ and $D_i\left(\mathbf{Y}_i^k, \mathbf{X}_i\right)$, $k \neq j$, which would make $F_i()$ depend on $\mathbf{Y}_i$.

3 Firm-Product Markups

We now use our output distance function approach to derive what objects must be observed to identify firm-product markups. We also show how past approaches have dealt with the relevant identification issues. We do this to spell out the restrictions that have been made in the past, and to compare these restrictions to those made here.

The identification of markups using the production-side approach of De Loecker and Warzynski (2012) requires three key assumptions: i) cost-minimization, ii) the existence of at least one *static* input in production where the cost-minimizing firms are price takers, and iii) knowledge of some features of each firm's technology. We first present a more precise representation of these three assumptions, and show how they generate a mapping from standard production data $(\mathbf{Y}_i, \mathbf{R}_i, \mathbf{X}_i, \mathbf{E}_i)$ to unobserved markups, similar to Grieco et al. (2018) and Dhyne et al. (2022).

We partition the set of inputs into two subvectors, $\mathbf{X}_i = (\mathbf{M}_i, \mathbf{K}_i)$. $\mathbf{M}_i$ represents a vector of *static* inputs, and $\mathbf{K}_i$ denotes the vector of *dynamic* inputs. The set $\mathbb{M}$ of static inputs is the set of inputs purchased each period at exogenous prices W_{si}. The set $\mathbb{K}$ of dynamic inputs may be accumulated over time through a dynamic investment process. We assume that there is at least one static input used by all firms in a given industry (e.g. materials), and assume that each firm's conditional cost function $C_i(\mathbf{Y}_i, \mathbf{K}_i, \mathbf{W}_i)$ is given by the following minimization problem:[11]

$$C_i(\mathbf{Y}_i, \mathbf{K}_i, \mathbf{W}_i) = \min_{\mathbf{M}_i} \sum_{M_{si} \in \mathbb{M}} W_{si} M_{si} \quad \text{s.t. } D_i(\mathbf{Y}_i, \mathbf{X}_i) = 1. \tag{7}$$

The first-order condition for any static input $M_{si} \in \mathbb{M}$ is given by:

$$W_{si} + \lambda_i \frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial M_{si}} = 0 \tag{8}$$

where λ_i is the relevant Lagrangian multiplier. The Envelope Theorem implies that conditional marginal costs $MC_i^j \equiv \frac{\partial C_i(\mathbf{Y}_i, \mathbf{K}_i, \mathbf{W}_i)}{\partial Y_i^j}$ can be written as:

$$MC_i^j = \lambda_i \frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial Y_i^j}. \tag{9}$$

[11] Costs related to the accumulation of dynamic inputs will, in general, show up in a firm's profit function. Since the quantity of dynamic inputs are being conditioned on in the conditional cost minimization problem, the exact structure of these costs can be ignored when solving for the conditional cost function.

Combining equations (8) and (9) yields

$$\frac{1}{MC_i^j} = -\frac{\frac{\partial D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial M_{si}}}{\frac{\partial D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial Y_i^j}}\frac{1}{W_{si}}, \tag{10}$$

and the firm-product markup $\mu_i^j \equiv \frac{P_i^j}{MC_i^j}$ is obtained by multiplying equation (10) by P_i^j:

$$\mu_i^j = -\frac{\frac{\partial D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial M_{si}}}{\frac{\partial D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial Y_i^j}}\frac{P_i^j}{W_{si}} = -\frac{\frac{\partial \ln D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial \ln M_{si}}}{\frac{\partial \ln D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial \ln Y_i^j}}\frac{R_i^j}{E_{si}} \tag{11}$$

where the second equality uses the fact that $P_i^j Y_i^j = R_i^j$, and $X_{si}W_{si} = E_{si}$.

Equation (11) extends the De Loecker and Warzynski (2012) firm markup formula to the firm-product level for a (potentially) multi-product firm. If the firm produces a single product so that$\mathbf{Y}_i = \mathbf{Y}_i^j$, their formula holds as long as the relevant output distance function satisfies Assumption 2, in which case $\frac{\partial \ln D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial \ln Y_i^j} = 1$ and $\frac{\partial \ln D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial \ln M_{si}} = -\frac{\partial \ln F_i(\mathbf{X}_i)}{\partial \ln M_{si}} \equiv -\theta_{si}(\mathbf{X}_i)$, so $\mu_i^j = \theta_{si}(\mathbf{X}_i)\frac{R_i^j}{E_{si}}$. With a single-product firm, the researcher only needs to estimate a single elasticity, $\theta_{si}(\mathbf{X}_i)$. Equation (11) extends the De Loecker and Warzynski (2012) result by showing that product-level markups at multi-product firms can be recovered from standard production data if the researcher can estimate $\frac{\partial \ln D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial \ln M_{si}}$ and $\frac{\partial \ln D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial \ln Y_i^j}$. We now highlight identification challenges that a econometrician faces when tackling this problem.

3.1 Identification Challenges

We use equation (11) to illustrate what needs to be known to identify the ratio of markups across product lines μ_i^j/μ_i^k:

$$\frac{\mu_i^j}{\mu_i^k} = \frac{\frac{\partial \ln D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial \ln Y_i^k}}{\frac{\partial \ln D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial \ln Y_i^j}}\frac{R_i^j}{R_i^k} = \frac{\frac{\partial D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial Y_i^k}}{\frac{\partial D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial Y_i^j}}\frac{P_i^j}{P_i^k} = -\Delta_i^{kj}(\mathbf{Y}_i,\mathbf{X}_i)\frac{P_i^j}{P_i^k}. \tag{12}$$

The final equality in this expression tells us that it is crucial to identify the marginal rate of transformation across goods when recovering within-firm markup differences. In particular, conditional on output prices, markup ratios are completely determined by marginal rates of transformation. Unfortunately, equation (12) also implies that MRT heterogeneity can generate a serious underidentification problem because equation (12) only provides a single equation to identify two unknowns; relative markups, μ_i^j/μ_i^k, and the marginal rate of transformation between goods k and j. In short, relative output prices within a firm can be rationalized in two different ways: differences in market power, or differences in technologies.

To see this more clearly, consider the following parameterization of a firm's output distance function, which we derive in Appendix B based on a model of rival and non-rival intermediates as in Baumol et al. (1982):[12]

$$D_i\left(\mathbf{Y}_i, \mathbf{X}_i\right) = \frac{\left(\sum_j \left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\phi\beta}}\right)^{\phi\beta}}{F(\mathbf{X}_i)}, \tag{13}$$

where $F(\mathbf{X}_i)$ is an arbitrary homogenous of degree $\phi > 0$ function and $\beta \in (0, 1]$ is the share of rival intermediates in total costs such that $1 - \beta$ is the share of non-rival, or public, intermediates. Applying Theorem 1 to equation (13), the technology is non-joint if and only if $\beta = 1$; otherwise, the technology involves joint production. The A_i^j terms function as product-line specific Hicks-neutral productivity shifters. By including these terms, we allow for MRT heterogeneity since $\Delta_i^{kj}\left(\mathbf{Y}_i, \mathbf{X}_i\right) = -\left(\frac{Y_i^j}{Y_i^k}\right)^{\frac{1-\phi\beta}{\phi\beta}}\left(\frac{A_i^k}{A_i^j}\right)^{\frac{1}{\phi\beta}}$ depends on i, conditional on $\mathbf{Y}_i$ and $\mathbf{X}_i$.

Evaluating equation (12) using the technology in equation (13) delivers:

$$\frac{\mu_i^j}{\mu_i^k} = \left(\frac{Y_i^j}{Y_i^k}\right)^{\frac{1-\phi\beta}{\phi\beta}}\left(\frac{A_i^k}{A_i^j}\right)^{\frac{1}{\phi\beta}}\frac{P_i^j}{P_i^k}. \tag{14}$$

There are two (i, j, k) specific unknowns in this expression: $\frac{\mu_i^j}{\mu_i^k}$ and $\frac{A_i^k}{A_i^j}$. De Loecker et al. (2016) deal with this under-identification problem by assuming $A_i^j = A_i^k\ \forall j, k$, in which case we obtain $\frac{\mu_i^j}{\mu_i^k} = \left(\frac{Y_i^j}{Y_i^k}\right)^{\frac{1-\phi\beta}{\phi\beta}}\frac{P_i^j}{P_i^k}$. This allows the researcher to identify product-level markups because markup-ratios rationalize the observed prices and quantities. However, if $A_i^j \neq A_i^k$, this can potentially generate a mismeasurement problem.

Valmari (2016), Gong and Sickles (2021), Orr (2022) offer another path. They first model demand and price setting behavior and recover markups μ_i^j/μ_i^k. This then allows them to use equation (14) to recover A_i^k/A_i^j.[13] However, such an approach requires the researcher to take a stand on a particular demand system and pricing setting behavior, which is precisely what the production-side approach wishes to avoid.

Equation (14) also points to another identification problem: researchers must address whether there is joint, or non-joint production before they can identify product-level markups.

[12]Note that this representation of a firm's production possibility frontier is similar to that used in Eslava et al. (2023) and De Roux et al. (2021), with the important difference that ϕ, β, and A_i^j are purely technology, rather than preference, parameters.

[13]These papers only consider non-joint production in which case $\beta = 1$. ϕ can then be recovered by estimating a firm-product specific production function.

Even if $A_i^j = A_i \; \forall j$ such that $\frac{\mu_i^j}{\mu_i^k} = \left(\frac{Y_i^j}{Y_i^k}\right)^{\frac{1-\phi\beta}{\phi\beta}} \frac{P_i^j}{P_i^k}$, a researcher must still know the values of β and ϕ to identify product-level markup ratios. De Loecker et al. (2016) address this by assuming that production is non-joint ($\beta = 1$), in which case $\beta = 1$, and ϕ can identified by estimating returns to scale for single-product firms.[14]

The severity of the potential identification and mismeasurement problems associated with the problems and potential solutions above depends crucially on how the magnitude with which MRT heterogeneity and joint production manifest in the data. We explore this further in the next subsection, where we use simple Monte-Carlo simulations to ask how biased the estimated plant-product markups might be if the researcher's model ignores MRT heterogeneity or joint production when they are present.

3.2 Monte Carlo Evidence

We base our Monte Carlo evidence around the output distance function in equation (13), which allows us parameterize MRT heterogeneity with A_i^j/A_i^k differences, and the importance of joint production with β; the share of non-rival intermediates in production. We present two Monte Carlo exercises to highlight how erroneous assumptions of no MRT heterogeneity and non-joint production quantitatively affect estimated markups. In the first exercise, we switch off joint production but keep within-firm productivity dispersion in the true model. We compare markups from the true model to one derived under the incorrect assumption of there being no within-firm productivity dispersion. In the second exercise, we switch off within-firm productivity dispersion but keep joint production in the true model. We then compare the true markups to those derived under the incorrect assumption of production being non-joint. In both settings, we assume that there are demand shocks both within and across firms.[15] We now discuss these two exercises in detail.

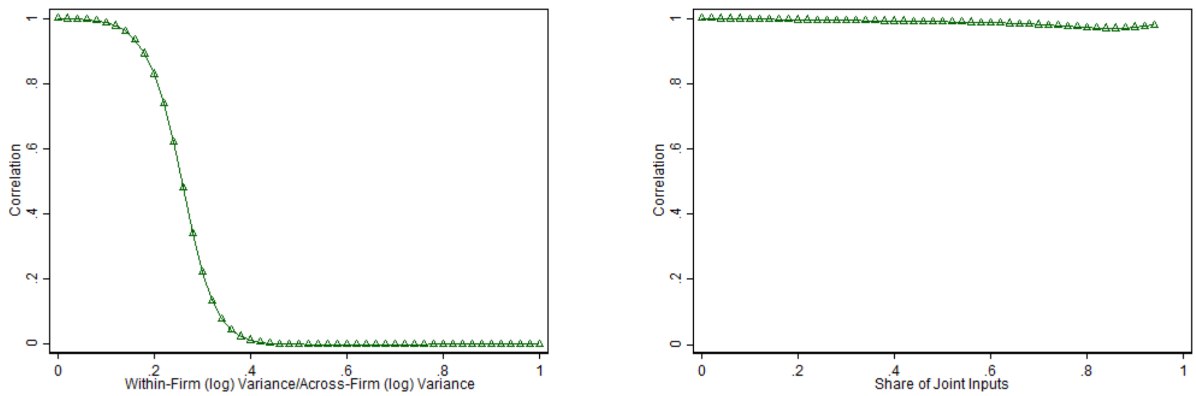
There are 1000 firms i, and 10 possible products j, a random subset of 5 of which each firm produces. Consumers spend a constant share of their income $1/J$ in each industry j. Labor is the only factor of production and is paid an exogenous wage w. We assume constant returns to scale in all industries for all firms such that $\phi = 1$, and $F(\mathbf{X}_i) = L_i$ where L_i is the total amount of labor used by the firm. This allows us to derive true markups using

[14]Recall that $F(\mathbf{X}_i)$ in equation (13) is an arbitrary homogenous of degree $\phi > 0$ function. In Appendix B, we show that this function also governs the product-line specific production technologies, and therefore can also be interpreted as the production function for single-product firms, with $\phi > 0$ governing overall returns to scale.

[15]We do this so that there is within-firm dispersion in output even when there is no within-firm productivity dispersion.

Figure 1: Monte Carlo

(A) Within-Firm TFP Heterogeneity (B) Joint Production



Notes: Each triangle in panel (A) presents the correlation of true markups from equation (16) and markups assuming no within-firm productivity heterogeneity from equation (17). The horizontal axis shows the ratio of within- to across-firm productivity heterogeneity for that correlation. There is no joint production in this panel. Each triangle is a separate correlation coefficient. Each triangle in panel (B) presents the correlation of true markups from equation (18) and markups assuming no within-firm productivity heterogeneity from equation (19). The horizontal axis shows the cost share of public inputs $1-\beta$. Each triangle is a separate correlation coefficient. There is no within-firm productivity heterogeneity in this panel.

equation (11) and the parameterization in equation (13):

$$\mu_i^j = \frac{\sum_k \left(\frac{Y_i^k}{A_i^k}\right)^{\frac{1}{\beta}}}{\left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\beta}}} \frac{R_i^j}{wL_i}. \tag{15}$$

Preferences of a representative consumer are given by $U = \prod_j \left(C^j\right)^{\alpha^j}$ where $C^j = \sum_{i \in \Omega_j} \left(c_i^j \nu_i^j\right)^{\frac{\sigma^j - 1}{\sigma^j}}$, ν_i^j is a demand shock, c_i^j is consumption of a variety of product j produced by firm i and Ω_j is the set of firms i that produce varieties of product j. We assume that the number of firms in each industry is sufficiently large that markups are constant within each industry and equal to the standard CES formula: $\mu_i^j = \frac{\sigma^j}{\sigma^j - 1}$.[16] Elasticities of substitution for the J industries σ^j run from 1.1 to 10.1 in even increments. Appendix C describes the setting in more detail.

For the first exercise, we impose no joint production ($\beta = 1$) in which case equation (15)

[16]It is easy to confirm that this will also equal the estimated markup using equation (15).

becomes

$$\mu_i^j = \frac{\sum_k \left(\frac{Y_i^k}{A_i^k}\right)}{\left(\frac{Y_i^j}{A_i^j}\right)} \frac{R_i^j}{wL_i}. \tag{16}$$

If one (falsely) assumes that there is no within-firm productivity heterogeneity ($A_i^j = A_i$), this expression becomes

$$\hat{\mu}_i^j = \frac{\sum_k Y_i^k}{Y_i^j} \frac{R_i^j}{wL_i}. \tag{17}$$

We now explore whether and how quickly markups derived from equations (16) and (17) diverge from one another as we allow for increasing amounts of within-firm productivity dispersion. If there is no within-firm productivity dispersion, and one correlates markups obtained from equations (16) and (17), the correlation is one because the assumption of $A_i^j = A_i$ is accurate. This is the first point on the far upper left of the panel (A) of Figure 1.

We then increase the within-firm variance of (log) productivity draws relative to the across-firm variance of (log) productivity draws as we move along the horizontal axis with the vertical axis reflecting the correlation between measures obtained between equations (16) and (17). When the variance of within-firm draws is small (10% of the across firm variance) relative to the across-firm variance, the assumption of no within-firm productivity heterogeneity is innocuous as reflected by a high correlation of 0.98. However, this changes quickly. When the ratio of the within- to across-firm variance is 0.3, the correlation falls to 0.22. We interpret this as evidence that markup estimation is sensitive to the precise degree of within-firm MRT heterogeneity.

In the second exercise (panel B), instead of steadily increasing the within-firm variance of productivity draws, we steadily increase the cost share of joint inputs ($1-\beta$) and compare true markups with markups obtained under the false assumption that production is non-joint. We assume there is no within-firm productivity dispersion in this second exercise. More formally, true markups in this case are given by

$$\mu_i^j = \frac{\sum_k \left(Y_i^k\right)^{\frac{1}{\beta}}}{\left(Y_i^j\right)^{\frac{1}{\beta}}} \frac{R_i^j}{wL_i}. \tag{18}$$

where $\beta \in (0,1]$. If one falsely assumes non-joint production, then one will believe that the following would accurately measure markups

$$\hat{\mu}_i^j = \frac{\sum_k Y_i^k}{Y_i^j} \frac{R_i^j}{wL_i}. \tag{19}$$

Analogous to panel (A) of Figure 1, panel (B) of Figure 1 presents the relevant correlations. When $1 - \beta = 0$, the share of joint inputs is zero, and there is no difference between restricted firm-product markups (from equation 19) and the true markups (from equation 18). Again, this is the dot at one in the upper left corner of panel (B). However, unlike with productivity dispersion, we find that the increasing presence of joint production ($\beta \downarrow$) does not materially affect one's ability to recover markups: the correlations remain well above 0.99 for cost shares of public inputs (1-β) up to 0.95. Appendix C shows plots of the variance of estimated markups instead of their correlation with true markups. As one might expect, the variance of estimates increases dramatically near the same parameter values when the correlation coefficients diminish. As a whole, these results suggest that within-firm productivity heterogeneity is more of the threat to identification of firm-product markups than is joint production.

4 Identification of Firm Markups

We now develop our *firm markup* and show that its identification does not require an econometrician to know whether or not there is joint production, nor the magnitude of MRT heterogeneity. Our object is a weighted average of firm-product markups and can be identified using only information on total firm revenues, expenditure on a static input, and a set of output distance function elasticities that can be estimated using techniques that have already been established in the literature subject to minor restrictions.

We obtain our object by multiplying the expression for a firm-product markup (equation 11) by $\frac{\frac{\partial \ln D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial \ln Y_i^j}}{\sum_{k\in\mathbb{Y}_i}\frac{\partial \ln D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial \ln Y_i^k}}$, and then summing over all $j \in \mathbb{Y}_i$, where $\mathbb{Y}_i$ is the set of all j such that $Y_i^j > 0$. This yields

$$\mu_i \equiv \sum_{j\in\mathbb{Y}_i} \rho_i^j \mu_i^j \qquad \text{where} \qquad \rho_i^j \equiv \frac{\frac{\partial \ln D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial \ln Y_i^j}}{\sum_{k\in\mathbb{Y}_i}\frac{\partial \ln D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial \ln Y_i^k}}. \tag{20}$$

Equation (20) shows that the firm markup defined here is simply a weighted average of firm-product level markups, μ_i^j, where the weights ρ_i^j sum to 1 by construction. These weights depend on the shape of a firm's output distance function, as well as equilibrium choices of product-level outputs and prices.

Slight manipulation of equation (20) delivers

$$\mu_i = -\frac{\frac{\partial \ln D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial \ln M_{is}}}{\sum_{k\in\mathbb{Y}_i}\frac{\partial \ln D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial \ln Y_i^k}}\frac{R_i}{E_{si}} \tag{21}$$

where $R_i \equiv \sum_j R_i^j$. This expression shows that an econometrician can identify μ_i using total firm revenue R_i, expenditure on a static input E_{si}, and output distance function elasticities $\frac{\partial \ln D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial \ln M_{is}}$ and $\sum_{k\in\mathbb{Y}_i}\frac{\partial \ln D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial \ln Y_i^k}$. Section 5 shows that these elasticities can be estimated using standard production data under minor restrictions. Importantly, estimation of these elasticities does not require a researcher rule out MRT heterogeneity or to take a stand on whether there is joint production.

Before we show how to estimate these elasticities, we first discuss interpretation of our firm-level markup. First, we provide an interpretation of the weights. Second, we show that the firm markup can be interpreted as the markup of a composite good produced by the firm in a world where consumers have general preferences over these firm-level composite outputs and the output distance function is separable. Third, we show that the firm markup is a meaningful object for a social planner who wants to eliminate across-firm misallocation but who takes within-firm misallocation as given.

4.1 Interpreting the Weights ρ_i^j

The weight ρ_i^j defined in equation (20) can be interpreted as the share of the percentage change in firm total cost when scaling up all products that can be attributed to scaling up product j. This makes it analogous to a cost share. To see this start by multiplying and dividing equation (21) by $\frac{\partial D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial M_{si}}\frac{1}{W_{si}}$:

$$\rho_i^j = \frac{\frac{\partial D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial Y_i^j}}{\frac{\partial D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial M_{si}}\frac{1}{W_{si}}}Y_i^j \Bigg/ \sum_{k\in\mathbb{Y}_i}\frac{\frac{\partial D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial Y_i^k}}{\frac{\partial D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial M_{si}}\frac{1}{W_{si}}}Y_i^k = \frac{MC_i^j Y_i^j}{\sum_{k\in\mathbb{Y}_i} MC_i^k Y_i^k} = \frac{\frac{\partial C_i(\mathbf{Y}_i,\mathbf{K}_i,\mathbf{W}_i)}{\partial Y_i^j}Y_i^j}{\sum_{k\in\mathbb{Y}_i}\frac{\partial C_i(\mathbf{Y}_i,\mathbf{K}_i,\mathbf{W}_i)}{\partial Y_i^k}Y_i^k}$$

where the second equality uses the definition of firm-product marginal cost $MC_i^j = \frac{\partial C_i(\mathbf{Y}_i,\mathbf{K}_i,\mathbf{W}_i)}{\partial Y_i^j}$ defined in equation (10) derived from the firm's cost minimization problem. This can be rewritten as:

$$\rho_i^j = \frac{\frac{\partial \ln C_i(\mathbf{Y}_i,\mathbf{K}_i,\mathbf{W}_i)}{\partial \ln Y_i^j}}{\sum_{k\in\mathbb{Y}_i}\frac{\partial \ln C_i(\mathbf{Y}_i,\mathbf{K}_i,\mathbf{W}_i)}{\partial \ln Y_i^k}}. \tag{22}$$

The numerator in equation (22) is the elasticity of the firm's cost function with respect to output j. While the denominator is simply the sum of the cost elasticities for all outputs within the firm, it is also equivalent to the elasticity of the cost function with respect to the scale of the firm. The denominator is the percentage change in the cost of the firm when it scales up all outputs by one percent.[17] The weight ρ_i^j therefore represents how costly it is to increase the output of product j within a multi-product firm.

The firm-level markup (equation 21) is therefore a weighted average of firm-product level markups where the weights correspond to the share of the product in total cost elasticity of the firm. This is analogous to the aggregate markup defined in Edmond et al. (2022), where single-product firm-level markups are weighted by the firm's cost share in the aggregate cost of production. Under their assumption of a common cost elasticity across firms, a firm's relative cost elasticity with respect to the cost elasticity of scaling up the entire economy is simply equal to the firm's cost share in total production. Our result generalizes the idea of a cost share to joint production settings where input shares are not well defined, due to the non-existence of product-line specific production technologies (i.e. joint production).

4.2 Single-Product Firm Equivalence Result

The firm markup acts as a firm-level wedge that scales down uniformly the production of all products within a firm. In order to develop this intuition, we examine a version of our model that is isomorphic to the single-product firm model of Edmond et al. (2022). They consider a world populated with single-product firms producing a product over which consumers have preferences. We first impose restrictions on consumer preferences and firms' technologies under which our multi-product firm model aggregates to their single-product firm model. We show that our aggregate firm-level markup is equivalent to their single-product firm markup under these restrictions.

Suppose that consumer preferences are given by $U(\{Y_i^j\}_{j,i}) = U(\{\mathcal{Y}_i\}_i)$ where the function $\mathcal{Y}_i = \mathcal{Y}_i(\{Y_i^j\}_j)$ is homogenous of degree 1 and $\mathcal{Y}_i(.)$ aggregates all products produced by a firm i. We refer to this as the firm's composite good. Further, under the assumption of separability, a firm's output distance function can be written as $D_i(\mathbf{Y}_i, \mathbf{X}_i) = G_i(\mathbf{Y}_i)/F_i(\mathbf{X}_i)$. If $G_i(.)$ is homogenous of degree one, we can collapse our multi-product firm to look like a

[17] To see this, start from the cost minimization problem defined in equation (7), and scale up the output vector by a factor λ. The cost elasticity with respect to the scale of the firm is given by the elasticity of the scaled cost function with respect to λ evaluated at $\lambda = 1$:

$$\left.\frac{\partial \ln C(\lambda \mathbf{Y}_i, \mathbf{K}_i, \mathbf{W}_i)}{\partial \lambda}\right|_{\lambda=1} = \sum_{k \in \mathbb{Y}_i} \frac{\partial \ln C(\mathbf{Y}_i, \mathbf{K}_i, \mathbf{W}_i)}{\partial Y_i^k} Y_i^k = \sum_{k \in \mathbb{Y}_i} \frac{\partial \ln C_i(\mathbf{Y}_i, \mathbf{K}_i, \mathbf{W}_i)}{\partial \ln Y_i^k}.$$

single-product firm that produces the composite good $\mathcal{Y}_i$. Assuming that firms operate on their production possibilities frontier, we can show the following:

$$\mathcal{Y}_i(\mathbf{Y}_i) = \frac{\mathcal{Y}_i(\mathbf{Y}_i)}{G_i(\mathbf{Y}_i)} F_i(\mathbf{X}_i) \qquad \Longrightarrow \tag{23a}$$

$$\mathcal{Y}_i(\mathbf{Y}_i) = \mathcal{A}_i(\boldsymbol{\lambda}_i) F_i(\mathbf{X}_i) \qquad \text{where} \qquad \mathcal{A}_i(\boldsymbol{\lambda}_i) \equiv \frac{\mathcal{Y}_i(\boldsymbol{\lambda}_i)}{G_i(\boldsymbol{\lambda}_i)} \tag{23b}$$

and $\boldsymbol{\lambda}_i = (\frac{Y_i^1}{Y_i^r}, \frac{Y_i^2}{Y_i^r}, ...)$ is the vector of firm outputs Y_i^j *relative* to some reference product within the firm $Y_i^r > 0$.[18]

Equations (23) relate any vector of firm outputs $\mathbf{Y}_i$ to a consumption aggregator $\mathcal{Y}_i(\mathbf{Y}_i)$. This relationship depends on two separate components of firm production: i) *within-firm output allocations*, which are captured by the relative output ratios $\boldsymbol{\lambda}_i$, and ii) *firm scale*, which is determined by its aggregate input use $\mathbf{X}_i$.

The production technology given in Edmond et al. (2022) is $Y_i = A_i F_i(\mathbf{X}_i)$ where A_i is a single-product-firm Hicks neutral productivity term that is independent of firm output. In our setting, firm productivity, $\mathcal{A}_i$ is determined by the within-firm output allocations, $\boldsymbol{\lambda}_i$. To link the two, we consider the case where within-firm output allocations are fixed such that relative output levels $\boldsymbol{\lambda}_i$ are taken at some initial reference level, $\boldsymbol{\lambda}_i^0$.[19] This makes $\mathcal{A}_i(\boldsymbol{\lambda}_i^0) = \mathcal{A}_i^0$ invariant to the scale of the firm. The firm production function can then be written as:

$$\mathcal{Y}_i = \mathcal{A}_i^0 F_i(\mathbf{X}_i). \tag{24}$$

In this environment, the De Loecker and Warzynski (2012) formula for the markup of the single-product firm defined in equation (24), μ_i^{SP}, with total revenue R_i and expenditure E_{si} on some static input X_{si} is then given by:

$$\mu_i^{SP} = \theta_{si} \frac{R_i}{E_{si}} = \frac{\partial \ln \left(\mathcal{A}_i^0 F_i(\mathbf{X}_i)\right)}{\partial \ln \left(X_{si}\right)} \frac{R_i}{E_{si}} = - \frac{\frac{\partial \ln(D_i(\mathbf{Y}_i, \mathbf{X}_i))}{\partial \ln(X_{is})}}{\sum_{k \in \mathbb{Y}_i} \frac{\partial \ln(D_i(\mathbf{Y}_i, \mathbf{X}_i))}{\partial \ln\left(Y_i^k\right)}} \frac{R_i}{E_{si}} = \mu_i$$

where the final equality follows from applying equation (21) to a separable output distance function. This implies that–in this setting–the markup of a single-product firm producing $\mathcal{Y}_i$ is equivalent to the weighted average of firm-product markups defined in equation (21). In this sense, our firm markup, μ_i, can be interpreted as the markup on a composite good

[18] We make use of the homogeneity of functions $\mathcal{Y}_i(\cdot)$ and $G_i(\cdot)$ in going from the first to the second line of equations (23).

[19] More formally, we assume that firms only produce output vectors such that $\{\mathbf{Y}_i : \boldsymbol{\lambda}_i = x \times \boldsymbol{\lambda}_i^0, x \in \mathbb{R}^+\}$

produced by its single-product counterpart.

4.3 Firm Markups and Across-Firm Misallocation

We now ask what the social planner's allocation looks like in this economy by multi-product firms under the conditions set out in section 4.2. From Edmond et al. (2015, 2022), we know that the social planner would allocate inputs across all of the single-product producers so as to equalize markups, i.e. $\mu_i = \mu \; \forall \; i$, so as to eliminate misallocation of resources across firms. This leads to the well-known result that variation in markups across firms is a symptom of misallocation.[20]

This result holds only if $\mathcal{A}_i$ is invariant to firm scale; otherwise, our model is not isomorphic to the single-product firm setting of Edmond et al. (2015, 2022). This requires that we hold within-firm output allocations fixed, i.e. $\boldsymbol{\lambda}_i$ remains unchanged when we consider counterfactual changes to firm scale through $\mathbf{X}_i$. We interpret this setting as one in which a social planner targets *across-firm* misallocation, taking *within-firm* output allocations as given. In this setting, we can then interpret our firm markup as equivalent to that in Edmond et al. (2015, 2022) and, therefore, as a summary statistic for the extent of across firm misallocation, even though our generalization allowed for multi-product firms with heterogeneous MRT and joint production.

We now extend the same result after relaxing the restrictions placed on consumer preferences and firm technology in the previous subsection. Suppose instead that the representative consumer has preferences over all products produced in the economy $U(\{Y_i^j\})$, and firm's technology is characterized by the output distance function $D_i(\mathbf{Y}_i, \mathbf{X}_i)$. A social planner who wants to eliminate across-firm misallocation taking within-firm output allocations as constant solves the following problem:

$$\max_{\{Y_i\}_i} U\left(\{Y_i \boldsymbol{\lambda}_i\}\right) \qquad \text{s.t.} \qquad D_i(Y_i, \boldsymbol{\lambda}_i, \mathbf{X}_i) \leq 1 \qquad \text{and} \qquad \sum_i X_{si} \leq X_s \qquad \forall \, i, s \quad (25)$$

where $\boldsymbol{\lambda}_i$ is the vector of relative output levels within the firm, and Y_i is the *scale* of the firm. The problem outlined in equation (25) states that the social planner chooses the scale of the firm, Y_i, by allocating resources across firms while keeping the ratio of outputs within each firm unchanged. The first order conditions from this problem gives the following condition

[20] If one extends this setting to have an endogenous supply of resources in the economy (e.g. by assuming labour is costly for a representative consumer as in Edmond et al. 2022), then you can further show that markups should all be equalized to 1.

that has to be met in order to eliminate across-firm misallocation:[21]

$$-\frac{\frac{\partial \ln D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial \ln X_{si}}}{\sum_{k\in\mathbb{Y}_i}\frac{\partial \ln D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial \ln Y_i^k}}\frac{\sum_{k\in\mathbb{Y}_i}\frac{\partial U(\{\mathbf{Y}_i\})}{\partial Y_i^k}Y_i^k}{X_{si}}=\gamma_s \qquad \forall\ i$$

where γ_s is the shadow price of the input X_{si}. The left-hand-side of the above equation in the decentralized equilibrium is proportional to the product of the price of input X_{si} that the firm faces and the firm markup, $w_{si}\mu_i$ (see Appendix D). Under the assumption that all firms face the same input prices, firms markups must be equalized, i.e. $\mu_i = \mu \ \forall\ i$ for an efficient allocation of resources to hold. Therefore, the constrained social planner would like to allocate inputs across firms such that firm markups are equalized. The firm markup can then be interpreted as a output wedge that decreases firm scale so that variation in the firm markup provides evidence of across firm misallocation, even under more general utility and production structures.

5 Estimating Firm Markups

Following De Loecker and Syverson (2021), we now discuss how two econometric techniques standard in the literature can be used to estimate the relevant output distance function elasticities needed to identify the firm markup given in equation (21). First, we can follow De Loecker et al. (2016) and use a sample of single-product firms to estimate an output elasticity for a multi-product firm that counterfactually chooses to be single-product. As long as the output distance function is separable, this single elasticity is sufficient to recover the firm markup. Second, we can use a cost-share estimator applied to the full sample of multi-product firms following Syverson (2004), Foster et al. (2008), Backus (2020), and Raval (2022).[22] We refer to these as *techniques 1* and *2*, respectively.

5.1 Technique 1: Relying on Single-Product Firms for Estimation

De Loecker et al. (2016) use single-product firms to estimate product-level output elasticities with non-joint production. While identification of firm-product markups in their context depends critically on MRT heterogeneity (see figure 1), an almost identical estimation approach can be used to identify the firm markup while not making any assumptions

[21] See derivation in Appendix D.

[22] See the excellent discussion in De Loecker and Syverson (2021) for more details of the merits of these two approaches.

about MRT heterogeneity even if there is joint production.[23]

Our extension requires two key restrictions. First, as previously discussed, each firm's technology must nest the possibility of a well-defined single-product production function (Assumption 3). This means that a firm uses the single-product production function whenever it chooses to produce a single product. This allows us to tie output elasticities of single-product firms to those of multi-product firms and does not preclude the possibility of joint production for multi-product firms whenever $\mathbf{Y}_i \neq \mathbf{Y}_i^j$ (Theorem 1). The second restriction is a set of restrictions on the firm's output distance function:

Assumption 4 *$D_i(\mathbf{Y}_i, \mathbf{X}_i)$ is separable, $G_i(\mathbf{Y}_i)$ is homogeneous of degree 1, and $F_i(\mathbf{X}_i) = F_{g(i)}(\mathbf{X}_i)$ for some mapping for firms i to industries, g.*

Combining Assumptions 3 and 4, the shape of the production function only differs at the *industry* level for single product firms.[24] This is a standard restriction imposed by nearly the entirety of the literature on production function estimation because estimation of production function parameters requires some sort of averaging across a set of firms operating in the same industry g.[25]

Assumption 4 does *not* restrict firms from having varying degrees of returns to scale, and does not require that single-product production be homogeneous of degree $\phi > 0$. The output aggregator $G_i(\mathbf{Y}_i)$ for for multi-product firms must be homogeneous of degree 1 but can otherwise vary flexibly across firms. This flexibility is key to allowing for MRT heterogeneity as well as joint/non-joint production.[26]

Assumptions 3 and 4 combined also imply that the firm markup can be recovered using an estimate of a single-product firm's output elasticity for a static input M_s. This is a key result in this paper which we now state and prove as a proposition:

Proposition 1 *Suppose Assumptions 3 and 4 hold. Then,*

$$\mu_i = \theta_{sg(i)}(\mathbf{X}_i) \frac{\sum_j R_i^j}{E_{si}} \tag{26}$$

where $\theta_{sg(i)}(\mathbf{X}_i) \equiv \frac{\partial \ln F_{g(i)}(\mathbf{X}_i)}{\partial \ln M_{si}}$ is the output elasticity for static input s in firms that choose

[23] This is because our approach sidesteps the need to solve an input allocation problem as in De Loecker et al. (2016).

[24] By "shape" we exclude TFP heterogeneity which we naturally allow to vary across firms and products.

[25] See for example Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg et al. (2015), Gandhi et al. (2020).

[26] For example, equation (13) provides an example of an output distance function consistent with Assumption 4. As we previously demonstrated, this class of technology allows for MRT heterogeneity, as well as joint or non-joint production. Since we are allowed to index the output aggregator, G_i, with i, we can allow for even greater flexibility than the output distance function in equation (13).

to produce a single product, i.e. if $\mathbf{Y}_i = \mathbf{Y}_i^j$*, then all cost minimizing firms operate using the single-product production technology* $Y_i^j = A_i^j F_{g(i)}(\mathbf{X}_i)$*.*

Proof. See Appendix E. □

The right-hand side of equation (26) is identical to the standard De Loecker and Warzynski (2012) markup formula, with the caveat that $\theta_{sg(i)}(\mathbf{X}_i)$ for multi-product firms is understood as the output elasticity for the firm if it counterfactually chose to be a single product producer. This shows that the De Loecker et al. (2016) estimation approach–using single product firms to estimate this output elasticity–can in principal be used to identify the firm markup.

An important concern is that firms may self-select into whether they are single product or multi-product producers depending on their productivity differences, potentially leading to selection bias. De Loecker et al. (2016) show that it is possible to correct for this bias in their setting. In Appendix Q, we show that their approach can be extended to environments that involve joint production and MRT heterogeneity based on a core competence model as in Eckel and Neary (2010), Mayer et al. (2014) and Arkolakis et al. (2021). However, one should not disregard that a key disadvantage to relying on single-product firms for estimation of the relevant elasticities in (21) is that there may be biases introduced by modelling the selection into single-product status incorrectly. This partially motivates Technique 2.

5.2 Technique 2: Cost Shares

This approach does not rely on using single-product firms as a counterfactual for multi-product firms and therefore is not subject to concerns about selection bias in estimation. This technique develops a cost-share approach to identify the unknown elasticities in equation (21) following Syverson (2004), Foster et al. (2008), Backus (2020), and Raval (2022). However, the validity of this approach requires restrictions not needed for Technique 1: *i*) static first-order conditions must hold for all inputs on average, and *ii*) each firm's (separable) input aggregator must be constant returns to scale Cobb-Douglas varying by industry. We formalize these as Assumptions 5 and 6 below:

Assumption 5 *For any input s, each firm's static cost minimizing first-order condition, expressed in terms of total expenditure on each input,*[27] *holds on average:*

$$\mathbb{E}\left(W_{si}X_{si}\right) = -\mathbb{E}\left(\lambda_i \frac{\partial D_i\left(\mathbf{Y}_i, \mathbf{X}_i\right)}{\partial X_{si}} X_{si}\right). \tag{27}$$

[27] i.e, each input's first order condition is multiplied by X_{si}, so the right hand side of equation (8) is expressed in terms of expenditures, rather than quantities.

Assumption 5 requires that firms adjust their inputs consistent with cost minimization *on average.* This may accommodate adjustment costs in capital (for example), which may prevent equation (8) from holding exactly each period or for each firm, even though a firm facing no adjustment costs will make equation (8) hold since current period costs would be lower for the firm if they set total capital costs to $\lambda_i \frac{\partial D_i(\mathbf{Y}_i,\mathbf{X}_i)}{\partial X_{si}} X_{si}$.

Assumption 6 *$D_i(\mathbf{Y}_i, \mathbf{X}_i)$ is separable, with $F_i(\mathbf{X}_i) = \Pi_s (X_{si})^{\beta_{s,g(i)}}$, where $\sum_s \beta_{s,g(i)} = 1$, for some mapping of firms to industries $g(i)$.*

Assumption 6 requires the output aggreagtor of the distance function to simplify to a standard constant returns to scale Cobb-Douglas production function whenever firms only produce a single product j. Since single-product production functions are industry specific, industry is a property of a firm, rather than a firm-product pair.[28] Together, Assumptions 5 and 6 lead to the following result:

Proposition 2 *Suppose Assumptions 5 and 6 hold. Then as long as at least one purely static input, X_{si}, exists and firms minimize their conditional costs as in equation (7), then:*

$$\mu_i = \frac{\mathbb{E}\left(W_{si}M_{si}|g = g(i)\right)}{\mathbb{E}\left(\sum_{s'} W_{s'i}X_{s'i}|g = g(i)\right)}\frac{R_i}{E_{si}}. \tag{28}$$

Proof. See Appendix G. □

A consistent estimate of $\frac{\mathbb{E}(W_{si}M_{si}|g=g(i))}{\mathbb{E}\left(\sum_{s'} W_{s'i}X_{s'i}|g=g(i)\right)}$ is then easily constructed using industry by industry aggregate cost shares, as in Syverson (2004), Foster et al. (2008), Backus (2020) and Raval (2022).

5.3 Comparing Techniques 1 and 2

While Technique 1 relies on single-product firms to estimate the relevant elasticities for identifying markups, Technique 2 uses observable cost shares and does *not* require that we condition on single-product firms. Therefore Proposition 2 from Technique 2 provides a useful alternative to Proposition 1 in Technique 1 for obtaining firm markups. However, Technique 2 relies on functional form restrictions not needed in Technique 1, most notably the requirement that output distance function be separable with a constant returns to scale Cobb-Douglas input aggregator.

[28] This restriction can be relaxed is one is comfortable assuming that *all* inputs are fully static, so that equation (8) holds firm-by-firm for all inputs. In Appendix F we show that if: i) the output distance function satisfies constant returns to scale, and ii) all inputs are fully static; then we can identify the firm markup using the ratio of revenues to costs. Importantly, this result allows for *non-separable* output distance functions, which can capture situations where firms operate non-joint production technologies that have different factor intensities (Hall 1973).

Because of this, Technique 1 and 2 require non-overlapping restrictions on a firm's technology and economic environment. While Technique 1 allows for rich models of dynamic input use and varying returns to scale, Technique 2 requires constant returns to scale Cobb-Douglas technologies, and that static first-order conditions hold across firms on average. This latter requirement limits dynamics. Technique 2 directly uses multi-product firms in its estimation, and does not require disciplining how firms choose their output sets as in Technique 1. Therefore determining which one is proper is difficult and case-specific. We therefore pursue both strategies in our estimation.

6 Data and Implementation

This section first briefly describes our data. We then describe how we operationalize the estimation of our output elasticities following the two techniques described above. Finally, we discuss how to estimate plant-product markups for comparison to our plant-level objects.

6.1 Data

We implement our methodology using the Indian Annual Survey of Industries (ASI) data set for the years 2001-2008. The ASI is a panel data set of manufacturing establishments with 10 or more permanent employees that use electricity, and of establishments 20 or more employees that do not use electricity. Large plants are always included and therefore ASI is a census of these firms. In addition, ASI contains a 5% random sample of small plants.[29] An observation is a plant-product-year triad. Each plant i reports a single 2-digit NIC industry $g(i)$ in which it operates. We take $g(i)$ to correspond to a particular production technology as in De Loecker et al. (2016).[30] Each plant reports up to ten products that it produces, with each entry corresponding to a single j in our framework.[31]

The primary appeal of the ASI is its inclusion of detailed product-line data, including revenues (measured in rupees) and physical outputs. Product-level output is the physical quantity of the good manufactured by the plant in a given year.[32] We observe capital,

[29] The cutoff for a plant to be considered "large" is based on the number of permanent employees, with the exact threshold varying from state to state.

[30] Besides major revisions in 1999 and 2009, the NIC system underwent a minor revision in 2005 from ISIC-3 to ISIC-3.1. While identical at the 2-digit level, ISIC-3.1 revised some 4-digit industry codes. In order to have a consistent set of industry codes, we drop plant-years from our pre-2005 data with a 4-digit industry code that does not have an exact match in the revised NIC. This results in the loss of about 11,000 plant-product-years.

[31] Associated with each product entry is a 5-digit ASICC code. Note that some plants report more than one entry for a given 5-digit ASICC code, in which case we treat each entry as a separate product.

[32] Each product level quantity is associated with a 5-digit product code as well as a particular unit of

Table 1: Summary statistics

Plant-product outcome	Mean	Std. dev.	Min.	Max.
Log Revenue: $\ln(R_{it}^j)$	15.926	2.681	1.386	27.557
Log Quantity Produced: $\ln(Q_{it}^j)$	0	2.482	-14.28	15.631
Plant-product-years:				355,624

Plant outcome	Mean	Std. dev.	Min.	Max.
Log Plant Revenue: $\ln(R_{it})$	17.051	2.235	6.774	27.603
Log Plant Capital: $\ln(K_{it})$	9.809	2.257	-6.697	19.813
Log Plant Labour: $\ln(L_{it})$	9.455	1.566	1.792	16.512
Log Plant Materials: $\ln(M_{it})$	11.328	1.959	2.313	20.816
Log Plant Wage: $\ln(W_{it})$	4.912	.622	0	11.832
Average products per plant-year:				1.77
Plant-years:				200,886
Distinct plants:				96,047

Notes: See main text and Appendix H for more detail on variable definitions. Log quantities demeaned within 5-digit product code.

labor, and intermediates input expenditures at the plant level, not the plant-product level. The plant capital stock is constructed using the perpetual inventory method as described in Appendix H. Labor input is in person-days. Person-days and the wage bill are plant totals for all employees including supervisors and managers. Intermediate inputs are the sum of expenditure on domestic and foreign goods. We deflate capital and intermediates using the Indian Wholesale Price Index (WPI) for all manufacturing industries although we offer a further input price correction described in section 6.2.[33] See Appendix H for more details.

Table 1 presents summary statistics. Our data comprise 355,624 plant-product-years, and 200,866 plant-years. The average number of products per plant-year is less than 2 suggesting the prominence of single-product plants. We now discuss how we use the plant data to estimate physical output elasticities and firm-level markups. We perform estimation using plants, not firms.[34]

measure (e.g. kilograms, number of, etc...). We demean these quantities within product code when presenting summary statistics for quantities.

[33]This follows De Loecker et al. (2016). For the years of our dataset, the WPI records price indices for some 2-digit NIC industries but not others. We opt to use the WPI for all manufacturing industries to deflate expenditures for all plants in our data.

[34]Since our data is plant (not firm) level, from now on we primarily let i index plants, rather than firms. We will compare our plant-product markups to those estimated in De Loecker et al. (2016). The latter are *firm*-product markups, and are estimated using the Prowess data set covering the years 1989-2003. See De Loecker et al. (2016) for more details on the Prowess sample frame.

6.2 Plant Markups

As discussed in Section 5, we consider two techniques to recover plant markups: the estimation of the production function for single product plants using a control function approach, as well as a cost-share estimator.

Estimation of the production function for single product plants requires specifying a functional form. We rely on a translog specification of the production function which imposes the constraint that the production function is globally homogenous of degree $\phi > 0$:[35]

$$y_{it}^j = \beta_L l_{it} + \beta_K k_{it} + \beta_M m_{it} + \beta_{LK}\left(l_{it}k_{it} - 0.5\left(l_{it}^2 + k_{it}^2\right)\right) + \beta_{LM}\left(l_{it}m_{it} - 0.5\left(l_{it}^2 + m_{it}^2\right)\right) + \beta_{KM}\left(k_{it}m_{it} - 0.5\left(k_{it}^2 + m_{it}^2\right)\right) + a_{it}^j \quad (29)$$

where lower case variables denote natural logs, and global returns to scale are given by $\phi = \beta_L + \beta_K + \beta_M$. In our baseline specification based on Technique 1, we focus on results where a single translog is estimated for the entire Indian economy, although we also report results where we allow production function parameters to vary by two-digit industry in Appendixes N and O.

We follow De Loecker et al. (2016) closely by implementing both a selection correction for single-product firms, and a control function approach to deal with the endogeneity of inputs.[36] To deal with measurement error in capital, our baseline results identify the coefficient on capital using lagged investment, as recommended by Collard-Wexler and De Loecker (2021). One minor point of departure is how we deal with input price bias. Since we observe plant-specific wages, we use this information directly to correct for input price bias in a way that does not require that the input price control function be estimated simultaneously with production function parameters- see Appendix I for more details. [37]

After having estimated each firm's (potentially counterfactual) single-product production function, we then construct plant markups. Under the assumptions underlying Propositions 1 and 2, we assume that materials is a static input ($s = M$), and write:

[35]See De Ridder et al. (2022) and Orr and Tabari (2022) for similar specifications of the translog with the added restriction that $\phi = 1$.

[36]We also consider dynamic panel based approaches to estimation in Appendixes J, N, and O to verify that our results are not driven by some of the concerns articulated by Bond et al. (2021) concerning proxy-variable approaches.

[37]To recover input allocations for multi-product firms, we will need the input price control function to recover output-specific input prices. However, since we observe wages at the plant level, we can obtain more precise estimates of the relationship between input prices and quality by directly regressing observed wages for single-product producers onto prices and market shares to obtain the relevant input price function. As a result, we can obtain the input price control function *outside* of the production function estimation algorithm. See Appendix I for more details.

$$\mu_{it} = \theta_{Mg(i)} \left(\mathbf{X}_{it}\right) \frac{\sum_j R_{it}^j}{E_{Mit}} \tag{30}$$

where $\theta_{Mg(i)}$ is either the output elasticity for single product firms in industry g (as in Proposition 1) or the industry cost share for materials (as in Proposition 2).[38]

The key decision that needs to be made to implement the cost-share estimator (Technique 2) is how to allocate plants to industries, where the input aggregator $F_{g(i)}(.)$ is assumed to have the same Cobb-Douglas form for all plants i belonging to industry g.[39] We consider two mappings of plants to industries $g(i)$. First, we simply take a plant's stated two-digit NIC code as their industry. Second, we follow Raval (2023) and for each two-digit NIC code, we split plants into five different quintiles, based on the ratio of labour to material expenditures, and treat each two-digit code × quintile as a separate "industry" g. Raval (2023) proposes this approach to account for unobserved factor-specific productivity differences across firms, and documents that this method alleviates some of the issues with the production-side approach to markup estimation described in Raval (2022). For each mapping of plants to industries, we then construct output elasticities using the industry-specific cost shares.

6.3 Plant-Product Markups

For comparison to our plant markups, we also estimate plant-product markups adhering as closely to De Loecker et al. (2016) which relies on a system of equations to solve for the unobserved input allocations.[40] The most basic variant of this approach imposes non-joint production and that $A_i^j = A_i \; \forall j$. In this case unobserved input allocations ρ_{it}^j can be obtained by solving the following system of equations for each plant-year (i, t):

$$\sum_{j \in \mathbb{Y}_i} \rho_{it}^j = 1, \tag{31}$$

and

$$Y_{it}^j = A_{it} F\left(\rho_{it}^j \frac{\mathbf{E}_{it}}{\widehat{W}_{it}^j}\right) \quad \forall j \tag{32}$$

[38] We always restrict our sample to plant-years where the estimated $\theta_{Mg(i)} \left(\mathbf{X}_{it}\right)$ is strictly positive. While this is always true for our baseline sample, we occasionally obtain $\theta_{Mg(i)} \left(\mathbf{X}_{it}\right) < 0$ for some industries and plants when we consider alternative production function estimators in the Appendices. We drop those plants whenever this happens.

[39] Although another important decision is how to obtain the rental rate of capital for each plant. For this purpose, we largely follow Raval (2022)— see Appendix H.

[40] We calculate plant-product markups instead of simply using the measures from the replication files to De Loecker et al. (2016) (which uses the Prowess data set) so that we can compare ASI-based plant markups to ASI-based plant-product markups.

where $\widehat{W}_{it}^{j}$ denotes the product line specific input-price control function.[41] The system of equations defined by equations (31) and (32) provide $J+1$ equations to obtain down $J+1$ unknowns; the J expenditure shares ρ_{it}^{j} and A_{it}. De Loecker et al. (2016) solve this system of equations numerically.[42]

Since the translog specification of the production function is homogeneous of degree ϕ, (31) and (32) provide a simple closed form expression for the input shares:[43]

$$\rho_{it}^{j} = \frac{\left(Y_{it}^{j}\right)^{\frac{1}{\phi}} \widehat{W}_{it}^{j}}{\sum_{k} \left(Y_{it}^{k}\right)^{\frac{1}{\phi}} \widehat{W}_{it}^{k}}. \tag{33}$$

Since production is non-joint in De Loecker et al. (2016), our estimates of ρ_{it}^{j} are sufficient to recover estimates of product-level markups, which using the standard De Loecker and Warzynski (2012) formula adapted to multi-product firms as in De Loecker et al. (2016):

$$\mu_{it}^{j} = \theta_{Mg(i)} \left(\rho_{it}^{j} \frac{\mathbf{E}_{it}}{\widehat{W}_{it}^{j}}\right) \frac{R_{i}^{j}}{\rho_{it}^{j} E_{Mi}} = \theta_{Mg(i)} \left(\mathbf{E}_{it}\right) \frac{R_{i}^{j}}{\rho_{it}^{j} E_{Mi}} \tag{34}$$

where the second equality uses the fact that $\theta_s = \frac{\partial F(\mathbf{X})}{\partial X_s} \frac{X_s}{F(\mathbf{X})}$ is homogeneous of degree 0 in $\mathbf{X}$ if the production function is homogeneous of degree $\phi > 0$.

We also consider specifications where ρ_{it}^{j} is given by the within-plant *revenue share* of product j, $R_{i}^{j} / \sum_{j} R_{i}^{j}$. This specification is of interest for two reasons; first, allocating inputs based on on revenue shares is common in the literature (Foster et al. 2008, Atalay 2014, Collard-Wexler and De Loecker 2015, Blum et al. 2018). Second, this input allocation rule rationalizes within-plant heterogeneity entirely through heterogeneity in marginal rates of transformation, rather than markup differences. Orr (2022) shows that revenue shares reflect within-plant productivity differences as long as production functions are identical and homogeneous of degree $\phi > 0$, and there is no within-firm markup dispersion. As a result, this allocation rule effectively imposes a constant within-plant markup, $\mu_{it}^{j} = \mu_{it}$, to allow for flexible A_{it}^{j} differences, rather than imposing constant within firm productivity,

[41] We obtain the input price control function outside of the production function estimation routine by regressing plant wages on various observables, such as prices and market shares, for single-product firms. See Appendix I for more details. If the input-price control function is correctly specified and estimated, $\mathbf{X}_{it}^{j} = \frac{\mathbf{E}_{it}}{\widehat{W}_{it}^{j}}$.

[42] De Loecker et al. (2016) rely on predicted output, $\widehat{\Phi}_{it}^{j}$, from a variant of their first stage estimating equation to generate (32), rather than output directly. We rely on output in our baseline specification since it is more tightly linked to our theory, but also consider whether using $\widehat{\Phi}_{it}^{j}$ or Y_{it}^{j} makes any difference for the recovered product-level markups as a robustness check.

[43] See Appendix K for derivation.

$A_{it}^j = A_{it}$ to identify flexible product-level markup differences μ_{it}^j.[44]

7 Results

We now examine plant-product markups using methods popular in the literature, and then compare them to the plant markups that are the focus of this paper.

Section 7.1 starts by showing that common estimators of plant-product markups deliver a large mass of markups close to zero. This implies nearly infinite marginal costs or zero prices. We show that such markups are nearly always found at multi-product firms suggesting potential misspecification of within-plant across-product input allocation rules. We then present data-driven exercises suggesting that omission of MRT heterogeneity may be behind this issue consistent with the Monte Carlo results of 3.2.[45]

Section 7.2 displays our plant-level markups where output elasticities are derived using Techniques 1 and 2 as discussed in section 5. Using both techniques, our plant-level markups display much tighter dispersion than those presented in section 7.1 even without trimming outliers. Further, we do not see a mass of markups at zero.

7.1 Product-level Markups and a Zeros Puzzle

7.1.1 Documenting the Problem of Outlier Markups

Figure 2 presents histograms of markups at the firm/plant-product level under the assumption that production is non-joint with no MRT heterogeneity (equations 33 and 34). Panel (A) uses the control function estimation approach as described in section 6.3 applied to our plant-level ASI data. Panel (B) plots firm-product markups obtained directly from De Loecker et al. (2016) replication files.[46]

Both panels display a large mass of markups near zero, another above 1, and a long right tail. While the second peak is expected, the mass of estimates near zero is striking, and can only be rationalized by very high marginal costs, or prices of almost zero. Since this is the *multiplicative* markup, $\mu \equiv \frac{P}{MC}$, perfect competition corresponds to $\mu = 1$. Therefore, $\mu = 0$

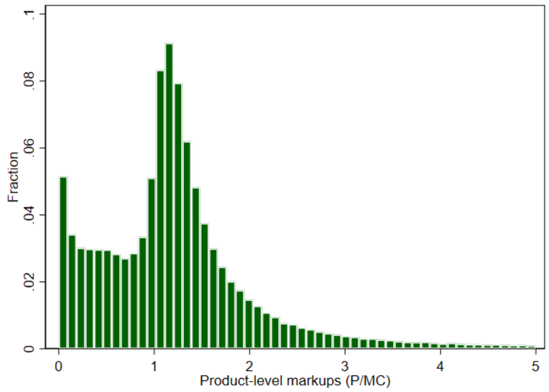
[44] To see this, substitute $\rho_{it}^j = \frac{R_{it}^j}{\sum_k R_{it}^k}$ into equation (34) to obtain $\mu_{it}^j = \theta_{Mg(i)}(\mathbf{E}_{it}) \frac{\sum_k R_i^k}{E_{Mi}}$, which is constant across j.

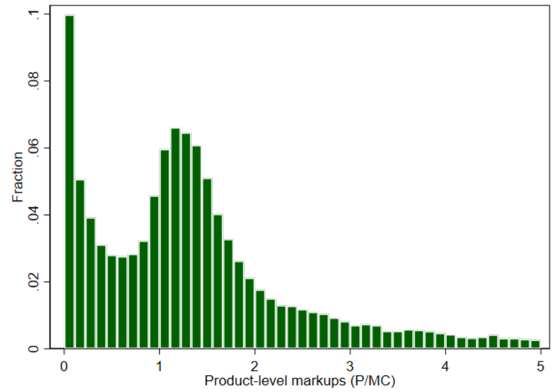
[45] These are data-driven in that they use actual data as opposed to the Monte Carlo simulations of section 3.2.

[46] Both sets of markups are for Indian producers. However, panel (A) uses the ASI and therefore only considers *plant*-product levels markups for the years 2002 to 2008, while panel (B) uses Prowess data to construct *firm*-product level markups for the years 1989 to 2003. In both figures we truncate the distribution at 5 to focus on the lowest set of markups we estimate.

Figure 2: Product-level Markups: All Plant-products

(A) ASI Data (B) Prowess Data





Notes: This figure displays histograms of firm/plant-product markups. Panel (A) presents plant-product markups calculated according to equations (33) and (34) using the ASI data. Panel (B) shows firm-product markups taken from replication files for De Loecker et al. (2016) that uses the Prowess data. We trim markups ≥ 5 for scale.

is unexpected in standard profit maximizing models of firm behavior and likely reflects some degree of model misspecification.

We quantify this possible misspecification by comparing the ratio of the number of estimated product-level markups that are close to zero ($\mu_{it}^j < 0.5$), to the number of markups above 0.5 but below 5.[47] We refer to the latter set as *reasonable markups* and to this ratio as the *zeros outlier ratio*. We can also use this metric to examine the long right tail in estimated markups.[48] We quantify the importance of the right tail by calculating the ratio of the number of markups greater than 5 to the number of reasonable markups.[49] We refer to this latter ratio as the *tail outlier ratio*.

To explore whether model misspecification is a plausible culprit, we systematically vary the assumptions used to estimate markups with the goal of pinning down the precise mechanisms that are generating these results. As a benchmark, the two far-left vertical bars in Figure 3 display the *zeros outlier ratio* and the *tail outlier ratio* (respectively) for the firm-

[47]While one might take $\mu_{it}^j \geq 1$ to be "reasonable", i) product-level markups can be less than 1 with complementary goods, and ii) when the correct model is perfect competition and there is some (classical) measurement error in revenues or expenditures, we may find estimated markups distributed above and below 1.

[48]While not directly plotted in Figure 2, estimated markups here have a fairly fat, right tail, with the 95th percentile of product level markups being 57.83, and the 99th percentile involving a markup of 546.

[49]We compare our outlier counts to the reasonable markup counts rather than simply examining proportions of markups of each type so that one outlier ratio becoming more likely does not mechanically decrease the severity of the other outlier ratio.

product markups from Figure 2, panel (A). The number of zero markup estimates is about 20% the number of reasonable markups estimates. Combined with the right tail outliers (slightly less than 20%), the sample of total outlier markups is roughly 40% the size of reasonable markup sample. The problem gets worse if we focus only on markups estimated at multi-product plants as seen in the second set of bars ("MP only"). Zero markup observations are over 40% of the size of the reasonable markups sample, and accounting for right tail markups generates a total outlier sample 70% the size of the sample reasonable markups, suggesting that to 40% of the plant-product markups are outliers $(0.7/(0.7+1) = 0.41)$. This suggests that mispecification of the input allocation rule may be partly responsible for the large number of markups with undesirable properties.

The next three sets of bars ("*No* W_j","*Phi*", "*Unit Adj*") show that adjusting the input allocation rule by *i*) removing the input price controls function, *ii*) purging output of measurement error, and *iii*) carrying out additional units adjustments do not substantively change these results. Appendix L describes these changes in detail. In the final two bars, we allocate inputs by revenue shares. This rule is interesting because it allows for MRT heterogeneity as in Orr (2022) but rules out within-plant markup differences. This simple allocation rule almost completely resolves our zero markup puzzle; the number of zero markup observations is only 2% the size of the reasonable sample and tail outliers are only 0.05%. Figure 4, panel (A) plots the histogram of product-level markups for our baseline multi-product sample, and compares it to panel (B) which shows the revenue share multi-product markup sample. Comparing panel B to panel A, both types of outliers become far less common.

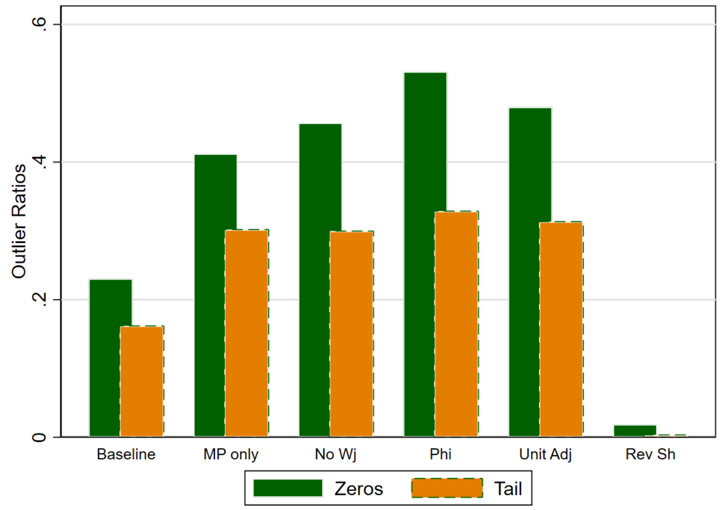
While allocating inputs based on revenue shares almost completely solves the zero markup puzzle, it is unsatisfying because this requires that we abstract from within-plant markup differences, defeating the point of estimating product-level markups. However, this suggests that establishing ways to allow for MRT heterogeneity that do not restrict within-firm markups may be fruitful.

7.1.2 Do Parameterized Productivity Ladders Help?

Can parameterized models allowing for MRT heterogeneity help in eliminating the mass of markups at zero? If the answer is "yes", then these fixes may allow researchers to use input allocation rules for existing models with minor changes. While we find that such parametrizations may help, results are too sensitive to provide conclusive support for whether productivity ladders can eliminate the outlier problems.

To see this, consider a simple parameterized input allocation rule that allows for MRT

Figure 3: Plant-Product Markup Outlier Ratios Across Methods



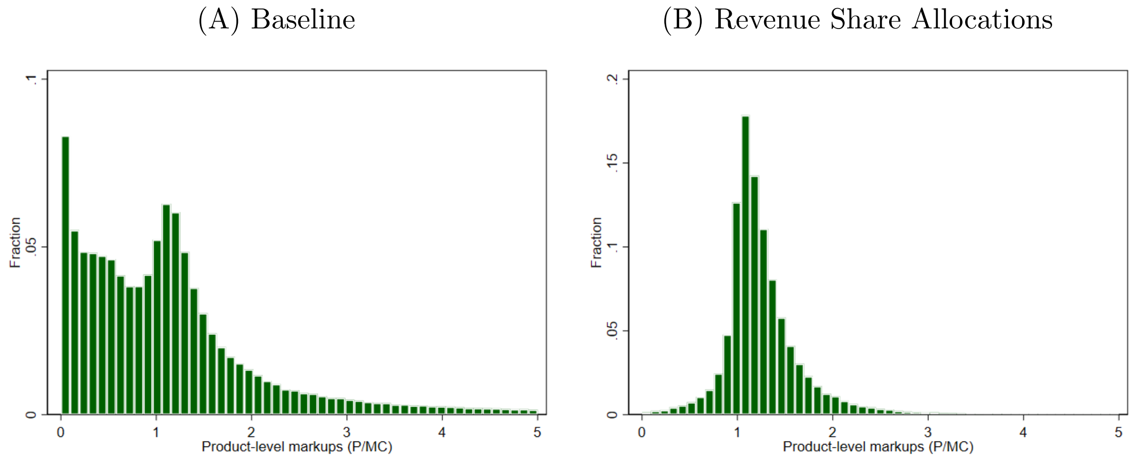
Notes: The above bar chart plots two outlier ratios for different models of product-level markups. Letting $\mathbb{Z}$ denote the set of product-level markups where $\mu \leq 0.5$, $\mathbb{T}$ denote the set of product-level markups where $\mu \geq 5$, $\mathbb{R}$ denote product-level markups between 0.5 and 5, and $|.|$ denote the number of elements in a given set, the *Zeros* outlier ratio is defined as $\frac{|\mathbb{Z}|}{|\mathbb{R}|}$ and the *Tail* outlier ratio is defined as $\frac{|\mathbb{T}|}{|\mathbb{R}|}$. *Baseline* refers to the product-level markups in Panel A of Figure 2. *MP Only* only considers the subset of markups produced by multi-product plants. *No* W_j removes the input price control function adjustment $\widehat{W_{it}^j}$ in equation (33). *Phi* uses predicted output $\widehat{\Phi}_{it}^j$ in place of realized output Y_{it}^j in equaiton (33). *Units Adj* uses the modified input allocation rule $\rho_{it}^j = \left[\left(\frac{Y_{it}^j}{\overline{A}^j}\right)^{\frac{1}{\phi}} \widehat{W}_{it}^j\right] / \left[\sum_k \left(\frac{Y_{it}^k}{\overline{A}^k}\right)^{\frac{1}{\phi}} \widehat{W}_{it}^k\right]$, where $\overline{A}^j$ is average TFPQ obtained from single product-firms by 5 digit product code j. *Rev Sh* uses revenue shares to allocate inputs in equation (34). All models except for *Baseline* look only at markups generated by multi-product plants. See Appendix L for further details.

heterogeneity (see Appendix K for derivation):

$$\rho_{it}^j = \frac{\left(\frac{Y_{it}^j}{A_{it}^j}\right)^{\frac{1}{\phi}} \widehat{W}_{it}^j}{\sum_k \left(\frac{Y_{it}^k}{A_{it}^k}\right)^{\frac{1}{\phi}} \widehat{W}_{it}^k}. \tag{35}$$

If the econometrician knows $\{A_{it}^j\}_j$, then they can derive the allocation rules necessary for estimating firm-product markups. However, $\{A_{it}^j\}_j$ is generally unknown making this unfeasible without further restrictions. To provide such structure, we consider a variant of the Mayer et al. (2014) model of within-firm heterogeneity based on core competencies. Each firm faces a productivity ladder over all potential products j, with $j = 0$ indexing a "core" product that has the highest productivity within the firm. We assume that for all $j \neq 0$

Figure 4: Plant-Product Markups: Multi-Product-Plant-Product



Notes: The above two panels display a histogram of product level markups, for the subsample of multi-product plants. Panel (A) shows our baseline plant-product level markups calculated according to equations (33) and (34). Panel (B) product-level markups where we use within-plant revenue shares to allocate inputs. We trim markups ≥ 5 for scale.

$$A_i^j = A_i^0 \times (\delta)^j \tag{36}$$

where $\delta \in (0,1)$ is a free parameter that implies greater within-plant productivity dispersion as $\delta \to 0$.[50] We consider two different rankings.[51] First, we assume that the product with the highest revenue is the firm's core product, $j = 0$, their second highest revenue product is $j = 1$, and so on. Alternately, we assume that a firm's lowest revenue product is the firm's core product, $j = 1$ is their second lowest revenue product, and so on. While this second case may seem perverse, previous work including Jaumandreu and Yin (2016), Forlani et al. (2016), Atkin et al. (2017), and Orr (2022) documents that TFPQ can be negatively correlated with demand shifters (e.g. quality), and that these demand shifters often explain revenues better than productivity differences (Hottman et al. 2016, Eslava et al. 2023). In short, it may be that the lowest revenue products within a firm are the highest A_{it}^j values.[52]

Figure 5 plots the zero and tail outlier ratios for various assumed values of $1/\delta$ under these two cases of rankings. Panel A assumes that top revenue products are $j = 0$ and panel B assumes that bottom revenue products are $j = 0$. As $1/\delta$ increases, MRT heterogeneity

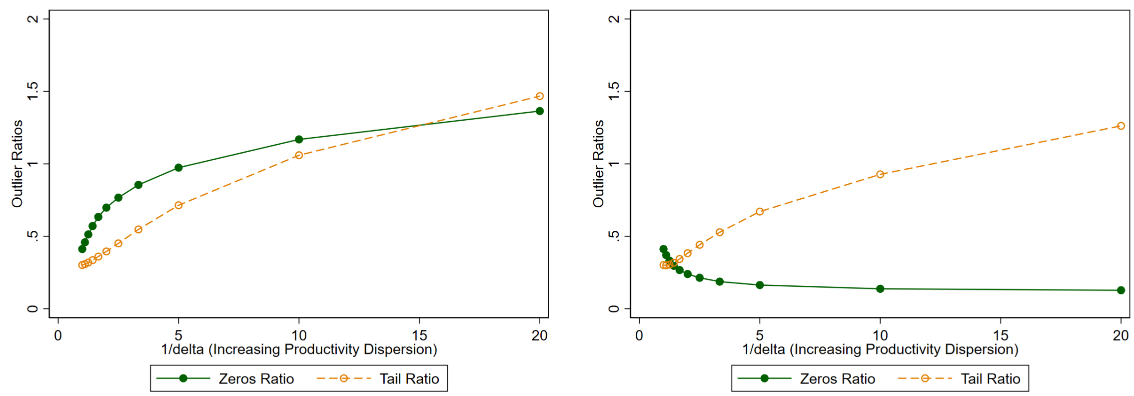
[50] With slight abuse of notation, j here indexes a plant-specific ranking of product-specific TFPQ.

[51] Unfortunately, we have no way of knowing what each plant's TFPQ ranking across products without first disciplining the input shares across product lines. We also report results based on quantity rankings in Appendix M. These results turn out to be similar qualitatively, although we find that revenue rankings we report in the main text perform better quantitatively in terms of solving the zeros puzzle.

[52] In this case, these are low value "cheap goods".

Figure 5: Product-level Markup Outlier Ratios with Productivity Ladders

(A) Core Product - Highest Revenue (B) Core Product - Lowest Revenue



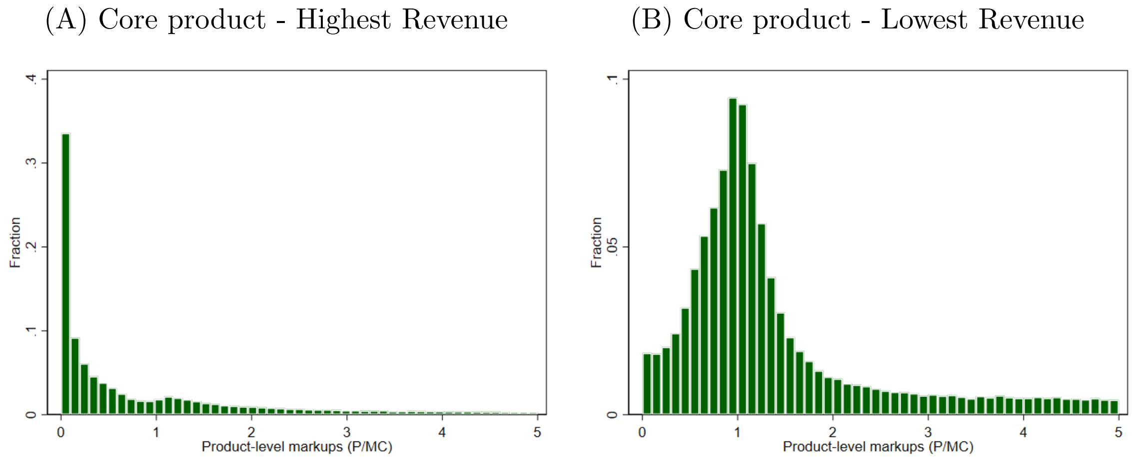
Notes: This figure reports outlier ratios for different models of product-level markups as a function of within firm productivity dispersion, parameterized by $1/\delta$. Letting $\mathbb{Z}$ denote the set of product-level markups where $\mu \leq 0.5$, $\mathbb{T}$ denote the set of product-level markups where $\mu \geq 5$, $\mathbb{R}$ denote product-level markups between 0.5 and 5, and $|.|$ denote the number of elements in a given set, the *Zeros* outlier ratio (represented by circles) is defined as $\frac{|\mathbb{Z}|}{|\mathbb{R}|}$, while the *Tail* outlier ratio (represented by triangles) is defined as $\frac{|\mathbb{T}|}{|\mathbb{R}|}$. Product-level markups calculated using equation (34) with $\rho^j_{it} = \left[\left(\frac{Y^j_{it}}{(\delta)^j}\right)^{\frac{1}{\phi}} \widehat{W}^j_{it}\right] / \left[\sum_k \left(\frac{Y^k_{it}}{(\delta)^k}\right)^{\frac{1}{\phi}} \widehat{W}^k_{it}\right]$. Panel (A) orders products j within a plant from highest to lowest revenue, with $j = 0$ having the highest revenue, $j = 1$ having the second highest revenue, and so on. Panel (B) orders products within a plant from lowest to highest revenue, with $j = 0$ denoting the lowest revenue product, $j = 1$ denoting the second lowest revenue product, and so on.

also increases as the steps along the productivity ladder become larger. The lines with triangles represent the tail outlier ratio while the lines with circles represent zeros outlier ratios. Assuming that products with small revenue shares have the highest productivity within the firm as in panel B helps reduce the share of zero ratio but increases the number of tail outlier ratios, diminishing its usefulness. When $1/\delta = 10$, the number of outliers actually exceeds the number of reasonable markups.[53] Assuming that high revenue products have the highest productivity as in panel A makes both outlier ratios worse.

Figure 6 plots the plant-product markup distribution for $1/\delta = 20$ for the case where the top-revenue product is a plant's core product (panel A), as well as the case where a firm's bottom revenue product is their core product (panel B). Again, when we treat high revenue products as having the highest TFPQ, the zero markups puzzle becomes much worse. If we assume low-revenue products have the highest TFPQ, the zero markups puzzle diminishes greatly but this proposed solution (again) generates a thicker right tail in markups consistent with continued model misspecification.

[53]More precisely, the ratio of the sum of zero and tail outliers to reasonable outliers is greater than 1.

Figure 6: Product-level Markups with Productivity Ladders



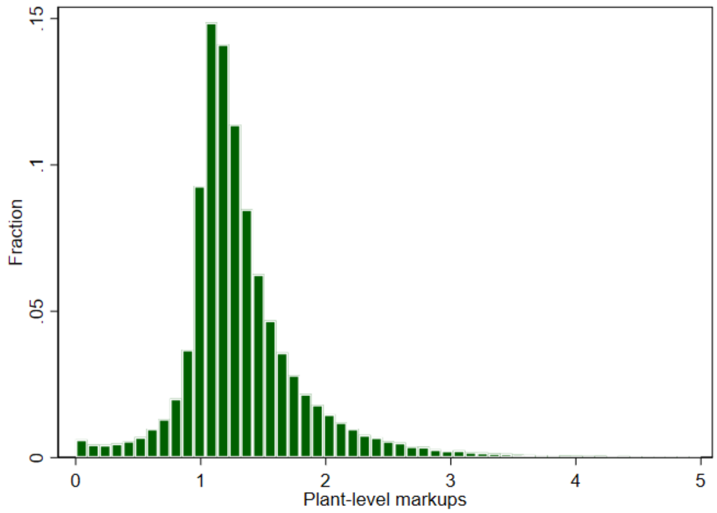
Notes: The above two panels display a histogram of product-level markups for the subsample of multi-product plants. Both panels calculate markups using equation (34) and $\rho_{it}^{j} = \left[\left(\frac{Y_{it}^{j}}{(\delta)^{j}}\right)^{\frac{1}{\phi}} \widehat{W}_{it}^{j}\right] / \left[\sum_{k}\left(\frac{Y_{it}^{k}}{(\delta)^{k}}\right)^{\frac{1}{\phi}} \widehat{W}_{it}^{k}\right]$ to calculate input shares, where $\delta = 0.05$. Panel (A) orders products within a plant from highest to lowest revenue, with $j = 0$ having the highest revenue, $j = 1$ having the second highest revenue, and so on. Panel (B) orders products within a plant from lowest to highest revenue, with $j = 0$ denoting the lowest revenue product, $j = 1$ denoting the second lowest revenue product, and so on. We trim markups ≥ 5 for scale.

MRT heterogeneity is limited in two important ways in the above exercises. First, in these exercises, within-plant heterogeneity takes on a very specific parametric form that rules out other forms of heterogeneity (see equation 36). Second, we force TFPQ to rise and fall across products within a firm using the coarse indicator of product performance. However, revenue is *not* a sufficient statistic for within-firm productivity dispersion as documented by Orr (2022).[54] Knowledge of the form of competition and within-firm markup dispersion is necessary to extract the required information from quantity and revenue data. The fact that this approach provides only a partial solution to the zero markup puzzle suggests that this simple model of productivity ladders is not sufficient to capture the true variation in the underlying data. However, this does suggest that allowing for more flexible patterns of MRT heterogeneity may be key to pinning down a reasonable estimate of product-level markups.[55]

[54]See also Appendix M for similar result where quantity rankings are used in place of revenue rankings.

[55]In Appendix O we consider two further sources of potential misspecification; assuming non-joint production when production may be joint, and misspecification of the production technology. We find that these cannot help resolve the zeros puzzle.

Figure 7: Baseline Plant-level Markups



Notes: The above histogram displays plant-level level markups according to equation (30) based on our baseline production function estimation routine described in the main text (technique 1). We trim markups ≥ 5 for scale.

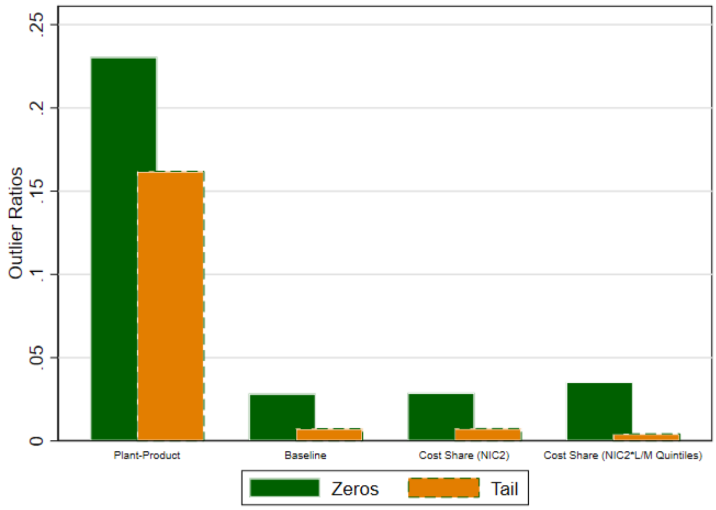
7.2 Plant Markups

Figure 7 displays a histogram of plant markups using technique 1: a control function estimator with a selection correction. Appendix N presents the estimated elasticities. Plant markups are well behaved, with a mean of around 1.5 without trimming outliers. Zero and right-tail outliers are now quite rate rare; the ratio of zero to reasonable markup estimates is 0.03, while the tail outlier ratio below 0.01.

Figures 8 and 9 compare outlier ratios and the distribution of plant markups using Technique 1 (the control function estimator for single-product firms) with two variants of Technique 2 (the cost-share approach): *i*) a simple cost-share approach where the production function parameters differ by two-digit NIC code; *ii*) the cost-share approach advocated in Raval (2023) which allows cost-share parameters to vary by labour-to-materials cost quintiles within each two-digit NIC code. Figure 8 reports outlier ratios for each method and Figure 9 reports average markups without any trimming with 90-10 percentiles. We also report the outlier ratios for our baseline plant-product markups in the first column of Figure 8 for scale.

Figures 8 and 9 show that results using Technique 1 and the first version of Technique 2 are similar. Outlier ratios appear to be almost identical (Figure 8) and the average markup only differs by 0.02 (Figure 9). It is also reassuring to see that the 90-10 ratios are remarkably similar across these two approaches, differing by less than 0.04 (Figure 9). For researchers unsure as to which approach to use in their own work— a cost share or an estimator based

Figure 8: Markup Outlier Ratios Across Methods



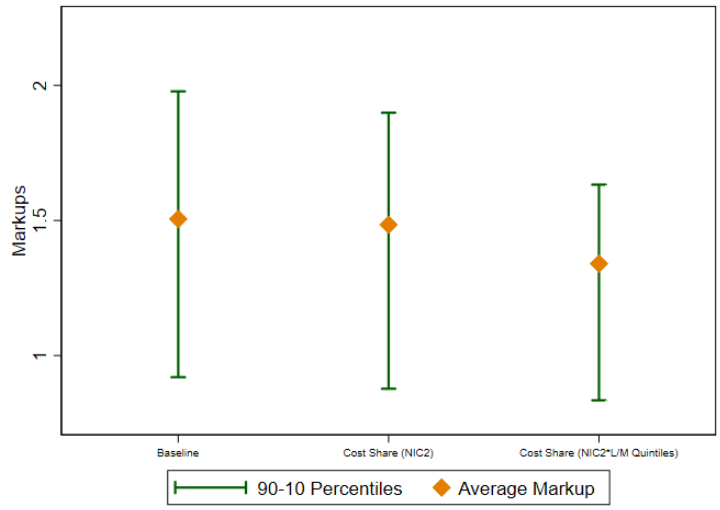
Notes: The above bar chart plots two outlier ratios for different models of product and plant-level markups. Letting $\mathbb{Z}$ denote the set of product-level markups where $\mu \leq 0.5$, $\mathbb{T}$ denote the set of product-level markups where $\mu \geq 5$, $\mathbb{R}$ denote product-level markups between 0.5 and 5, and $|.|$ denote the number of elements in a given set, the *Zeros* outlier ratio is defined as $\frac{|\mathbb{Z}|}{|\mathbb{R}|}$ and the *Tail* outlier ratio is defined as $\frac{|\mathbb{T}|}{|\mathbb{R}|}$. *Plant-Product* refers to the product-level markups from panel (A) of Figure 2. *Baseline* refers to the plant-level markups in Figure 7. *Cost Share (NIC2)* obtains plant-level markups after estimating the materials output elasticity using industry level cost shares for each two-digit NIC industry code. *Cost Share (NIC2*L/M Quintiles)* obtains plant-level markups using a seperate cost-share estimator for each labour-to-materials quintile $\times$ two-digit NIC industry.

on single-product firms — it is useful to know that both approaches find consistent evidence for some degree of across-plant misallocation due to market power.[56]

The more flexible cost-share approach recommended by Raval (2023) to deal with factor augmenting productivity differences implies slightly lower markups on average. In particular, once we allow the Cobb-Douglas parameters to vary by labour-to-materials quintiles within each industry, the average markup falls from 1.5 to 1.34. Similarly, dispersion in markups also falls using this approach, with the 90th to 10th percentile difference falling to around 0.80, which is about 25% smaller than the baseline dispersion. These differences are important

[56]However, it is worth emphasizing that our baseline approach using single-product firms estimates a single set of production-function parameters for the entire Indian economy, while the cost-share approach estimates a separate Cobb-Douglas technology by two-digit NIC code. Note, however, that since the translog specification allows output elasticities to vary across *all* plants (since output elasticities are functions of input use), while this cost-share approach forces all output elasticities to be identical for all firms in the same 2-digit NIC code, it's not clear which approach should be regarded as "more" flexible. In Appendix P, we also report results for single-product firm translog production functions for a subset of India's largest industries, but find that occasionally our approach leads to some unreasonable (e.g. negative) output elasticities for many plants (See Appendix N), and as a result there are more zero and outlier issues, which we believe are driven by insufficient sample size concerns.

Figure 9: Plant-level Markup Dispersion Across Methods



Notes: The above plots the mean plant-level markup, as well as the 90th to 10th percentile of markups, for four different ways of measuring plant-level markups. *Baseline* refers to the plant-level markups in Figure 7 (technique 1). *Cost Share (NIC2)* obtains plant-level markups after estimating the materials output elasticity using industry level cost shares for each two-digit NIC industry code (technique 2i). *Cost Share (NIC2*L/M Quintiles)* obtains plant-level markups using a separate cost-share estimator for each labour-to-materials quintile × two-digit NIC industry (technique 2ii).

and potentially point to the importance of allowing for factor-specific productivity differences as noted by Raval (2022, 2023). On the other hand, note that all methods clearly imply sizeable differences in plant markups across plants, which indicates that output market power is responsible for some degree of misallocation in the Indian economy.

8 Conclusion

In this paper, we considered two important problems. First, we asked how to identify plant-product markups using standard production data. We found that disciplining the magnitude of MRT heterogeneity is of first-order importance, while determining whether production was joint or non-joint was less empirically relevant. We see further work on new and novels ways deal with MRT heterogeneity as a very valuable path for future research. In particular, in this paper we only considered some simple, stylized approaches to deal with this problem, but expect that further progress can be made beyond our modest first-pass at this problem.

Another key problem we tackled in this paper was how to interpret the firm markups uncovered by the De Loecker and Warzynski (2012) approach. We showed that the firm markup can be interpreted as a cost-weighted average of firm-product markups under fairly general conditions, and that this object is a welfare relevant statistic that can help uncover the magnitude of across-firm misallocation due to market power. We also showed that a number of popular estimators, including the cost-share approach, as well as a control function approach developed in De Loecker et al. (2016), can be applied to uncover this object, even in joint production settings. Empirically, we found that both cost-share and control function approaches generated similar plant markup estimates in the Indian manufacturing sector, although we also found that adjusting for factor-specific productivity differences can lead to lower markups levels as well as dispersion. We hope these identification results shall prove useful for uncovering market power in other settings.

References

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer, "Identification Properties of Recent Production Function Estimators," *Econometrica*, 2015, *83*, 2411–2451.

Argente, David, Sara Moreira, Ezra Oberfield, and Venky Venkateswaran, "Scalable Expertise," 2020.

Arkolakis, Costas, Sharat Ganapati, and Marc-Andreas Muendler, "The extensive margin of exporting products: A firm-level analysis," *American Economic Journal: Macroeconomics*, 2021, *13* (4), 182–245.

Atalay, Enghin, "Materials prices and productivity," *Journal of the European Economic Association*, 2014, *12* (3), 575–611.

Atkin, David, Amit K Khandelwal, and Adam Osman, "Exporting and firm performance: Evidence from a randomized experiment," *The quarterly journal of economics*, 2017, *132* (2), 551–615.

Backus, Matthew, "Why is productivity correlated with competition?," *Econometrica*, 2020, *88* (6), 2415–2444.

Baldwin, John and Wulong Gu, "The impact of trade on plant scale, production-run length and diversification," in "Producer Dynamics: New Evidence from Micro Data," University of Chicago Press, 2009, pp. 557–592.

Baumol, J. William, John C. Panzar, and Robert D. Willig, *Contestable Markets and the Theory of Industry Structure*, New York: Harcourt College Pub, 1982.

Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott, "Multiproduct firms and trade liberalization," *Quarterly Journal of Economics*, 2011, *126* (3), 1271–1318.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63*, 242–262.

Blum, Bernardo S., Sebastian Claro, Ignatius Horstmann, and David A. Rivers, "The ABCs of Firm Heterogeneity: The Effects of Demand and Cost Differences on Exporting," *Working Paper*, 2018.

Boehm, Johannes and Ezra Oberfield, "Misallocation in the Market for Inputs: Enforcement and the Organization of Production," *Working Paper*, 2018.

Bond, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch, "Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data," *Journal of Monetary Economics*, 2021, *121*, 1–14.

Bresnahan, Timothy F, "Empirical studies of industries with market power," *Handbook of industrial organization*, 1989, *2*, 1011–1057.

Caves, Douglas W., Laurits R. Christensen, and W. Erwin Diewert, "The economic theory of index numbers and the measurement of input, output, and productivity," *Econometrica*, 1982, pp. 1393–1414.

Collard-Wexler, Allan and Jan De Loecker, "Reallocation and technology: Evidence from the US steel industry," *American Economic Review*, 2015, *105* (1), 131–171.

_ and _, "Productivity and Capital Measurement Error," *Working Paper*, 2021.

Dhyne, Emmanuel, Amil Petrin, Valerie Smeets, and Frederic Warzynski, "Theory for Extending Single-Product Production Function Estimation to Multi-Product Settings," *Working Paper*, 2022.

_ , _ , _ , and _, "Multi Product Firms, Import Competition, and the Evolution of Firm-product Technical Efficiencies," *Working Paper*, 2023.

Ding, Xiang, "Industry linkages from joint production," *Work. Pap., Georgetown Univ., Washington, DC*, 2022.

Doraszelski, Ulrich and Jordi Jaumandreu, "Using cost minimization to estimate markups," *CEPR Discussion Paper No. DP14114*, 2019.

Eckel, Carsten and J. Peter Neary, "Multi-product firms and flexible manufacturing in the global economy," *Review of Economic Studies*, 2010, *77* (1), 188–217.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, "Competition, markups, and the gains from international trade," *American Economic Review*, 2015, *105* (10), 3183–3221.

_ , _ , and _, "How costly are markups?," 2022.

Eslava, Marcela, John Haltiwanger, and Nicolas Urdaneta, "The Size and Life-Cycle Growth of Plants: The Role of Productivity, Demand and Wedges," *Working Paper*, 2023.

Flynn, Zach, James Traina, and Amit Gandhi, "Measuring markups with production data," *Available at SSRN 3358472*, 2019.

Forlani, Emanuele, Ralf Martin, Giordiano Mion, and Mirabelle Muûls, "Unraveling firms: Demand, productivity and markups heterogeneity," *Working Paper*, 2016.

Foster, Lucia, John Haltiwanger, and Chad Syverson, "Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?," *American Economic Review*, 2008, *98* (1), 394–425.

Gandhi, Amit, Salvador Navarro, and David Rivers, "On the Identification of Gross Output Production Functions," *Journal of Political Economy*, 2020, *128*, 2973–3016.

Gong, Binlei and Robin C Sickles, "Resource allocation in multi-divisional multi-product firms," *Journal of Productivity Analysis*, 2021, *55* (2), 47–70.

Grieco, Paul, Joris Pinkse, and Margaret Slade, "Brewed in North America: Mergers, marginal costs, and efficiency," *International Journal of Industrial Organization*, 2018, *59*, 24–65.

Grieco, Paul LE and Ryan C McDevitt, "Productivity and quality in health care: Evidence from the dialysis industry," *The Review of Economic Studies*, 2017, *84* (3), 1071–1105.

_ **, Shengyu Li, and Hongsong Zhang**, "Production function estimation with unobserved input price dispersion," *International Economic Review*, 2016, *57* (2), 665–690.

Hall, Robert E, "The specification of technology with several kinds of output," *Journal of Political Economy*, 1973, *81* (4), 878–892.

_ , "Market structure and macroeconomic fluctuations," *Brookings papers on economic activity*, 1986, *1986* (2), 285–338.

Hottman, Colin J., Stephen J. Redding, and David E. Weinstein, "Quantifying the sources of firm heterogeneity," *Quarterly Journal of Economics*, 2016, *131* (3), 1291–1364.

Jaumandreu, Jordi and Heng Yin, "Cost and product advantages: A firm-level model for the chinese exports and industry growth," *Working Paper*, 2016.

Johnes, Geraint, "Costs and industrial structure in contemporary British higher education," *The Economic Journal*, 1997, *107* (442), 727–737.

Kasahara, Hiroyuki and Yoichi Sugita, "Nonparametric Identification of Production Function, Total Factor Productivity, and Markup from Revenue Data," *Total Factor Productivity, and Markup from Revenue Data (October 28, 2020)*, 2020.

Khmelnitskaya, Ekaterina, Guillermo Marshall, and Scott Orr, "Identifying Scope Economies using Demand-Side Data," *Working Paper*, 2023.

Kimball, Miles S, "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit and Banking*, 1995, *27* (4), 1241–1277.

Kirov, Ivan and James Traina, "Measuring markups with revenue data," *Available at SSRN 3912966*, 2021.

Levinsohn, James and Amil Petrin, "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 2003, *70*, 317–341.

Loecker, Jan De and Chad Syverson, "An industrial organization perspective on productivity," in "Handbook of industrial organization," Vol. 4, Elsevier, 2021, pp. 141–223.

_ **and Frederic Warzynski**, "Markups and Firm-Level Export Status," *American Economic Review*, 2012, *102*, 2437–2471.

_ **and Jan Eeckhout**, "Global market power," Technical Report, National Bureau of Economic Research 2018.

_ **and Pinelopi K. Goldberg**, "Firm Performance in a Global Market," *The Annual Review of Economics*, 2014.

_ **, Jan Eeckhout, and Gabriel Unger**, "The rise of market power and the macroeconomic implications," *The Quarterly Journal of Economics*, 2020, *135* (2), 561–644.

__ , **Pinelopi Goldberg, Amit Khandelwal, and Nina Pavcnik**, "Prices,Markups and Trade Reform," *Econometrica*, 2016, *84* (2), 445–510.

Maican, Florin and Matilda Orth, "Determinants of economies of scope in retail," *International Journal of Industrial Organization*, 2021, *75*, 102710.

Mayer, Thierry, Marc J. Melitz, and Gianmarco IP Ottaviano, "Market size, competition, and the product mix of exporters," *American Economic Review*, 2014, *104* (2), 495–536.

Mundlak, Yair, "Specification and estimation of multiproduct production functions," *Journal of Farm Economics*, 1963, *45* (2), 433–443.

Olley, G. Stephen and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 1996, *64*, 1263–1297.

Orr, Scott, "Within-Firm Productivity Dispersion: Estimates and Implications," *Journal of Political Economy*, 2022, *130* (11), 2771–2828.

__ **and Mokhtar Tabari**, "Decomposing the Within-Firm Productivity Gains From Import Competition: Evidence from India," *Working Paper*, 2022.

Raval, Devesh, "Testing the production approach to markup estimation," *Review of Economic Studies (Forthcoming)*, 2022.

__ , "A Flexible Cost Share Approach to Markup Estimation," *Working Paper*, 2023.

Ridder, Maarten De, Basile Grassi, and Giovanni Morzenti, "The Hitchhiker's Guide to Markup Estimation," *Working Paper*, 2022.

Roux, Nicolas De, Marcela Eslava, Santiago Franco, and Eric Verhoogen, "Estimating production functions in differentiated-product industries with quantity information and external instruments," Technical Report, National Bureau of Economic Research 2021.

Samuelson, Paul A, "The fundamental singularity theorem for non-joint production," *International Economic Review*, 1966, *7* (1), 34–41.

Syverson, Chad, "Market structure and productivity: A concrete example," *Journal of Political Economy*, 2004, *112* (6), 1181–1222.

Valmari, Nelli, "Estimating Production Functions of Multiproduct Firms," Technical Report, Research Institute of the Finnish Economy 2016.

Appendix

A Proof of Theorem 1

The Proof of Theorem 1 relies on the following intermediate results:

Lemma 1 *Suppose a firm operates a non-joint technology with product-line production functions* $Y_i^j = A_i^j F_i(\mathbf{S}_i^j \circ \mathbf{X}_i)$*, where* $F_i(\mathbf{X}_i)$ *continuous, differentiable, strictly increasing in all arguments, quasi-concave, and homogeneous of degree* $\phi_i > 0$*. Then the firm's output distance function is given by:*

$$D_i^{NJ}(\mathbf{Y}_i, \mathbf{X}_i; F_i(.), \{A_i^j\}_j) = \frac{\left(\sum_j \left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\phi}}\right)^{\phi}}{F_i(\mathbf{X}_i)} \tag{37}$$

Proof. Given the proposed product-line production functions, the firm's output distance function will satisfy:

$$\begin{aligned} D_i^{NJ}(\mathbf{Y}_i, \mathbf{X}_i; F_i(.), \{A_i^j\}_j) \equiv \min_{\delta, \{\mathbf{S}_i^j\}_j} \quad & \delta \\ \text{s.t.:} \quad & \frac{Y^j}{\delta} = A_i^j F_i\left(\mathbf{S}_i^j \circ \mathbf{X}_i\right) \ \forall j \\ & \sum_j S_{ni}^j = 1, \ \forall n \end{aligned} \tag{38}$$

The Lagrangian for the optimization problem in (38) is given by:

$$L_i = \delta + \sum_j \lambda_i^j \left(\frac{Y^j}{\delta} - A_i^j F_i\left(\mathbf{S}_i^j \circ \mathbf{X}_i\right)\right) + \sum_n \mu_{ni} \left(\sum_j S_{ni}^j - 1\right) \tag{39}$$

Since $F_i(.)$ is quasi-concave, any solution $\Gamma_i \equiv \left(\{\mathbf{S}_i^j\}_j, \delta, \{\lambda_i^j\}_j, \{\mu_{ni}\}_n\right)$ to (38) will satisfy the following first order conditions:

$$FOC_{S_{ni}^j}(\Gamma_i) \equiv -\lambda_i^j A_i^j \frac{\partial F_i}{\partial X_i}\left(\mathbf{S}_i^j \circ \mathbf{X}_i\right) X_i + \mu_{ni} = 0 \quad \forall (j,n) \tag{40}$$

$$FOC_{\delta}(\Gamma_i) \equiv 1 - \sum_j \lambda_i^j \frac{Y_i^j}{\delta^2} = 0 \tag{41}$$

$$FOC_{\lambda_i^j}(\Gamma_i) \equiv \frac{Y^j}{\delta} - A_i^j F_i\left(\mathbf{S}_i^j \circ \mathbf{X}_i\right) = 0 \quad \forall j \tag{42}$$

$$FOC_{\mu_{ni}}(\Gamma_i) \equiv \sum_j S_{ni}^j - 1 = 0 \quad \forall n \tag{43}$$

It is easily verified that the following $\Gamma_i^* \equiv \left(\{\mathbf{S}_i^{j*}\}_j, \delta^*, \{\lambda_i^{j*}\}_j, \{\mu_{ni}^*\}_n\right)$ satisfies equations (40) through (43), and therefore solves (38):

$$S_{ni}^{j*} = \frac{\left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\phi}}}{\sum_k \left(\frac{Y_i^k}{A_i^k}\right)^{\frac{1}{\phi}}} \tag{44}$$

$$\delta^* = \frac{\left(\sum_j \left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\phi}}\right)^{\phi}}{F_i(\mathbf{X}_i)} \tag{45}$$

$$\lambda_i^{j*} = \left(\sum_j \left(\frac{Y_i^j}{A_i^j F\left(\mathbf{X}_i\right)}\right)^{\frac{1}{\phi}}\right)^{2\phi} \frac{\left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\phi}}}{\sum_k \left(\frac{Y_i^k}{A_i^k}\right)^{\frac{1}{\phi}}} \frac{1}{Y_i^j} \tag{46}$$

$$\mu_{ni}^* = \frac{\partial F\left(\mathbf{X}_i\right)}{\partial X_i} \frac{X_i}{F\left(\mathbf{X}_i\right)} \left(\sum_j \left(\frac{Y_i^j}{A_i^j F\left(\mathbf{X}_i\right)}\right)^{\frac{1}{\phi}}\right)^{\phi} \tag{47}$$

It immediately follows that $D_i^{NJ}(\mathbf{Y}_i, \mathbf{X}_i; F_i(.), \{A_i^j\}_j) = \frac{\left(\sum_j \left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\phi}}\right)^{\phi}}{F_i(\mathbf{X}_i)}$ by equation (45). □

Lemma 2 *Suppose Assumptions 1 and 2 hold. If there exists $\widetilde{F}_i^j(.)$ such that for all $(\mathbf{Y}_i, \mathbf{X}_i) \in \mathbb{P}_i^F$ there exist input shares $\{\mathbf{S}_i^j\}_j$ such that $Y_i^j = \widetilde{F}_i^j(\mathbf{S}_i^j \circ \mathbf{X}_i)$, the only $\widetilde{F}_i^j(.)$ satisfying this condition are $\widetilde{F}_i^j(.) = A_i^j F_i(.)$, where $F_i(.)$ is defined in Assumptions 1 and 2; otherwise, no such $\widetilde{F}_i^j(.)$ exist.*

Proof. We establish this by contradiction. In particular, suppose not, so that there exists $\widetilde{F}_i^j(.)$ such that for all $(\mathbf{Y}_i, \mathbf{X}_i) \in \mathbb{P}_i^F$ there exist input shares $\{\mathbf{S}_i^j\}_j$ such that $Y_i^j = F_i^j(\mathbf{S}_i^j \circ \mathbf{X}_i)$, and there exists at least one $\widehat{\mathbf{X}}_i$ such that $F_i^j(\widehat{\mathbf{X}}_i) \neq A_i^j F_i(\widehat{\mathbf{X}}_i)$. By Assumption 3, for any $\mathbf{X}_i$ we can find $\left(\widehat{\mathbf{Y}}_i^j, \mathbf{X}_i\right) \in \mathbb{P}_i^F$. Denote this value of Y_i^j, given $\mathbf{X}_i$, as $\widehat{Y}_i^j\left(\mathbf{X}_i\right)$. Assumption 2 implies that $\widehat{Y}_i^j\left(\mathbf{X}_i\right) = A_i^j F_i\left(\mathbf{X}_i\right)$. Since $F_i\left(\mathbf{X}_i\right)$ is strictly increasing in all it's arguments, this implies that $\widehat{Y}_i^j\left(\mathbf{X}_i\right)$ is also strictly increasing in all it's arguments.

Let $\mathbf{S}_i^j\left(\widehat{Y}_i^j\left(\mathbf{X}_i\right), \mathbf{X}_i\right)$ denote the optimal input shares for the proposed production function $F_i^j(.)$. The fact that $\widehat{Y}_i^j\left(\mathbf{X}_i\right)$ is strictly increasing in all it's arguments implies that for any $\left(\widehat{Y}_i^j\left(\mathbf{X}_i\right), \mathbf{X}_i\right) \in \mathbb{P}_i^F$, we must have $\mathbf{S}_i^j\left(\widehat{Y}_i^j\left(\mathbf{X}_i\right), \mathbf{X}_i\right) = 1$. To prove this, suppose not, so that $Y_i^j\left(\mathbf{X}_i\right) = F_i^j(\mathbf{S}_i^j\left(\widehat{Y}_i^j\left(\mathbf{X}_i\right), \mathbf{X}_i\right) \circ \mathbf{X}_i)$, where $\mathbf{S}_i^j\left(\widehat{Y}_i^j\left(\mathbf{X}_i\right), \mathbf{X}_i\right)$ is strictly less than 1 for at least one input. Denote $\widetilde{\mathbf{X}}_i = \mathbf{S}_i^j\left(\widehat{Y}_i^j\left(\mathbf{X}_i\right), \mathbf{X}_i\right) \circ \mathbf{X}_i$, where $\widetilde{\mathbf{X}}_i$ is strictly less than

$\mathbf{X}_i$ in at least one dimension. Since $\widehat{Y}_i^j(\mathbf{X}_i)$ is strictly increasing in all it's arguments, this means that $Y_i^j(\mathbf{X}_i) = F_i^j(\widetilde{\mathbf{X}}_i) > Y_i^j\left(\widetilde{\mathbf{X}}_i\right)$,which further implies

$$\frac{Y_i^j\left(\widetilde{\mathbf{X}}_i\right)}{F_i^j(\widetilde{\mathbf{X}}_i)} < 1$$

Next, consider $\left(\widehat{Y}_i^j\left(\widetilde{\mathbf{X}}_i\right), \widetilde{\mathbf{X}}_i\right) \in \mathbb{P}_i^F$. Since the production possibility frontier provides maximal quantities of output, given inputs, it must be that:

$$F_i^j(\mathbf{S}_i^j \circ \widetilde{\mathbf{X}}_i) \leq F_i^j\left(\mathbf{S}_i^j\left(\widehat{Y}_i^j\left(\widetilde{\mathbf{X}}_i\right), \widetilde{\mathbf{X}}_i\right) \circ \widetilde{\mathbf{X}}_i\right)$$

for any feasible $\mathbf{S}_i^j$. Note that this implies, when combined with the above inequality:

$$\frac{Y_i^j\left(\widetilde{\mathbf{X}}_i\right)}{F_i^j\left(\mathbf{S}_i^j\left(\widehat{Y}_i^j\left(\widetilde{\mathbf{X}}_i\right), \widetilde{\mathbf{X}}_i\right) \circ \widetilde{\mathbf{X}}_i\right)} \leq \frac{Y_i^j\left(\widetilde{\mathbf{X}}_i\right)}{F_i^j(\widetilde{\mathbf{X}}_i)} < 1$$

Which is a contradiction since if $\left(\widehat{Y}_i^j\left(\widetilde{\mathbf{X}}_i\right), \widetilde{\mathbf{X}}_i\right) \in \mathbb{P}_i^F$ we must have $\frac{Y_i^j(\widetilde{\mathbf{X}}_i)}{F_i^j(\mathbf{S}_i^j(\widehat{Y}_i^j(\widetilde{\mathbf{X}}_i),\widetilde{\mathbf{X}}_i)\circ\widetilde{\mathbf{X}}_i)}$=1. Therefore, $\mathbf{S}_i^j\left(\widehat{Y}_i^j(\mathbf{X}_i), \mathbf{X}_i\right) = 1$ for all $\left(\widehat{Y}_i^j(\mathbf{X}_i), \mathbf{X}_i\right) \in \mathbb{P}_i^F$.

Next, choose any $\widehat{\mathbf{X}}_i$ such that $F_i^j(\widehat{\mathbf{X}}_i) \neq A_i^j F_i(\widehat{\mathbf{X}}_i)$, and consider $\left(\widehat{Y}_i^j\left(\widehat{\mathbf{X}}_i\right), \widehat{\mathbf{X}}_i\right) \in \mathbb{P}_i^F$. The previous argument has established $\widehat{Y}_i^j = F_i^j(\mathbf{S}_i^j \circ \widehat{\mathbf{X}}_i) = F_i^j(\widehat{\mathbf{X}}_i)$, i.e. $\mathbf{S}_i^j = 1$, while Assumption 2 implies $\widehat{Y}_i^j = A_i^j F(\widehat{\mathbf{X}}_i)$, which together imply $F_i^j(\widehat{\mathbf{X}}_i) = A_i^j F(\widehat{\mathbf{X}}_i)$, generating a contradiction. Therefore, the only $\widetilde{F}_i^j(.)$ that may support non-joint production are $\widetilde{F}_i^j(.) = A_i^j F_i(.)$.

□

We now turn to the statement and proof of Theorem 1, which will make use of these two lemmas.

Theorem (1)**.** *Suppose Assumptions 2 and 3 hold. Then a firm's technology is non-joint if and only if:*

$$D_i(\mathbf{Y}_i, \mathbf{X}_i) = \frac{\left(\sum_j \left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\phi_i}}\right)^{\phi_i}}{F_i(\mathbf{X}_i)} \tag{48}$$

Proof. The proof proceeds in two steps. First, in **Step 1** we show that if the firm's input distance function satisfies (48), then the technology is non-joint. Then in **Step 2** we show that if the technology is non-joint, it must have an output distance function given by (48).

Step 1: Suppose a firm's input distance function satisfies (48). To prove that this production possibilities set involves non-joint production, we must show that there exist $\widetilde{F}_i^j(.)$ such that for all $(\mathbf{Y}_i, \mathbf{X}_i) \in \mathbb{P}_i^F$, there exist input shares $\{\mathbf{S}_i^j\}_j$ such that $Y_i^j = \widetilde{F}_i^j(\mathbf{S}_i^j \circ \mathbf{X}_i)$.

Take $\widetilde{F}_i^j(.) = A_i^j F_i(.)$. Lemma 1 tells us that the output distance function for this proposed class of non-joint technologies $(F_i(.), \{A_i^j\}_j)$ has an output distance function equal to $D_i^{NJ}(\mathbf{Y}_i, \mathbf{X}_i; F_i(.), \{A_i^j\}_j) = \frac{\left(\sum_j \left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\phi_i}}\right)^{\phi_i}}{F_i(\mathbf{X}_i)}$.

Since the firm's output distance function satisfies (48), $(\mathbf{Y}_i, \mathbf{X}_i) \in \mathbb{P}_i^F$ is the same as the set of all $(\mathbf{Y}_i, \mathbf{X}_i)$ such that $\frac{\left(\sum_j \left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\phi_i}}\right)^{\phi_i}}{F_i(\mathbf{X}_i)} = 1$. However, since $D_i^{NJ}(\mathbf{Y}_i, \mathbf{X}_i; F_i(.), \{A_i^j\}_j) = \frac{\left(\sum_j \left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\phi_i}}\right)^{\phi_i}}{F_i(\mathbf{X}_i)}$ was generated under the assumption that $Y_i^j = A_i^j F_i(\mathbf{S}_i^j \circ \mathbf{X}_i)$, then if we consider the set of $(\mathbf{Y}_i, \mathbf{X}_i)$ such that $D_i^{NJ}(\mathbf{Y}_i, \mathbf{X}_i; F_i(.), \{A_i^j\}_j) = \frac{\left(\sum_j \left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\phi_i}}\right)^{\phi_i}}{F_i(\mathbf{X}_i)} = 1$, this immediately implies that for all $(\mathbf{Y}_i, \mathbf{X}_i) \in \mathbb{P}_i^F$, there exist input shares $\{\mathbf{S}_i^j\}_j$ such that $Y_i^j = \widetilde{F}_i^j(\mathbf{S}_i^j \circ \mathbf{X}_i)$

Step 2: Suppose the technology is non-joint. By Lemma 2, $Y_i^j = A_i^j F_i(.)$ is the only set of non-joint technologies consistent with Assumptions 1 and 2. Lemma 1 then immediately implies $D_i(\mathbf{Y}_i, \mathbf{X}_i) = \frac{\left(\sum_j \left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\phi}}\right)^{\phi}}{F_i(\mathbf{X}_i)}$. □

B A model of input sharing in production

In this Appendix, we show that the output distance function (13) can be obtained from a model of production where inputs are shared across product lines, similar to the CES cost function described in Chapter 15 of Baumol et al. (1982).

Suppose that each firm assembles final output using two classes of intermediate inputs; private intermediates, I_i^{rj}, and a public intermediate, I_i^p. Private intermediates are specialized inputs, such as product-line specific equipment or machinery, that can only be used in the production of product line j. Public intermediates, on the other hand, are non-rival and can be used in the production of all product lines at once— these might be thought of as managerial inputs, or buildings and structures that are useful for producing many things at once.

Each of these intermediate inputs are built entirely in-house, using the firm's vector of inputs $\mathbf{X}_i$. We let $\mathbf{X}_i^{rj}$ denote the vector of inputs allocated to producing the specialized private intermediate I_i^{rj}, and $\mathbf{X}_i^p$ denote a vector of inputs allocated to producing the public, non-rival intermediate, I_i^p. Because $\mathbf{X}_i^p$ contains public inputs across product lines, it does not have an index j. The production of the public and private intermediates within the firm are given by:

$$I_i^p = G(\mathbf{X}_i^p); \quad I_i^{rj} = G(\mathbf{X}_i^{jr})$$

Note that the aggregation of inputs to make the public and private intermediates are determined by the same function $G(\cdot)$. We then assume that the production function for product j at a given firm i is

$$Y_i^j = \frac{A_i^j}{\kappa}(I_i^{rj})^{\beta\phi}(I_i^p)^{(1-\beta)\phi} = \frac{A_i^j}{\kappa}G(\mathbf{X}_i^{rj})^{\beta\phi}G(\mathbf{X}_i^p)^{(1-\beta)\phi}. \tag{49}$$

where $\beta \in [0,1]$ governs the intensity of private intermediate use relative to the public intermediate use, $\beta\phi$ and $(1-\beta)\phi$ are the private and public returns to scale respectively, A_i^j is a Hicks-neutral productivity shifter specific to product line j, and $\kappa = (\beta)^{\beta\phi}(1-\beta)^{(1-\beta)\phi}$ is a constant. Note that even though there exist product-specific production functions inside the firm, the above defines a joint production technology because of the presence of public inputs. Therefore, product line production functions within the firm vary only due to the productivity shifters A_i^j.[57] We now make two assumptions that will facilitate the derivation of the output distance function $D_i(\mathbf{Y}_i, \mathbf{X}_i)$ for this firm.

Assumption 7 $G(\cdot) : \mathbb{R}^N \to \mathbb{R}$ *continuous and twice differentiable, equal to zero if any of its arguments are equal to zero, strictly increasing in all arguments, quasi-concave, and homogeneous of degree 1.*

Assumption 8 *Both static and dynamic inputs can be costlessly transferred between public and private use.*

The firm's output distance function specified in definition 2 can be characterised as

[57]We can also derive a closed form expression for a firm's output distance function that has varying total returns to scale by product line, i.e. ϕ_j. We show the derivation for the simple case here.

follows:

$$
\begin{aligned}
D_i(\mathbf{Y}_i, \mathbf{X}_i) \quad \equiv \quad & \min_{\delta, \mathbf{X}_i^p, \{\mathbf{X}_i^{rj}\}_j} \delta \\
\text{s.t.} \quad & \frac{Y_i^j}{\delta} \leq \frac{A_i^j}{\kappa} G(\mathbf{X}_i^{rj})^{\beta\phi} G(\mathbf{X}_i^p)^{(1-\beta)\phi} \quad \forall\, j \\
& X_{si}^p + \sum_j X_{si}^{rj} \leq X_{si} \quad \forall\, s
\end{aligned} \tag{50}
$$

The above optimization problem has the following Lagrangian:

$$
\mathcal{L} = \delta + \sum_j \lambda_i^j \left(\frac{Y_i^j}{\delta} - \frac{A_i^j}{\kappa} G(\mathbf{X}_i^{rj})^{\beta\phi} G(\mathbf{X}_i^p)^{(1-\beta)\phi} \right) + \sum_s \mu_{si} \left(X_{si}^p + \sum_j X_{si}^{rj} - X_{si} \right) \tag{51}
$$

Since the production functions are increasing in all inputs, all constraints will bind with equality. The first order necessary conditions where all constraints bind are given by:

$$
\mathrm{FOC}_\delta \left(\delta, \mathbf{X}_i^r, \mathbf{X}_i^p, \boldsymbol{\lambda}_i, \boldsymbol{\mu}_i\right) \quad \equiv \quad 1 - \sum_j \frac{\lambda_i^j Y_i^j}{\delta^2} = 0 \tag{52}
$$

$$
\mathrm{FOC}_{X_{si}^{rj}} \left(\delta, \mathbf{X}_i^r, \mathbf{X}_i^p, \boldsymbol{\lambda}_i, \boldsymbol{\mu}_i\right) \quad \equiv \quad \beta\phi \frac{G_{(s)}(\mathbf{X}_i^{rj})}{G(\mathbf{X}_i^{sr})} \frac{\lambda_i^j Y_i^j}{\delta} - \mu_{si} = 0 \quad \forall\, \{s, j\} \tag{53}
$$

$$
\mathrm{FOC}_{X_{si}^{p}} \left(\delta, \mathbf{X}_i^r, \mathbf{X}_i^p, \boldsymbol{\lambda}_i, \boldsymbol{\mu}_i\right) \quad \equiv \quad (1-\beta)\phi \frac{G_{(s)}(\mathbf{X}_i^{p})}{G(\mathbf{X}_i^{p})} \sum_j \frac{\lambda_i^j Y_i^j}{\delta} - \mu_{s} = 0 \quad \forall\, \{s\} \tag{54}
$$

$$
\begin{aligned}
\mathrm{FOC}_{\lambda_i^j} \left(\delta, \mathbf{X}_i^r, \mathbf{X}_i^p, \boldsymbol{\lambda}_i, \boldsymbol{\mu}_i\right) \quad \equiv \quad & \frac{Y_i^j}{\delta} - \frac{A_i^j}{\kappa} G(\mathbf{X}_i^{rj})^{\beta\phi} G(\mathbf{X}_i^p)^{(1-\beta)\phi} = 0 \\
& \& \quad \lambda_i^j > 0 \quad \forall\, \{j\}
\end{aligned} \tag{55}
$$

$$
\begin{aligned}
\mathrm{FOC}_{\mu_{si}} \left(\delta, \mathbf{X}_i^r, \mathbf{X}_i^p, \boldsymbol{\lambda}_i, \boldsymbol{\mu}_i\right) \quad \equiv \quad & X_{si}^p + \sum_j X_{si}^{rj} - X_{si} = 0 \\
& \& \quad \mu_{si} > 0 \quad \forall\, \{s\}
\end{aligned} \tag{56}
$$

where $G_{(s)}(\cdot) = \partial G(\cdot)/\partial X_{si}$, and $\boldsymbol{\lambda}_i$ and $\boldsymbol{\mu}_i$ denote the vectors of Lagrangian multipliers. Stacking equations (52) - (56) allows us to represent the solution to the cost minimization problem as a vector $(\delta, \mathbf{X}_i^r, \mathbf{X}_i^p, \boldsymbol{\lambda}_i, \boldsymbol{\mu}_i)$ that satisfies $\mathbf{FOC}\left(\delta, \mathbf{X}_i^r, \mathbf{X}_i^p, \boldsymbol{\lambda}_i, \boldsymbol{\mu}_i\right) = 0$ with $(\boldsymbol{\lambda}_i, \boldsymbol{\mu}_i) >> 0$.

Consider the following candidate solution for the input allocation rule:

$$X_{si}^{rj*} = \beta \frac{\left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\beta\phi}}}{\sum_k \left(\frac{Y_i^k}{A_i^k}\right)^{\frac{1}{\beta\phi}}} X_{si} \qquad \forall \{s, j\} \tag{57}$$

$$X_{si}^{p*} = (1-\beta) X_{si} \qquad \forall \{s\} \tag{58}$$

Substituting the above into the product line production function that is binding (equation 55), and using the homogeneity of $G(\cdot)$:

$$\frac{Y_i^j}{\delta^*} = \frac{A_i^j}{\kappa} \beta^{\beta\phi} \left(\frac{\left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\beta\phi}}}{\sum_k \left(\frac{Y_i^k}{A_i^k}\right)^{\frac{1}{\beta\phi}}} \right)^{\beta\phi} G(\mathbf{X}_i)^{\beta\phi} (1-\beta)^{(1-\beta)} G(\mathbf{X}_i)^{(1-\beta)\phi}$$

$$(\delta^*)^{\frac{1}{\beta\phi}} \frac{\left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\beta\phi}}}{\sum_k \left(\frac{Y_i^k}{A_i^k}\right)^{\frac{1}{\beta\phi}}} = \left(\frac{Y_i^j}{A_i^j G(\mathbf{X}_i)^\phi}\right)^{\frac{1}{\beta\phi}}$$

$$(\delta^*)^{\frac{1}{\beta\phi}} \sum_j \frac{\left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\beta\phi}}}{\sum_k \left(\frac{Y_i^k}{A_i^k}\right)^{\frac{1}{\beta\phi}}} = (\delta^*)^{\frac{1}{\beta\phi}} = \sum_j \left(\frac{Y_i^j}{A_i^j G(\mathbf{X}_i)^\phi}\right)^{\frac{1}{\beta\phi}}$$

$$\implies \delta^* = \left(\sum_j \left(\frac{Y_i^j}{A_i^j F(\mathbf{X}_i)}\right)^{\frac{1}{\beta\phi}} \right)^{\beta\phi} \quad \text{where} \quad F(\mathbf{X}) = G(\mathbf{X})^\phi \tag{59}$$

Further, also consider the following solutions for the Lagrangian multipliers:

$$\lambda_i^j = \left(\sum_j \left(\frac{Y_i^j}{A_i^j F(\mathbf{X}_i)}\right)^{\frac{1}{\beta\phi}} \right)^{2\beta\phi} \frac{\left(\frac{Y_i^j}{A_i^j}\right)^{\frac{1}{\beta\phi}}}{\sum_k \left(\frac{Y_i^k}{A_i^k}\right)^{\frac{1}{\beta\phi}}} \frac{1}{Y_i^j} \qquad \forall \{j\} \tag{60}$$

$$\mu_{si} = \phi \left(\sum_j \left(\frac{Y_i^j}{A_i^j F(\mathbf{X}_i)}\right)^{\frac{1}{\beta\phi}} \right)^{\beta\phi} \frac{G_{(s)}(\mathbf{X}_i)}{G(\mathbf{X}_i)} \qquad \forall \{s\} \tag{61}$$

It is easily verified that the candidate solution $(\delta^*, \mathbf{X}_i^{r*}, \mathbf{X}_i^{p*}, \boldsymbol{\lambda}_i^*, \boldsymbol{\mu}_i^*)$ defined in equations (57) - (61) satisfies the first order conditions stated in equations (52) - (56). Therefore, the output distance function for a firm that shares public intermediates across product lines is given by:

$$D_i(\mathbf{Y}_i, \mathbf{X}_i) = \left(\sum_j \left(\frac{Y_i^j}{A_i^j F(\mathbf{X}_i)} \right)^{\frac{1}{\beta\phi}} \right)^{\beta\phi} \tag{62}$$

where $F(\mathbf{X})$ is homogeneous of degree ϕ. An important takeaway from the above expression is that β governs whether there is joint-production or not.

C Monte Carlo Details

This appendix discusses in more detail the Monte Carlo exercises presented in section 3.2. There are J product-specific CES aggregates C^j each comprised of firm-specific varieties c_i^j as described in the text. Because the ASI data lists a maximum of ten products a firm can list, we set J=10. We also assume that each firm produces a random subset of 5 of these 10 products. Production of other products is exogenously set to zero. For the Monte Carlo exercises, we assume that the output distance function takes the form of equation (13), with $F(\mathbf{X}_i) = L_i$ and therefore $\phi = 1$. Labour, L_i, is paid an exogenous wage w=1 in both uses. β is the cost share of rival (non-public) inputs in production at the firm as in Appendix B. The constant elasticity of substitution demand structure combined with goods market clearing implies that

$$c_i^j = \frac{(\nu_i^j)^{\sigma-1}(P_i^j)^{-\sigma^j}}{\sum_i (P_i^j/\nu_i^j)^{1-\sigma}} \gamma^j E \tag{63}$$

where E is aggregate expenditure, γ^j is a Cobb-Douglas expenditure share $1/J$, and ν_i^j are idiosyncratic demand shifters whose properties are discussed below. The true CES markup is given by

$$P_i^j = \frac{\sigma^j}{\sigma^j - 1} MC_i^j. \tag{64}$$

where

$$MC_i^j = -\frac{\frac{\partial D_i(\mathbf{Y}_i, L_i)}{\partial Y_i^j}}{\frac{\partial D_i(\mathbf{Y}_i, L_i)}{\partial L_i}} w. \tag{65}$$

Using equations (13), (64), and (65), and substantially manipulating, we obtain:

$$P_i^j = \frac{\sigma}{\sigma - 1} \left[\frac{Y_i^j}{L_i} \right]^{\alpha - 1} \frac{w}{(A_i^j)^{\alpha}}. \tag{66}$$

where $\alpha = \frac{1}{\phi\beta}$. If returns to scale are constant and there is no joint production, then $\phi\beta = 1$, and this simplifies to the familiar CES constant markup of price over marginal cost: $P_i^j = \frac{\sigma}{\sigma-1}\frac{w}{A_i^j}$.

Parameters

Lower case letters represent log transformations such that $x = \ln X$. We assume that firm-product productivity can be parameterized by $a_i^j = a_i + \tilde{a}_i^j$. Firms draw a_i from a random normal distribution $N(0, \sigma_F^2)$ where we set $\sigma_F^2 = 5$. $\tilde{a}_i^j$ is drawn from a different random normal distribution $N(0, \sigma_{FJ}^2)$. The ratio σ_{FJ}^2/σ_F^2 is the key parameter that we vary to assess the importance of within-firm productivity heterogeneity. We assume that quality can be written as $\nu_i^j = \exp\ d_i^j$ where $d_i^j = d_i + \tilde{d}_i^j - a_i^j$. d_i and $\tilde{d}_i^j$ are random variables drawn from distributions $N(0, \sigma_D^2)$ and $N(0, \sigma_{DJ}^2)$. We set $\sigma_D^2 = 5$ and $\sigma_{DJ}^2 = 0.5$ in all cases. The final term $(-a_i^j)$ captures the negative correlation of productivity and quality documented in Orr (2022). We assume that the cost of one unit of labor is exogenous and equal to 1, and that aggregate expenditure is also exogenous, $E = 100$. We assume that there are 10,000 firms such that $N = 10,000$. We assume constant returns to scale in all specifications $\phi = 1$. Unless otherwise stated, we assume that $\beta = 1$ such that production is non-joint.

Fixed Point

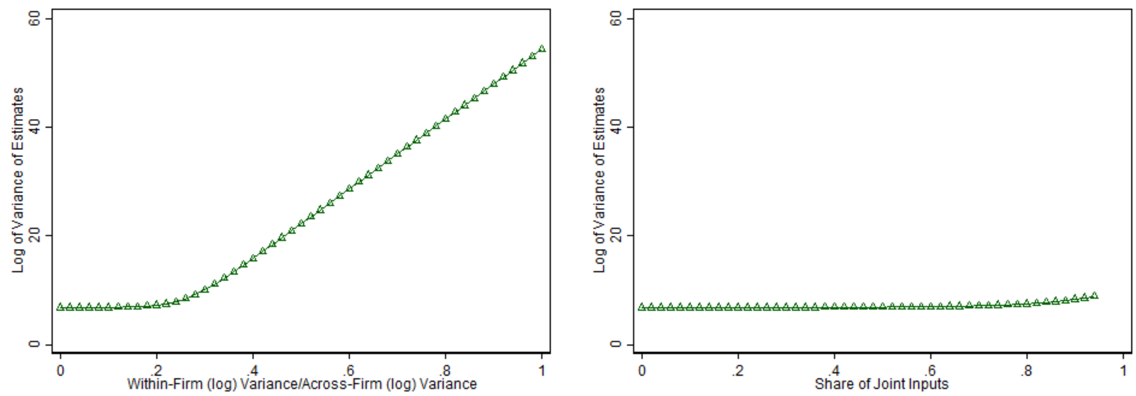
We start by nominating a $N \times J$ matrix of prices P_i^j. Using equation (63) and goods market clearing, we obtain $N \times J$ matrices of revenue and output where typical elements are R_i^j and Y_i^j, respectively. Using equation (13) with $F(\mathbf{X}_i) = L_i$, we obtain a $N \times 1$ vector of total labor demands which we then use to generate a new vector of firm-product prices $\widehat{P}_i^j$ using (66). We then update P_i^j until P_i^j and $\widehat{P}_i^j$ converge. When there is no joint production ($\beta = 1$), this convergence occurs immediately. This delivers values of Y_i^j, R_i^j, and L_i that we combine with exogenous parameters A_i^j and β to calculate true and measured markups using equations (16)-(19) as we discuss below.

Within-Firm Productivity Heterogeneity (Figure 1, panel a)

All parameters are as described above except σ_{FJ} which we initially set to zero imposing no productivity heterogeneity across product lines within a firm. We then calculate the correlation between markups implied by equations (16) and (17). This correlation is 1 which is the first triangle seen in the upper left of panel a. We then increase σ_{FJ} by 0.1, take the relevant draws solve the model again, and calculate the correlation between markups implied

Figure 10: Monte Carlo: Log Variance of Estimates

(A) Within-Firm Productivity Heterogeneity (B) Joint Production



Notes: Each triangle in panel a presents the log of the variance of estimated markups. The horizontal axis shows the ratio of within- to across-firm productivity heterogeneity for that correlation. There is no joint production in this panel. Each triangle is a separate variance term. Each triangle in panel b presents the log of the variance of estimated markups. The horizontal axis shows the cost share of public inputs $1-\beta$. Each triangle is a separate log of variance term. There is no within-firm productivity heterogeneity in this panel. The log of the variance of markups when there is no within-firm productivity heterogeneity and no joint production is 6.78=ln(885.4).

by equations (16) and (17), plotting it as the second triangle in this figure. Seeds are set such that the distribution of across firm draws is the same for each point in the figure. We continue increasing σ_{FJ}, and calculating/plotting the relevant correlation until $\sigma_{FJ}/\sigma_F = 1$.

Joint Production (Figure 1, panel b)

All parameters are as described above except β which we initially set to 1 imposing no joint production. We then calculate the correlation between markups implied by equations (18) and (19). This correlation is 1 which is the first triangle seen in the upper left of panel b. We then decrease β by 0.1, take the relevant draws, solve the model again, and calculate the correlation between markups implied by equations (18) and (19). Again, seeds are set such that draws of the relevant exogenous variables are held constant. We continue decreasing β, and calculating/plotting the relevant correlation until $\beta = 0.2$. At this point, we begin to encounter scaling issues in the shares.

Variance of Estimated Markups

Figure 10 presents the log of the variance of estimated markups analogous to figure 1. The horizontal axis of each panel is exactly the same in each figure. We present the y-axis in logs because of the scale of panel (a). Panel (a) shows that the variance starts to increase dramatically when within-firm productivity heterogeneity is roughly 20 percent of across firm heterogeneity. Panel (b) again shows that the omission of joint production produces less dramatic results although the log of the variance of estimated results increases slightly to 7.41 inputs for joint production are rougly 80 percent of total costs.

D Constrained Social Planner and Decentralized Economy

We will show that the object that the constrained social planner outlined in section 4.3 cares about is the aggregate firm-level markup μ_i. We first re-state the social planners problem here and derive the first order conditions. Then we characterize the decentralised economy and show that firm-level aggregate markups are the object that the social planner wants to equalize across firms.

Constrained Social Planner: A social planner that wants to eliminate across-firm misallocation while holding within-firm output allocations constant faces the following constrained optimization problem:

$$\begin{aligned}
\max_{\{Y_i\}_i} \quad & U\left(\{Y_i \boldsymbol{\lambda}_i\}\right) \\
\text{s.t.} \quad & D_i(Y_i \boldsymbol{\lambda}_i, \mathbf{X}_i) = 1 \qquad \forall\, i \\
& \sum_i X_{si} \leq X_s \qquad \forall\, s
\end{aligned}$$

where $\boldsymbol{\lambda}_i$ is the vector of relative output levels within the firm, and Y_i is *scale* of the firm. Suppose δ_i and γ_s are the lagrangian multipliers associated with firm i's technology constraint and input s's resource constraint respectively, then the first order conditions for the above problem are given by:

$$\text{FOC wrt } Y_i: \qquad \sum_k \frac{\partial U(\{\mathbf{Y}_i\})}{\partial Y_i^k} \lambda_i^k = \delta_i \sum_j \frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial Y_i^k} \lambda_i^k \qquad \forall\, i \tag{67}$$

$$\text{FOC wrt } X_{si}: \qquad \delta_i \frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial X_{si}} = -\gamma_s \qquad \forall\, \{s, i\} \tag{68}$$

Substituting equation 68 into 67, and substituting for $\lambda_i^k = Y_i^k / Y_i^r$ for some reference product r, we can rewrite the above conditions to be:

$$\begin{aligned}
&\sum_{k \in \mathbb{Y}_i} \frac{\partial U(\{\mathbf{Y}_i\})}{\partial Y_i^k} Y_i^k = - \frac{\gamma_s}{\frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial X_{si}}} \sum_{k \in \mathbb{Y}_i} \frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial Y_i^k} Y_i^k \\
&- \frac{\frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial X_{si}} \frac{X_{si}}{D_i(\mathbf{Y}_i, \mathbf{X}_i)}}{\sum_{k \in \mathbb{Y}_i} \frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial Y_i^k} \frac{Y_i^k}{D_i(\mathbf{Y}_i, \mathbf{X}_i)}} \frac{\sum_{k \in \mathbb{Y}_i} \frac{\partial U(\{\mathbf{Y}_i\})}{\partial Y_i^k} Y_i^k}{X_{si}} = \gamma_s \\
&\implies - \frac{\frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln X_{si}}}{\sum_{k \in \mathbb{Y}_i} \frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln Y_{ik}}} \frac{\sum_{k \in \mathbb{Y}_i} \frac{\partial U(\{\mathbf{Y}_i\})}{\partial Y_i^k} Y_i^k}{X_{si}} = \gamma_s
\end{aligned} \tag{69}$$

Therefore the social planner will allocate resources across firms such that the object on the left hand side of equation (69) is equalized across all firms.

Decentralized Economy: Consumers maximize their utility given prices, and firms will minimize costs. We do not require disciplining how firms set prices for deriving markups, but rather use prices we observe in the data as the set of prices that maximize firms' profits. The consumer problem is given by:

$$\begin{aligned}
&\max_{\{Y_i^j\}_{i,j}} U(\{\mathbf{Y}_i\}) \quad \text{s.t.} \quad \sum_i \sum_j P_i^j Y_i^j \leq I \\
&\implies \quad \frac{\partial U(\{\mathbf{Y}_i\})}{\partial Y_i^j} = \theta_u P_i^j
\end{aligned} \tag{70}$$

where θ_u is the Lagrangian multiplier associated with the consumer budget constraint. From the firm's cost minimization problem specified in section 3 and the firm-level aggregate markup defined in section 4, we have:

$$\mu_i = - \frac{\frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln X_{is}}}{\sum_{k \in \mathbb{Y}_i} \frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln Y_i^k}} \frac{\sum_k R_i^k}{E_{si}} = - \frac{\frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln X_{is}}}{\sum_{k \in \mathbb{Y}_i} \frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln Y_i^k}} \frac{\sum_k P_i^k Y_i^k}{w_{si} X_{si}}$$

Substituting for price P_i^k from the consumer FOC, we have:

$$\begin{aligned}
&\mu_i = - \frac{\frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln X_{is}}}{\sum_{k \in \mathbb{Y}_i} \frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln Y_i^k}} \frac{\sum_{k \in \mathbb{Y}_i} \frac{\partial U(\{\mathbf{Y}_i\})}{\partial Y_i^j} Y_i^k}{X_{si}} \frac{1}{\theta_u w_{is}} \\
&\implies \quad - \frac{\frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln X_{is}}}{\sum_{k \in \mathbb{Y}_i} \frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln Y_i^k}} \frac{\sum_{k \in \mathbb{Y}_i} \frac{\partial U(\{\mathbf{Y}_i\})}{\partial Y_i^j} Y_i^k}{X_{si}} = \theta_u w_{is} \mu_i
\end{aligned} \tag{71}$$

The left hand side of equation 71 is exactly equal to the object that the constrained social planner wants to equalize across all firms. Therefore, under the scenario where all firms face the same input costs, $w_{si} = w_s$, the social planner wants to equalize firm-level markups μ_i across all firms in order to eliminate across firm misallocation while holding the within-firm (mis)allocation fixed.

E Proof of Proposition 1

Proof. By Assumption 4, we have

$$D_i\left(\mathbf{Y}_i, \mathbf{X}_i\right) = \frac{G_i(\mathbf{Y}_i)}{F_{g(i)}(\mathbf{X}_i)}, \tag{72}$$

where $G_i(\mathbf{Y}_i)$ is homogenous of degree 1. This implies

$$\sum_{k \in \mathbb{Y}_i} \frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln Y_i^k} = \sum_{k \in \mathbb{Y}_i} \frac{\partial \ln G_i(\mathbf{Y}_i)}{\partial \ln Y_i^k} = \frac{1}{G_i} \sum_{k \in \mathbb{Y}_i} \frac{\partial G_i}{\partial Y_i^k} Y_i^k = 1$$

Substituting this into (21) yields:

$$\mu_i = \frac{\partial \ln D_i\left(\mathbf{Y}_i, \mathbf{X}_i\right)}{\partial M_{is}} \frac{\sum_j R_i^j}{E_{si}} \tag{73}$$

While (72) implies:

$$\mu_i = \frac{\partial \ln F_{g(i)}\left(\mathbf{X}_i\right)}{\partial \ln M_{si}} \frac{\sum_j R_i^j}{E_{si}} = \theta_{sg(i)}\left(\mathbf{X}_i\right) \frac{\sum_j R_i^j}{E_{si}}. \tag{74}$$

The last part of the Theorem then follows from Assumptions 3 and 4, where if $\mathbf{Y}_i = \mathbf{Y}_i^j$, then $D_i\left(\mathbf{Y}_i^j, \mathbf{X}_i\right) = \frac{Y_i^j}{A_i^j F_{g(i)}(\mathbf{X}_i)}$. Since cost minimizing firms will operate on their production possibilities frontier, and therefore $D_i\left(\mathbf{Y}_i^j, \mathbf{X}_i\right) = 1$, this immediately implies $Y_i^j = A_i^j F_{g(i)}\left(\mathbf{X}_i\right)$, and therefore $\theta_{sg(i)}\left(\mathbf{X}_i\right) \equiv \frac{\partial \ln F_{g(i)}(\mathbf{X}_i)}{\partial \ln M_{si}}$ can be interpreted as the output elasticity for static input s for firms that choose to produce a single product.

□

F Firm-level Cost Shares

In this Appendix, we consider an alternative set of restrictions that will allow researchers to identify the firm-markup for any output distance function satisfying constant returns to scale. In a multi-output context, constant returns to scale is defined as follows:

Assumption 9 *Each firm's production possibility frontier involves* ***constant returns to scale****, i.e. for any $\lambda > 0$ and $(\mathbf{Y}_i, \mathbf{X}_i) \in \mathbb{P}_i^f$*

$$D_i(\lambda \mathbf{Y}_i, \lambda \mathbf{X}_i) = 1 \tag{75}$$

Equation (75) essentially states that if we scale inputs by λ from any previously feasible production vector $(\mathbf{Y}_i, \mathbf{X}_i)$, then we can also produce an output vector that is scaled by λ.

To identify the firm-markup for all output distance functions satisfying Assumption 9, we require the following further restriction on the firm's operating environment.

Assumption 10 *All inputs are static, i.e.* $\mathbf{X}_i = \mathbf{M}_i$

Assumption 10 is a very strong restriction, requiring that all inputs, including capital, be available for purchase for some (potentially firm varying) input price W_i^X. A key benefit of this restriction is that with it, we can prove that the firm markup is identified for *any* technology that satisfies constant returns to scale; importantly, this includes technologies with *non-separable* output distance functions. This is important, since non-separable specifications of a firm's technology are needed to capture situations where firms operate non-joint production technologies that have different factor intensities (Hall 1973)— for example, situations where a high quality good may be more capital intensive than a low quality good.

Proposition 3 *If firms minimize total costs and Assumptions 9 and 10 hold, then :*

$$-\frac{\frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln X_{si}}}{\sum_{k \in \mathbb{Y}_i} \frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln Y_i^k}} = \frac{W_{si} X_{si}}{\sum_s W_{si} X_{si}} \tag{76}$$

Proof. Since firms cost minimize, $D_i(\mathbf{Y}_i, \mathbf{X}_i) = 1$. Totally differentiating this expression yields:

$$\sum_{j \in \mathbb{Y}_i} \frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial Y_i^j} dY_i^j + \sum_s \frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial X_{si}} dX_{si} = 0$$

or:

$$\sum_s \frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial X_{si}} dX_{si} = -\sum_{j \in \mathbb{Y}_i} \frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial Y_i^j} dY_i^j$$

Divide by D_i and multiply the left hand side by $\frac{X_{si}}{X_{si}}$

$$\sum_s \frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial X_{si}} \frac{X_{si}}{D_i} \frac{dX_{si}}{X_{si}} = -\sum_{j \in \mathbb{Y}_i} \frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial Y_i^j} \frac{Y_i^j}{D_i} \frac{dY_i^j}{Y_i^j}$$

$$\sum_s \frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln X_{si}} \frac{dX_{si}}{X_{si}} = -\sum_{j \in \mathbb{Y}_i} \frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln Y_i^j} \frac{dY_i^j}{Y_i^j}$$

Assumption 9 implies that if $\frac{dX}{X} = 1 + \delta \; \forall X$, then $\frac{dY_i^j}{Y_i^j} = 1 + \delta \; \forall j$, which implies:

$$\sum_s \frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln X_{si}} = -\sum_{j \in \mathbb{Y}_i} \frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln Y_i^j} \tag{77}$$

Next, since firms cost minimize and all inputs are static by Assumption 10, the static cost minimization condition (8) holds for all inputs s, which implies:

$$W_{is} = -\lambda_i \frac{\partial D_i}{\partial X_{si}} \rightarrow \frac{W_{si} X_{si}}{D_i} = -\lambda_i \frac{\partial D_i}{\partial X_{si}} \frac{X_{si}}{D_i}$$

Since $D_i = 1$

$$W_{si} X_{si} = -\lambda_i \frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln X_{si}} \tag{78}$$

Sum over all s

$$\sum_s W_{si} X_{si} = -\lambda_i \sum_s \frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln X_{si}} \tag{79}$$

Divide (78) by (79), yielding:

$$\frac{W_{si} X_{si}}{\sum_s W_{si} X_{si}} = \frac{\frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln X_{si}}}{\sum_s \frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln X_{si}}}$$

Substituting using (77) to substitute out $\sum_s \frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln X_{si}}$ in this expression then yields:

$$\frac{W_{si} X_{si}}{\sum_s W_{si} X_{si}} = -\frac{\frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln X_{si}}}{\sum_{k \in \mathbb{Y}_i} \frac{\partial \ln D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial \ln Y_i^k}} \tag{80}$$

□

Proposition 3 tells us that the series of output distance function elasticities that are necessary to recover the firm markup in (21) can be point identified, firm-by-firm, by simply examining the cost shares of one of the inputs. This is a powerful result, as the *only* restriction on the shape of the output distance functions is that it be constant returns to scale, which incorporates an extremely broad class of joint and non-joint technologies, including non-joint technologies with different product-line production functions. However, it requires the strong restriction that all inputs be static, which may not be appropriate in many settings. Since the requirement that all inputs be static means that the static first-order conditions hold for all inputs, Proposition 3 also implies to following corrolary:

Corollary 1 *If firms minimize total costs and Assumptions 9 and 10 hold, then:*

$$\mu_i = \frac{\sum_j R_i^j}{\sum_s W_{si} X_{si}} \tag{81}$$

Proof. Implied by substituting (76) into (21) and cancelling out $\frac{W_{si}X_{si}}{E_{si}} = 1$ □

This result is reminiscent of a well known result that the ratio of revenues to costs measures the markup when there are constant returns to scale and all inputs are static (e.g. De Loecker et al. 2020). However, a key distinction is that previous variants of this result focused on settings where revenue and cost data were interpreted as if they applied to a single product. Here, we allow firms to produce arbitrary numbers of products, and allow firms to operate using an extremely wide class of joint and non-joint production technologies. In this more general setting, the "markup" cannot always be interpreted as the price over marginal cost of a single product (this is only true if a firm produces a single product), but rather a weighted average of a series of production level markups, $\mu_i = \sum_j \rho_i^j \mu_i^j$.[58]

G Proof of Proposition 2

Proof. Since $D_i(\mathbf{Y}_i, \mathbf{X}_i) = 1$, then

$$\frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial X_{si}} X_{si} = \frac{\partial D_i(\mathbf{Y}_i, \mathbf{X}_i)}{\partial X_{si}} \frac{X_{si}}{D_i(\mathbf{Y}_i, \mathbf{X}_i)} = -\beta_{s,g(i)},$$

where the last equality follows from Assumption 6. Combining this with Assumption 5 then implies

$$\mathbb{E}\left(W_{si}X_{si} | g = g(i)\right) = \beta_{s,g(i)} \mathbb{E}\left(\lambda_i | g = g(i)\right).$$

Since this holds for all inputs:

$$\mathbb{E}\left(\sum_s W_{si}X_{si} | g = g(i)\right) = \sum_s \beta_{s,g(i)} \mathbb{E}\left(\lambda_i | g = g(i)\right) = \mathbb{E}\left(\lambda_i | g = g(i)\right).$$

[58] De Loecker and Warzynski (2012) emphasize a similar result when addressing the issue of multiple products shipped to different markets. Specifically, they show that the firm level markup derived in their setting is an input share weighted average of product-level markups. However, their results apply only to non-joint production settings with product line production function.

Takign the ratio of the previous two expression then yields:

$$\frac{\mathbb{E}\left(W_{si}M_{si}|g=g(i)\right)}{\mathbb{E}\left(\sum_{s'}W_{s'i}X_{s'i}|g=g(i)\right)}=\beta_{s,g(i)}$$

Since s is assumed to be a static input, (21) holds for this input, and the result immediately follows. □

H Data Appendix

- Labour (L_{it}) : *Labour input*: We use labour input in person-days, as recorded in ASI Block E. We use the total labour input for all classes of employees, including managers. *Wage bill, wL*: The wage bill is the total expenditure on wages for all classes of employees. Dividing the wage bill by labour input gives the implied wage, w.

- Capital (K_{it}) : *Stock of Capital*: We rely on the perpetual inventory method, use the heterogeneous depreciation rates of 0%, 5%, 10%, 20%, and 40% for land, buildings, machinery, transportation equipment, and computers & software, respectively, when constructing the capital stock over time as in Boehm and Oberfield (2018). The starting value for each component of a plant's capital stock is given by its book net opening value as recorded in block C of the ASI, and investment is taken as the value of gross additions plus gross revaluations minus gross deductions. The perpetual inventory method is only used to construct the capital stock for plants operating in multiple sequential years. We use the book net opening value as an estimate of the capital stock for all plants that are missing past values. *Value of Capital, rK*: For the cost share approach, we need to calculate the total user cost of capital for each firm. We largely follow Raval (2022) for this purpose. In particular, we obtain the real interest rate by year as $R_t=\frac{i_t-\pi_t}{1+\pi_t}$, where i_t is the private sector lending rate and π_t is the inflation rate.[59] The rental rate of capital is then defined as the sum of the real interest rate plus depreciation for each type of capital (buildings, machinery, transportation equipment, and computers software). Then value of the capital stock is then the rental rates times the value of the capital stock, summed over each type of capital.

- Materials (M_{it}) : Domestic and directly imported inputs are recorded in ASI Blocks H and I, respectively. For domestic inputs, we keep the total value of intermediates used,

[59] We obtain the private sector lending rate from IMF Financial Statistics, and the inflation rate for India from World Development Indicators.

including electricity. For directly-imported intermediates we also keep the total, and impute zero imports for firms that do not fill out the block. Total materials expenditure is the sum of expenditure on domestic and foreign intermediates.

I Baseline Estimation Details

In this Appendix, we outline in detail the construction of our control function estimator for the production function using single product firms. In particular, we describe how we deal with input price bias, selection into single product firm status, and finally how we construct out estimating equation based on innovations to productivity.

I.1 Input prices

Although we observe output quantities in the ASI, we do not observe input prices for all inputs. As De Loecker and Goldberg (2014) explain, using deflated input expenditures as a proxy for physical inputs can bias elasticity estimates when physical output is observed. De Loecker et al. (2016) deal with unobserved input prices using a control function approach. Modeling input prices as a function of output prices, market shares, and other state variables, they estimate the parameters of this function simultaneously with translog parameters. Recovering input price control function parameters allows them to predict input "prices," which are used to deflate expenditures and construct output elasticities.

We exploit a unique feature of the ASI to correct for input price bias during the production function estimation routine, without having to use a control function. The ASI includes data on both labor expenditure and physical input in man-days, which allows us to recover labor unit values (wages). Following De Loecker et al. (2016) in assuming common deviations from the national price index for each of capital, labor and intermediates, we use these unit values to deflate plant expenditures for capital and intermediates, in addition to using labor input in man-days.[60] More formally, we assume that prices for input s can be decomposed into time- and time-plant varying components, or

$$W_{sit} = W_{st} \times W_{it}. \tag{82}$$

Recall that we deflated nominal expenditures using the Indian WPI, which is time-varying only. By subsequently deflating using the observed wage, we capture components of input

[60] De Loecker et al. (2016) require that input "prices" be the same across inputs within each product line in order to identify input price control function parameters.

prices that vary cross-sectionally. We refer to the plant-level wage as an input "price," it is properly interpreted as the plant's idiosyncratic input price deviation from the WPI.

Since we only observe plant-level wages, not plant-product wages, this approach does not allow for plant-product input prices as in De Loecker et al. (2016). Since we only use single-product firms for estimation, this is without a loss of generality at the estimation stage. However, this distinction matters once we turn to identifying markups in multi-product producers. In particular, firm-product input price control functions can incorporate information on unobserved quality differences by product line. This is a key piece of identifying information that can, and should, be incorporated in the algorithm for identifying the unobserved input allocations. As a result, when using their approach to uncover input allocations, we follow De Loecker et al. (2016) and allow for firm-product input prices. In practice, we do this by directly estimating a firm-product input price controls function using *observed* plant-level wages for single-product producers. See Section K below for details.

On the other hand, when we consider our measure of firm-level markups, we do not allow for firm-product specific input prices. The reason for this is that our approach allows for joint production, i.e. situations where inputs are not directly allocated to particular product lines, but rather enter the production of all goods simultaneously. In these cases, a firm-product input price does not have a clear meaning. Therefore, to truly have our approach nest both joint and non-joint production, we do not allow for within-firm input price variation.

I.2 Selection correction

A concern with using single-product plants is that this may introduce selection bias due to the non-random addition of product lines. As De Loecker et al. (2016) note, plants may choose to add product lines if their realized productivity exceeds an idiosyncratic threshold. Moreover, this threshold may be negatively correlated with past levels of dynamic inputs (for example, floor space in a factory) because firms with large stocks of these inputs at $t-1$ would rather add a product at t, even if their productivity is low, rather than leave inputs underutilized. This may induce a negative correlation between the firm's time-t productivity shock and levels of dynamic inputs at t, which are positively correlated with levels at $t-1$. This violates the exclusion restrictions described below.

We correct for potential selection bias of this form using a similar procedure to De Loecker et al. (2016). An important component of this approach is that selection into multi-product firm status at time t be based on the productivity of a single-product firm through a simple cutoff rule based on their productivity at time $t-1$. For this purpose, we assume that the decision to add a product at time t is made at time $t-1$. We then require that this decision be based on a simple threshold productivity rule, where, letting J_{it} denote the number of products by i at time t:[61]

$$J_{it} > 1 \quad \text{if} \quad a_{i,t-1} > \overline{a}_{i,t-1} = \overline{a}_{t-1}(l_{i,t-1}, k_{i,t-1}, i_{i,t-1}, \Gamma_{i,t-1}), \quad \text{otherwise} \quad J_{it} = 1 \tag{83}$$

where $i_{i,t-1}$ is investment, $a_{i,t}$ is log TFPQ for the single product produced by firm i, and $\Gamma_{i,t-1}$ is a vector of firm-level state variables at time $t-1$, excluding dynamic inputs and a_{it}.[62]

If this is the case, we can then write:

$$\begin{aligned} Pr(J_{it} = 1 | J_{i,t-1} = 1) &= Pr\left(a_{i,t-1} \leq \overline{a}_{i,t-1}\right) \\ &= Pr\left(a_{i,t-1} \leq \overline{a}_{t-1}(l_{i,t-1}, k_{i,t-1}, i_{i,t-1}, \Gamma_{i,t-1}\right) \\ &= \kappa_{t-1}\left(l_{i,t-1}, k_{i,t-1}, i_{i,t-1}, \Gamma_{i,t-1}\right) \\ &\equiv SP_{it} \end{aligned} \tag{84}$$

Importantly, this expression can be estimated using a standard binary outcome model (e.g. a probit) on a dummy for whether a firm changes their status from single product to multi-product in the subsequent period. This allows us to then obtain probabilities that a firm exits the sample, which can then be used to correct for selection. We provide the details of this procedure I.3 below.

An important consideration is whether the more general setting of MRT heterogeneity and joint production considered in this paper is likely to generate the required threshold selection rules as in (83). In particular, De Loecker et al. (2016) appeal to the model in Mayer et al. (2014), which is explicitly non-joint. In Appendix Q, we consider a particular parametric example that extends the Mayer et al. (2014) environment to our setting with joint production, and show that this model generates exactly this class of selection rule. Note, however, that this particular parametric structure is not strictly necessary; rather, any model that generates a selection rule as in (83) will be valid.[63]

[61] De Loecker et al. (2016) consider a slightly different formulation of the this problem, based on firm's productivity at time t, rather than $t-1$. Note however, that since the decision to add products is made at time $t-1$, it is natural to state the selection rule in terms of realized productivity at time $t-1$ (when the firm makes their decision to add a product), rather than time t. Note, however, that both approaches to the problem lead to basically identical selection corrections.

[62] We set $\Gamma_{it} \equiv (w_{it}, p_{it}, ms_{it}, \mathbf{D}_{it})$ where w_{it} is the firm wage, p_{it} is a log output price, ms_{it} is the firm's quantity market share for product j, $\mathbf{D}_{it}$ is a vector of year and 5-digit ASICC product code dummies for the product produced by the single product firm.

[63] An important condition for this type of selection rule to be valid is that the degree of joint-production not vary across firms; otherwise, the selection cutoff will depend on other unobservables that would need to be estimated.

I.3 Proxy-variable estimator

We rely on a proxy-variables approach as in De Loecker et al. (2016), where we assume that the physical materials demand function can be written as

$$m_{it} = m_{g(i)}\left(a_{it}, k_{it}, l_{it}, \Gamma_{it}\right), \tag{85}$$

Under the assumption that this function is monotonic in a_{it} for single product plants, we can invert this expression, yielding a_{it}^{j} as a function of observables:

$$a_{it} = h_{g(i)}\left(m_{it}, k_{it}, l_{it}, \Gamma_{it}\right) = h_{g(i)}\left(\mathbf{x}_{it}, \Gamma_{it}\right). \tag{86}$$

Estimation relies of a two step procedure as in De Loecker et al. (2016), In the first step we estimate the following model for all single product firms:

$$y_{it} = \ln\left[F_{g(i)}(\mathbf{X}_{it};\beta)\right] + a_{it} + \epsilon_{it} = \ln\left[F_{g(i)}(\mathbf{X}_{it};\beta)\right] + h_{g(i)}\left(\mathbf{X}_{it}, \Gamma_{it}\right) + \epsilon_{it} = \Phi_{g(i)}\left[\mathbf{X}_{it}, \Gamma_{it}\right] + \epsilon_{it} \tag{87}$$

where $F_{g(i)}(\mathbf{X}_{it};\beta)$ has the translog form in (29), ϵ_{it} is an error term capturing unanticipated shocks when inputs are chosen at time t, or measurement error in y_{it}, and $\Phi_{g(i)}\left[\mathbf{X}_{it}, \Gamma_{it}\right] \equiv \ln\left[F_{g(i)}(\mathbf{X}_{it})\right] + h_{g(i)}\left(\mathbf{X}_{it}, \Gamma_{it}\right)$

We use (87) to estimate $\Phi_{g(i)}\left[\mathbf{X}_{it}, \Gamma_{it}\right]$, which allows us to then construct an estimate of $a_{i,t}$ for single product firms as $\widehat{a}_{i,t} = \widehat{\Phi}_{g(i)}\left[\mathbf{X}_{it}, \Gamma_{it}\right] - \ln\left[F_{g(i)}(\mathbf{X}_{it};\boldsymbol{\beta})\right]$.[64]

The production function parameters are then estimated in the second step. We assume that productivity follows a Markov process, so $\mathbb{E}\left(a_{i,t}|\mathbb{I}_{i,t-1}\right) = \mathbb{E}\left(a_{i,t}|a_{i,t-1}\right)$, where $\mathbb{I}_{i,t-1}$ is firm i's information set at time t. Since in this step we need to condition of firms that were single product in both time periods, and we assume that selection in multi-product firm status, conditional on being single product at time $t-1$, follows selection rule (83) and probability (84), we work with the following moment:

$$\begin{aligned}
a_{i,t} &= \mathbb{E}\left(a_{i,t}|a_{i,t-1}, a_{i,t-1} \leq \overline{a}_{i,t-1}\right) + \xi_{it} \\
a_{i,t} &= \mathbb{E}\left(a_{i,t}|a_{i,t-1}, a_{i,t-1} \leq f_{g(i)}(SP_{it})\right) + \xi_{it} \\
a_{i,t} &= m_{g(i)}\left(a_{i,t-1}, SP_{it};\boldsymbol{\gamma}\right) + \xi_{it}
\end{aligned} \tag{88}$$

where $\boldsymbol{\gamma}$ denotes a vector of parameters governing the shape of the unknown function $m_{g(i)}(,)$, and ξ_{it} is a productivity shocks unknown to firms at time $t-1$.[65] Note that the second

[64] In practice, we approximate the unknown function with a second order polynomial in all continuous variables, plus the various fixed effects.

[65] We approximate $g(.)$ with a second order polynomial in all it's arguments.

equality uses the fact that according to (84), $SP_{it} = Pr\left(a_{i,t-1} \leq \overline{a}_{i,t-1}\right)$, and the fact that $Pr\left(a_{i,t-1} \leq \overline{a}_{i,t-1}\right)$ will be a monotonic function of $\overline{a}_{i,t-1}$, and therefore we can invert this expression to obtain $\overline{a}_{i,t-1} = f_{g(i)}(SP_{it})$.

We simultaneously estimate $\boldsymbol{\gamma}$, and the production function parameters $\boldsymbol{\beta}$, by substituting $\widehat{a}_{i,t} = \widehat{\Phi}_{g(i)}\left[\mathbf{X}_{it}, \Gamma_{it}\right] - \ln\left[F_{g(i)}(\mathbf{X}_{it};\beta)\right]$ into (88), which yields the following expression for ξ_{it} as a function of unknown parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma})$:

$$\xi_{it}(\boldsymbol{\beta},\boldsymbol{\gamma}) = \hat{\Phi}_{it} - \ln[F_{g(i)}(\mathbf{X}_{it};\boldsymbol{\beta})] - m_{g(i)}\left(\hat{\Phi}_{i,t-1} - \ln[F_{g(i)}(\mathbf{X}_{i,t-1};\boldsymbol{\beta})], \widehat{SP}_{it};\boldsymbol{\gamma}\right) \quad (89)$$

where $\widehat{SP}_{it}$ is the predicted values obtained by estimating a binary choice model for remaining single product based on (84). In practice, we do this by estimating the following probit model

$$DSP_{it} = \kappa_{t-1}\left(l_{i,t-1}, k_{i,t-1}, i_{i,t-1}, \Gamma_{i,t-1}\right) \quad (90)$$

where DSP_{it} is a dummy for whether a firm remains single product at time t, and $\kappa_{t-1}(.)$ is approximated with a second degree polynomial in all continuous variables.

We rely on equation (89) to construct sample equivalents of moments based on

$$\mathbb{E}[\xi_{it} \times \mathbf{Z}_{it}] \quad (91)$$

where $\mathbf{Z}_{it}$ is a vector of instruments.[66] We instrument labour and materials with their lagged values, and to deal with measurement error in k_{it}, we instrument k_{it} with lagged investment, as suggested by Collard-Wexler and De Loecker (2015). Together $(l_{i,t-1}, m_{i,t-1}, i_{i,t-1})$ identify the linear terms in (29), while we generate further instruments for identifying the nonlinear translog terms of the form $x_1x_2 - 0.5\left((x_1)^2 + (x_2)^2\right)$ for each $(x_1, x_2) \in (l, m, k)$ by replacing x_s with the instrument chosen for that particular input.[67]

[66] In practice, we concentrate out $\widehat{\gamma}$ in an "inner-loop", given $\widehat{\boldsymbol{\beta}}$, as in Levinsohn and Petrin (2003), where conditional on a guess $\widehat{\boldsymbol{\beta}}$, we estimate $m_{g(i)}()$ using OLS, and then use the error terms for from this to construct the GMM criterion function. To account for differences in units across product codes, we demean all variables within 5-digit ASICC code when constructing the GMM criterion function.

[67] So, for example, we generate an instrument for $l_{it}m_{it} - 0.5\left((l_{it})^2 + (m_{it})^2\right)$ using $l_{i,t-1}m_{i,t-1} - 0.5\left((l_{i,t-1})^2 + (m_{i,t-1})^2\right)$, and an instrument for $k_{it}m_{it} - 0.5\left((k_{it})^2 + (m_{it})^2\right)$ using $i_{i,t-1}m_{i,t-1} - 0.5\left((i_{i,t-1})^2 + (m_{i,t-1})^2\right)$.

J Linear dynamic panel estimator

As an alternative to the proxy variable approach with a selection-correction for single product plants, we also apply a linear dynamic panel style estimator to the sample of single-product firms. While this approach does not deal with selection into single-product plant status, it does address some concerns articulated in Bond et al. (2021) concerning internal consistency of proxy-variable methods when there are variable markups.

The estimation approach relies on the assumption that the productivity of single product plant follows a linear AR(1) process:

$$a_{it} = a_0 + \rho a_{i,t-1} + \xi_{it} \tag{92}$$

We can use the above to ρ-difference $y_{it} = \ln[F_{g(i)}(\mathbf{X}_{it}; \boldsymbol{\beta})] + a_{it}$ yielding:

$$y_{it} = \rho y_{i,t-1} + \ln[F_{g(i)}(\mathbf{X}_{it}; \boldsymbol{\beta})] - \rho \ln[F_{g(i)}(\mathbf{X}_{it}; \boldsymbol{\beta})] + \xi_{it} \tag{93}$$

We estimate the above model using a non-linear GMM routine, using the same set of instruments as used in the control function estimator, plus $y_{i,t-1}$ to identify ρ. To account for differences in units across product codes, we demean all variables within 5-digit ASICC code when estimating this expression.

K Recovering Shares in Closed Form

In our baseline procedure for recovering markups by product line, we follow De Loecker et al. (2016) as allow for product-line specific input prices, to potentially account for quality differences across products. First, we assume as in De Loecker et al. (2016) that plant-product input prices can be decomposed into time-varying and time- and plant-product varying components, or

$$\widehat{W}_{sit}^{j} = W_{st} \times \widehat{W}_{it}^{j}. \tag{94}$$

where W_{st} is simply an industry-level deflator for input s. We then assuming that the cross-section varying component can be modeled as a function of firm state variables, including output prices, market shares, vectors of state and product dummies, and a trade status indicator, or

$$\ln W_{it}^{j} = w(p_{it}^{j}, ms_{it}^{j}, IMP_{it}, \mathbf{D}_{it}, \mathbf{G}_{it}). \tag{95}$$

Since we actually observe input prices (plant wages) in the data, we can use single-product plants to predict input prices for multi-product plants. In particular, we estimate the model

$$\ln W_{it}^{j} = \boldsymbol{\delta} \cdot \mathbf{C}_{it}^{j} + \epsilon_{it} \tag{96}$$

on the sample of single-product plants, where $\mathbf{C}_{it}^{j}$ contains a low-order polynomial in log output prices and market shares, plus product, state, import status, and year dummies. Indicators are permitted to enter log-additively, identically to De Loecker et al. $\boldsymbol{\delta}$ is a vector of parameters. We estimate (96) by OLS, and use the fitted values $\widehat{W}_{it}^{j}$ as product-level deviations from the price index, for all firms.

With the predicted product-line specific input prices in hand, we can then derive our closed for expression for input shares. We start with generalization of (32) that allows for MRT heterogeneity, and then obtain our baseline by imposing $A_{it}^{j} = A_{it}$.

Since $F(.)$ is homogenous of degree $\phi > 0$, we can write:

$$Y_{it}^{j} = A_{it}^{j} F\left(\rho_{it}^{j} \frac{\mathbf{E}_{it}}{\widehat{W}_{it}^{j}}\right) = A_{it}^{j} \left(\frac{\rho_{it}^{j}}{\widehat{W}_{it}^{j}}\right)^{\phi} F\left(\mathbf{E}_{it}\right) \tag{97}$$

Which implies

$$\left(\frac{Y_{it}^{j}}{A_{it}^{j}}\right)^{\frac{1}{\phi}} \widehat{W}_{it}^{j} = \rho_{it}^{j} \left(F\left(\mathbf{E}_{it}\right)\right)^{\frac{1}{\phi}} \tag{98}$$

Sum over j

$$\sum_{j} \left(\frac{Y_{it}^{j}}{A_{it}^{j}}\right)^{\frac{1}{\phi}} \widehat{W}_{it}^{j} = \left(F\left(\mathbf{E}_{it}\right)\right)^{\frac{1}{\phi}} \tag{99}$$

where we use the fact that $\sum_{j} \rho_{it}^{j} = 1$. Divide (100) by (99):

$$\frac{\left(\frac{Y_{it}^{j}}{A_{it}^{j}}\right)^{\frac{1}{\phi}} \widehat{W}_{it}^{j}}{\sum_{k} \left(\frac{Y_{it}^{k}}{A_{it}^{k}}\right)^{\frac{1}{\phi}} \widehat{W}_{it}^{k}} = \rho_{it}^{j} \tag{100}$$

Equation (33) in the main text is then obtained by imposing $A_{it}^{j} = A_{it}$.

L Alternative Share Procedures for Figure 3

In this Appendix, we describe in detail three alternative to our baseline input allocation rule that we report in Figure 3 of the main text.

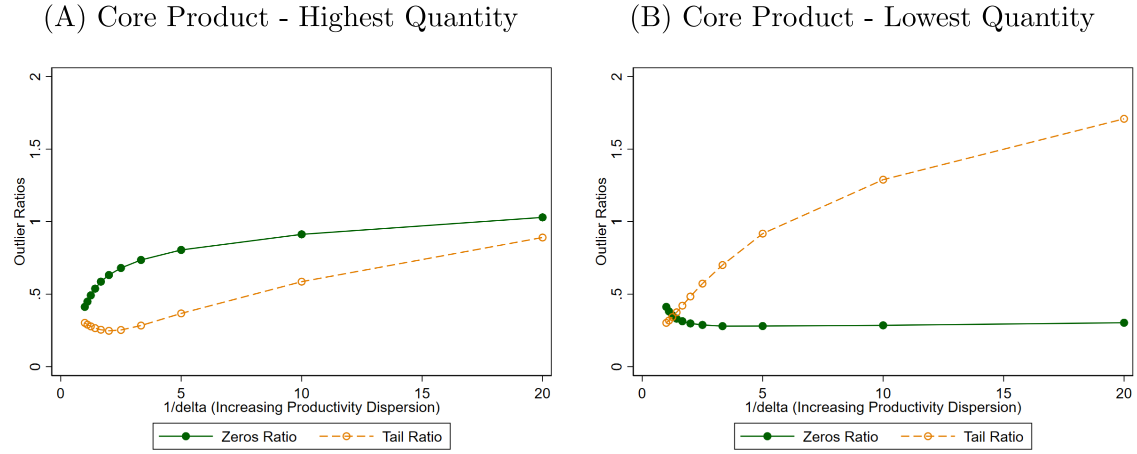
1: We consider modifying the input allocation rule by dropping the input price control function, $\widehat{W_{it}^j}$ from the problem, generating the modified input share rule $\rho_{it}^j = \frac{\left(Y_{it}^j\right)^{\frac{1}{\phi}}}{\sum_k \left(Y_{it}^k\right)^{\frac{1}{\phi}}}$ ("No Wj"). This case corresponds to the standard setting considered in this paper where inputs are the same when used across various product lines. We find that dropping the input-price control function does little to change the proportion of right tail outliers, while it slightly increases the proportion of zero outliers.

2: We replace realized output in (33) with "predicted" output, as suggested by De Loecker et al. (2016) to adjust for measurement error. More precisely, in this specification ("Phi"), we re-rerun the first step of our control function estimation routine for the *full* sample of single- and multi-product plants. We then use this regression to generated predicted output $\widehat{\Phi}_{it}^j$, which we use in place of Y_{it}^j in (33)— this specification appears to increase both the proportion of tail, as well as zero, outliers.

3: We consider whether the $A_{it}^j = A_{it}$ restriction is generating issues due to differences in quantity units across product lines. In particular, A_{it}^j is a quantity-based TFP measure, and therefore will partly vary with the units product j is measured in. To correct for this, we consider a units adjusted input allocation rule $\frac{\left(\frac{Y_{it}^j}{\overline{A}^j}\right)^{\frac{1}{\phi}} \widehat{W}_{it}^j}{\sum_k \left(\frac{Y_{it}^k}{\overline{A}^k}\right)^{\frac{1}{\phi}} \widehat{W}_{it}^k}$, ("Units Adjustment"), where $\overline{A}^j$ is average TFPQ for single-product plants producing product code j. Unfortunately, we find that this approach does little to alleviate the zero markup puzzle.

M Product Markups with Quantity Ranked TFPQ

Figure 11: Product-level Markup Outlier Ratios with Productivity Ladders



Notes: The above reports outlier ratios for different models of product-level markups. Letting $\mathbb{Z}$ denote the set of product-level markups where $\mu \leq 0.5$, $\mathbb{T}$ denote the set of product-level markups where $\mu \geq 5$, and $\mathbb{R}$ denote product-level markups between 0.5 and 5, the **Zeros** outlier ratio (represented by circles) is defined as $\frac{|\mathbb{Z}|}{|\mathbb{R}|}$, while the **Tail** outlier ratio (represented by triangles) is defined as $\frac{|\mathbb{T}|}{|\mathbb{R}|}$, where $|.|$ denotes the number of elements in a given set. These statistics are report as a function of within firm productivity dispersion, parameterized by $\frac{1}{\delta}$. Product-level markups calculated using equation (34) with $\rho^j_{it} = \left[\left(\frac{Y^j_{it}}{(\delta)^j} \right)^{\frac{1}{\phi}} \widehat{W}^j_{it} \right] / \sum_k \left(\frac{Y^k_{it}}{(\delta)^k} \right)^{\frac{1}{\phi}} \widehat{W}^k_{it} \Bigg]$. Panel (A) orders products j within a plant from highest to lowest quantity sold, with $j = 0$ having the highest quantity sold, $j = 1$ having the second highest quantity, and so on. Panel (B) orders products within a plant from lowest to highest quantity, with $j = 0$ denoting the lowest quantity product, $j = 1$ denoting the second lowest quantity product, and so on.

N Production Function Estimates

In this Appendix, we report output elasticities and overall returns to scale for the various approaches used to estimate plant and plant-product markups. The approaches considered in the paper are:

1. **Baseline:** We rely on a translog specification of the technology as in equation (29), and use the control function estimation approach outlined in Appendix I. In particular, we only use single product plants for estimation and correct for this with a De Loecker et al. (2016)-style selection correction, and also correct for input price bias using plant-level wages. We include lagged investment as an instrument for capital in this specification, as recomended by Collard-Wexler and De Loecker (2021). We only estimate a single set

of production function parameters for the entire Indian economy in this specification; however, this still allows for heterogeneity in output elasticities across firms due to the translog form. Estimates reported in Table 2.

2. **Linear dynamic panel:** We rely on a translog specification of the technology as in equation (29), and use the dynamic panel approach outlined in Appendix J. We only use single product plants for estimation but *do not* correct for this with a De Loecker et al. (2016)-style selection correction. However, we *do* correct for input price bias using plant-level wages. We include lagged investment as an instrument for capital in this specification, as recomended by Collard-Wexler and De Loecker (2021). We only estimate a single set of production function parameters for the entire Indian economy in this specification. Estimates reported in Table 3.

3. **Baseline by industry:** Same as **Baseline**, except we estimate production function parameters by 2-digit NIC industry. Estimates reported in Table 4.

4. **Linear dynamic panel by industry:** Same as **Linear dynamic panel**, except we estimate production function parameters by 2-digit NIC industry. Estimates reported in Table 5.

5. **Baseline alternative instruments:** Same as **Baseline**, except we use k_{it} as an instrument, rather than lagged investment. Estimates reported in Table 6.

6. **Baseline Cost Share:** Cost-share approach described in the main text, where production function parameters only differ across 2-digit NIC industries. Estimates reported in Table 7.

7. **Heterogeneous Cost Share:** Cost-share approach described in the main text, where production function parameters differ across 2-digit NIC industries × labour-to-material quintiles, as suggested by Raval (2023). Estimates reported in Table 8.

For each specification, we report the mean and standard deviation for each output elasticity, as well as the mean and standard deviation of returns to scale.[68] In the first column, we report these statistics for all available industries. In the subsequent columns, we report statistics by industry, for a subset of India's largest industries. The output elasticities are calculated over single-product plants when we rely on approaches that only use single-product plants for estimation (Tables 2 through 6). In the cost-share approaches (Tables 7 and 8, we rely on all plant-years, including multi-product plants.

[68]Sometimes the standard deviation is zero, if that particular specification holds the statistic constant within a particular industry-group.

Table 2: Output Elasticities: Baseline estimation using single product plants

	All	15	17	18	21	24	25	26	27	28	29	31
θ_k	0.187	0.186	0.188	0.189	0.188	0.184	0.188	0.194	0.179	0.185	0.187	0.183
	(0.020)	(0.021)	(0.020)	(0.019)	(0.015)	(0.021)	(0.016)	(0.017)	(0.018)	(0.019)	(0.018)	(0.020)
θ_l	0.057	0.049	0.055	0.064	0.051	0.054	0.050	0.071	0.040	0.056	0.060	0.051
	(0.022)	(0.020)	(0.021)	(0.020)	(0.014)	(0.020)	(0.017)	(0.018)	(0.021)	(0.019)	(0.019)	(0.020)
θ_m	0.900	0.909	0.901	0.890	0.905	0.905	0.906	0.879	0.925	0.902	0.897	0.909
	(0.037)	(0.037)	(0.037)	(0.036)	(0.026)	(0.037)	(0.030)	(0.030)	(0.035)	(0.035)	(0.034)	(0.037)
RTS	1.144	1.144	1.144	1.144	1.144	1.144	1.144	1.144	1.144	1.144	1.144	1.144
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)

Notes: The above table reports the estimated average output elasticities and their standard deviation across single-product plants. Output elasticities are based on a single translog estimated for the entire Indian economy. The first column reports the statistics for all sectors, while the remaining columns report statistics for India's largest manufacturing sectors (organized by 2-digit NIC code). The bottom row reports estimated returns to scale.

Table 3: Output Elasticities: Linear dynamic panel estimation using single product plants

	All	15	17	18	21	24	25	26	27	28	29	31
θ_k	0.219	0.223	0.224	0.219	0.226	0.213	0.227	0.228	0.210	0.215	0.218	0.211
	(0.051)	(0.052)	(0.051)	(0.046)	(0.039)	(0.057)	(0.041)	(0.048)	(0.045)	(0.048)	(0.046)	(0.049)
θ_l	0.030	0.012	0.025	0.046	0.017	0.028	0.015	0.057	0-.001	0.030	0.037	0.022
	(0.042)	(0.035)	(0.040)	(0.035)	(0.028)	(0.041)	(0.033)	(0.040)	(0.039)	(0.035)	(0.034)	(0.036)
θ_m	0.860	0.873	0.860	0.844	0.866	0.869	0.868	0.825	0.901	0.864	0.855	0.876
	(0.065)	(0.065)	(0.064)	(0.062)	(0.046)	(0.064)	(0.053)	(0.052)	(0.061)	(0.061)	(0.060)	(0.065)
RTS	1.109	1.109	1.109	1.109	1.109	1.109	1.109	1.109	1.109	1.109	1.109	1.109
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)

Notes: The above table reports the estimated average output elasticities and their standard deviation across single-product plants. Output elasticities are based on a single translog estimated for the entire Indian economy. The first column reports the statistics for all sectors, while the remaining columns report statistics for India's largest manufacturing sectors (organized by 2-digit NIC code). The bottom row reports estimated returns to scale.

Table 4: Output Elasticities: Multi-industry baseline estimation using single product plants

	All	15	17	18	21	24	25	26	27	28	29	31
θ_k	0.219	0.170	0.129	0.056	0.425	–0.049	0.122	0.262	0.257	0.298	0.311	0.706
	(0.765)	(0.062)	(0.095)	(0.400)	(0.384)	(0.663)	(2.308)	(0.670)	(0.220)	(0.597)	(0.508)	(0.606)
θ_l	0.162	0.069	–0.459	0.049	0.087	–0.006	0.624	0.853	0.052	0.073	–0.161	–0.182
	(0.602)	(0.025)	(0.341)	(0.447)	(0.373)	(0.156)	(1.131)	(0.274)	(0.251)	(0.504)	(0.168)	(0.811)
θ_m	0.828	0.897	1.497	0.914	0.686	1.064	0.659	0.181	0.888	0.867	1.217	0.740
	(0.658)	(0.043)	(0.246)	(0.326)	(0.666)	(0.508)	(1.247)	(0.686)	(0.261)	(0.106)	(0.365)	(0.312)
RTS	1.209	1.135	1.167	1.019	1.198	1.008	1.404	1.296	1.197	1.238	1.367	1.264
	(.114)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)

Notes: The above table reports the estimated average output elasticities and their standard deviation across single-product plants. Output elasticities come from seperate translog production functions estimated for each of the listed 2-digit NIC sectors. The first column reports the statistics for all sectors, while the remaining columns report statistics for by sector. The bottom row reports estimated returns to scale.

Table 5: Output Elasticities: Multi-industry linear dynamic panel estimation using single product plants

	All	15	17	18	21	24	25	26	27	28	29	31
θ_k	0.284	0.170	0.104	0.235	0.561	0.292	0.550	0.275	0.153	0.186	0.366	0.692
	(0.462)	(0.211)	(0.213)	(0.760)	(0.695)	(0.217)	(0.292)	(0.560)	(0.230)	(0.486)	(0.338)	(0.557)
θ_l	0.140	0.096	–0.134	0.161	0.122	0.066	–0.054	0.704	0.048	0.233	–0.313	–0.266
	(0.704)	(0.108)	(0.570)	(1.305)	(0.177)	(0.059)	(0.460)	(0.166)	(0.183)	(1.779)	(0.123)	(0.435)
θ_m	0.710	0.853	1.109	0.462	0.482	0.692	0.640	0.247	0.918	0.902	1.103	0.702
	(0.668)	(0.198)	(0.676)	(0.585)	(0.856)	(0.233)	(0.479)	(0.534)	(0.409)	(1.439)	(0.215)	(0.180)
RTS	1.134	1.12	1.079	.857	1.164	1.05	1.136	1.225	1.119	1.321	1.156	1.128
	(.099)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)

Notes: The above table reports the estimated average output elasticities and their standard deviation (in parentheses) across single-product plants. Output elasticities come from seperate translog production functions estimated for each of the listed 2-digit NIC sectors. The first column reports the statistics for all sectors, while the remaining columns report statistics for by sector. The bottom row reports estimated returns to scale.

Table 6: Output Elasticities: Baseline estimation using single product plants— Alternative Instruments

	All	15	17	18	21	24	25	26	27	28	29	31
θ_k	–0.13	–0.10	–0.117	–0.152	–0.103	–0.133	–0.10	–0.159	–0.093	–0.134	–0.141	–0.125
	(0.081)	(0.067)	(0.078)	(0.063)	(0.059)	(0.087)	(0.063)	(0.085)	(0.070)	(0.067)	(0.062)	(0.066)
θ_l	0.160	0.136	0.147	0.175	0.136	0.167	0.133	0.174	0.142	0.167	0.169	0.163
	(0.089)	(0.081)	(0.087)	(0.072)	(0.068)	(0.100)	(0.071)	(0.093)	(0.076)	(0.078)	(0.072)	(0.077)
θ_m	1.067	1.062	1.067	1.074	1.065	1.063	1.064	1.083	1.049	1.065	1.069	1.06
	(0.031)	(0.031)	(0.030)	(0.029)	(0.022)	(0.031)	(0.025)	(0.025)	(0.029)	(0.029)	(0.028)	(0.031)
RTS	1.097	1.097	1.097	1.097	1.097	1.097	1.097	1.097	1.097	1.097	1.097	1.097
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)

Notes: The above table reports the estimated average output elasticities and their standard deviation (in parentheses) across single-product plants. Output elasticities are based on a single translog estimated for the entire Indian economy. The first column reports the statistics for all sectors, while the remaining columns report statistics for India's largest manufacturing sectors (organized by 2-digit NIC code). The bottom row reports estimated returns to scale.

Table 7: Output Elasticities: Cost Shares by 2-digit NIC

	All	15	17	18	21	24	25	26	27	28	29	31
θ_k	0.064	0.039	0.080	0.039	0.104	0.081	0.062	0.133	0.075	0.035	0.039	0.037
	(0.033)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
θ_l	0.068	0.047	0.079	0.138	0.068	0.040	0.057	0.071	0.048	0.080	0.090	0.070
	(0.029)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
θ_m	0.868	0.914	0.842	0.823	0.828	0.878	0.881	0.796	0.877	0.884	0.871	0.892
	(0.043)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
RTS	1	1	1	1	1	1	1	1	1	1	1	1
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)

Notes: The above table reports the estimated average output elasticities and their standard deviation (in parentheses) across all plants. Output elasticities obtained by the industry cost shares for each NIC2 industry. The first column reports the statistics for all sectors, while the remaining columns report statistics for India's largest manufacturing sectors (organized by 2-digit NIC code). The bottom row reports estimated returns to scale.

Table 8: Output Elasticities: Cost Shares by 2-digit NIC× Labour-to-materials quintiles

	All	15	17	18	21	24	25	26	27	28	29	31
θ_k	0.066	0.044	0.079	0.045	0.105	0.096	0.066	0.103	0.078	0.042	0.050	0.044
	(0.031)	(0.017)	(0.011)	(0.020)	(0.026)	(0.016)	(0.015)	(0.027)	(0.024)	(0.014)	(0.021)	(0.014)
θ_l	0.116	0.063	0.108	0.168	0.074	0.101	0.080	0.218	0.052	0.120	0.138	0.100
	(0.112)	(0.051)	(0.091)	(0.110)	(0.043)	(0.083)	(0.058)	(0.155)	(0.048)	(0.092)	(0.098)	(0.077)
θ_m	0.817	0.892	0.813	0.787	0.821	0.803	0.854	0.680	0.870	0.838	0.812	0.856
	(0.122)	(0.066)	(0.085)	(0.130)	(0.058)	(0.097)	(0.072)	(0.131)	(0.070)	(0.106)	(0.119)	(0.088)
RTS	1	1	1	1	1	1	1	1	1	1	1	1
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)

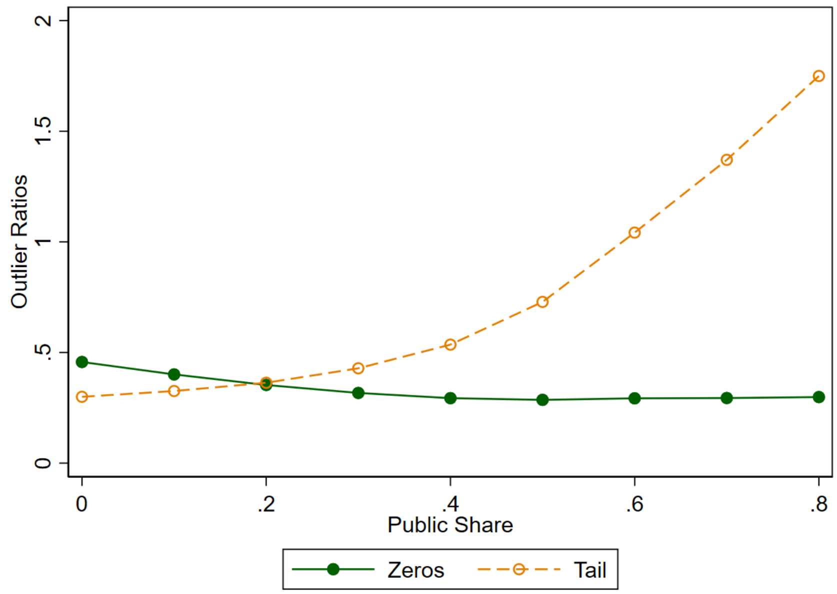
Notes: The above table reports the estimated average output elasticities and their standard deviation (in parentheses) across all plants. Output elasticities obtained by the industry cost shares for each materials-to-labour quintile in each NIC2 industry. The first column reports the statistics for all sectors, while the remaining columns report statistics for India's largest manufacturing sectors (organized by 2-digit NIC code). The bottom row reports estimated returns to scale.

O Allowing for Scope and Varying Estimators

In this Appendix, we consider whether two further sources of potential misspecification can explain the zeros puzzle; assuming non-joint production when production may be joint, and misspecification of the production technology. Figure 12 considers the case of joint production by relying on output distance function (13), in which case $1 - \beta$ captures the share of public, non-rival intermediates in production. While we find a slight decrease in the degree of zeros as we increase the public share, this is more than counterbalanced by a significant increase the the proportion of tail outliers, suggesting that joint production, on its own, is unlikely to solve the zero puzzle.

In Figure 13, we examine whether alternative methods of estimating material's output elasticity can resolve the zeros puzzle. We report outlier ratios derived from: (1) our baseline estimates, (2) a linear dynamic panel estimator, (3) estimating a separate production function for each of the largest 2-digit NIC industries, (4) a linear dynamic panel estimator applied by industry to the same set of industries as in (3), (5) our baseline approach but using k_{it} as an instrument to identify β_K rather than lagged investment, (6) a cost share estimator by 2-digit NIC industry], and (7), a cost share estimator applied to each 2-digit NIC industry $\times$ five separate labour to material expenditure quintiles, as recommended by Raval (2023). Each set of bars in Figure 13 reports outlier ratios where input shares are calculated using (33) as well as using revenue shares. Each set of bars tells a similar story: there is a sizeable number of zero and tail outliers when assume that there is no MRT heterogeneity, but many fewer when we allocate inputs using revenue shares.

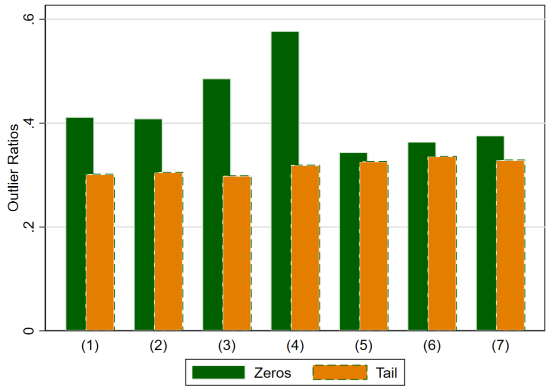
Figure 12: Product-level Markup Outlier Ratios with Scope



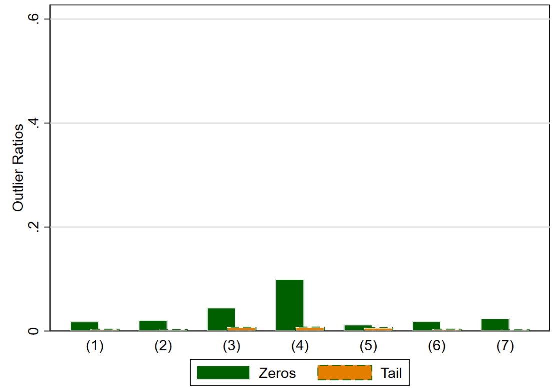
Notes: The above reports outlier ratios for different models of product-level markups. Letting $\mathbb{Z}$ denote the set of product-level markups where $\mu \leq 0.5$, $\mathbb{T}$ denote the set of product-level markups where $\mu \geq 5$, and $\mathbb{R}$ denote product-level markups between 0.5 and 5, the **Zeros** outlier ratio (represented by circles) is defined as $\frac{|\mathbb{Z}|}{|\mathbb{R}|}$, while the **Tail** outlier ratio (represented by triangles) is defined as $\frac{|\mathbb{T}|}{|\mathbb{R}|}$, where $|.|$ denotes the number of elements in a given set. These statistics as a function of the public input share, $1-\beta$, where markups are calculated using equation (34) with $\rho_{it}^{j} = \frac{\left(Y_{it}^{j}\right)^{\frac{1}{\beta\phi}}}{\sum_k \left(Y_{it}^{k}\right)^{\frac{1}{\beta\phi}}}$..

Figure 13: Product-level Markup outlier Ratios Across Methods

(A) Baseline



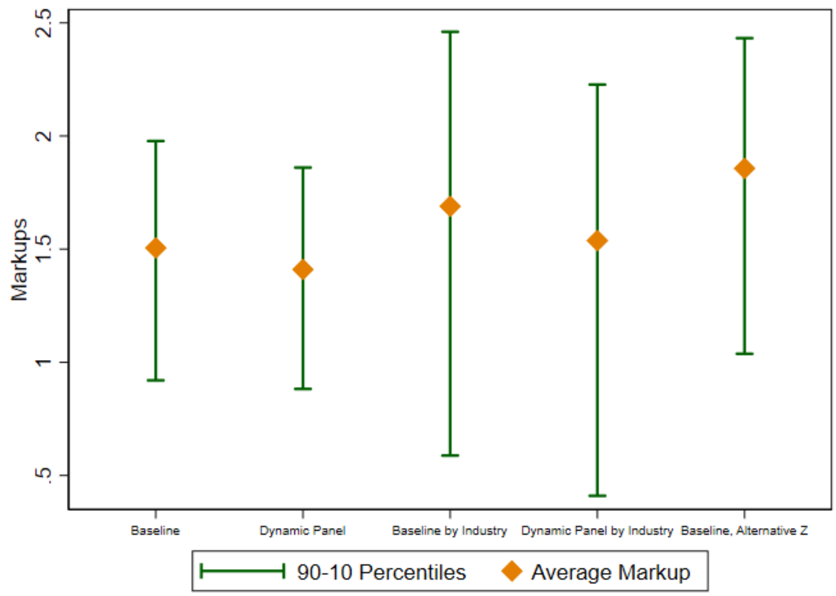
(B) Revenue Share



Notes: The above bar chart plots two outlier ratios for different models of product-level markups for the sub sample of multi-product plants. Letting $\mathbb{Z}$ denote the set of product-level markups where $\mu \leq 0.5$, $\mathbb{T}$ denote the set of product-level markups where $\mu \geq 5$, and $\mathbb{R}$ denote product-level markups between 0.5 and 5, the **Zeros** outlier ratio is defined as $\frac{|\mathbb{Z}|}{|\mathbb{R}|}$, while the **Tail** outlier ratio is defined as $\frac{|\mathbb{T}|}{|\mathbb{R}|}$, where $|.|$ denotes the number of elements in a given set. Panel (A) reports outlier statistics when markups are calculated using (33) and (34), as in our baseline. Panel (B) reports outlier statistics when markups are calculated using (34) with ρ_{it}^{j} equalling a product's revenue share (**Rshare**). The numbers of the x axis denote different ways of estimating relevant output elasticities to recover product-level markups. **Specification (1)** refers to our baseline estimates, where a single production function is estimated for all of India, using a control function approach with a selection correction. **Specification (2)** recovers output elasticities using a linear dynamic panel estimator. **Specification (3)** applies a control function with a selection correction by 2-digit NIC industry (industries 15, 17, 18, 21, 24, 25, 26, 27, 28, 29, and 31). **Specification (4)** applies the linear dynamic panel estimator by 2-digit NIC for the same set of industries. **Specification (5)** applies the control function estimator with the selection correction using, using the current capital k_{it} as instrument to identify β_K, rather than lagged investment as in our baseline. **Specification (6)** estimates the materials output elasticity by 2-digit NIC code using the cost share approach. **Specification (7)** estimates the materials output elasticity by 2-digit NIC code $\times$ labour to material expenditure quintiles using the cost share approach. Specifications (1) through (5) are rely on a homogeneous translog specification of the production function, while Specification (6) and (7) rely on Cobb-Douglas.

P Plant-Markups: Alternative Estimation Routines

Figure 14: Plant-level Markup Dispersion Across Methods



Notes: The above plots the mean plant-level markup, as well as the 90th to 10th percentile of markups, across various methods for calculating the plant-level materials elasticities. **Baseline** uses the estimates from Table 2 . **Dynamic Panel** obtains plant-level markups using the elasticities in Table 3. **Baseline by Industry** obtains plant-level markups using the elasticities in Table 4. **Dynamic Panel by Industry** obtains plant-level markups using the elasticities in Table 5. **Baseline, Alternative Z** obtains plant-level markups using the elasticities in Table 6.

Q De Loecker et al. (2016) selection correction when there is joint production and markup heterogeneity

De Loecker et al. (2016) justify their selection correction with reference to a model similar to Mayer et al. (2014), which rules out joint production. In this appendix, we show that their approach to correcting for selection can still be valid even if firms produce using joint production technologies. In particular, we work through a simple parametric example of demand which generates a selection rule as in (83).

Q.0.1 Market Structure, Preferences, and Technologies

We consider a model of monopolistic competition with variable markups, based on Kimball (1995). Specifically, we consider a representative consumer whose overall welfare, V, is implicitly defined by the following equation:

$$\int_{\omega \in \Omega} \Upsilon \left(\frac{Y(\omega)}{V} \right) d\omega = 1, \tag{101}$$

where $Y(\omega)$ is the total quantity consumed of variety ω, Ω is the set of all available varieties, and $\Upsilon(\psi)$ for $\psi \equiv \frac{Y(\omega)}{V}$ is a function satisfying $\Upsilon^{'}(\psi) > 0$ and $\Upsilon^{'}(\psi) < 0$. For our purposes, it will be convenient to work with the following specification of $\Upsilon(\psi)$:

$$\Upsilon(\psi) = \eta \ln(\psi) + \beta \frac{\sigma}{\sigma - 1} (\psi)^{\frac{\sigma-1}{\sigma}}, \tag{102}$$

where η, β and σ are utility parameters.[69]

We assume that the representative consumer minimizes their expenditure on all goods in Ω, subject to the preference relation (101). This leads to the following inverse demand function for variety ω:

$$P(Y(\omega)) = \frac{\lambda}{V} \Upsilon^{'} \left(\frac{Y(\omega)}{V} \right) = \frac{\lambda}{V} \left(\frac{\eta V}{Y_i} + \beta \left(\frac{Y(\omega)}{V} \right)^{\frac{-1}{\sigma}} \right) \tag{103}$$

where λ is the consumer's Lagrangian multiplier from their expenditure minimization problem. We assume that there are infinitely many firms in the market, so Ω is infinitely large, and therefore all firms take λ and V as given It is then straightforward to show that these preferences lead to the following variable markup rule:

$$\mu(\omega) = \frac{P(\omega)}{MC(\omega)} = \frac{\sigma}{\sigma - 1} \left(1 + \frac{\eta}{\beta} \left(\frac{Y(\omega)}{V} \right)^{-\frac{\sigma-1}{\sigma}} \right) \tag{104}$$

While we assume there infinitely many firms in the market, we assume that each firm produces at most $N + 1$ products, where N is a finite integer. We then assume that each firm produces using the following production possibility frontier.

$$\left(\sum_j \left(\frac{Y_i^j}{A_i^j} \right)^{\frac{1}{\alpha}} \right)^{\alpha} = (K_i)^{1-\beta_L} (L_i)^{\beta_L} \tag{105}$$

Here $\alpha \neq 1$, which allows for joint production. We assume, similar to Olley and Pakes

[69]It is straightforward to verify that if $\eta = 0$, then (101) and (102) together imply that the representative consumer has standard CES preferences.

(1996), that capital predetermined at time t, while labour is static.[70]

Following Mayer et al. (2014), we assume that each firm has a core variety, which is simply the product that highest the highest single-product productivity index A_i^j. For simplicity, we index this product by $j = 0$ for each firm. Products $j = 1, 2, ...$ decline in productivity according to the following deterministic productivity ladder:

$$\frac{A_i^j}{A_i^0} = \delta^j \tag{106}$$

Where $\delta \in (0, 1)$.

Since capital is pre-determined, a firm's total costs for producing some vector of outputs $\mathbf{Y}_i \equiv (Y_i^0, Y_i^1, ..., Y_i^N)$, conditional on K_i, firm specific wages w_i, and their vector of productivity shocks $\mathbf{A}_i \equiv (A_i^0, A_i^1, ..., A_i^N)$ is given by:

$$C(\mathbf{Y}_i, \mathbf{A}_i, K_i, w_i) = \frac{w_i}{(K_i)^{\frac{1-\beta_L}{\beta_L}}} \left(\sum_{j=0}^{J_i} \left(\frac{Y_i^j}{A_i^j} \right)^{\frac{1}{\alpha}} \right)^{\frac{\alpha}{\beta_L}} = \frac{w_i}{\left(K_i^{1-\beta_L} A_i^0 \right)^{\frac{1}{\beta_L}}} \left(\sum_{j=0}^{J_i} \left(\frac{Y_i^j}{\delta^j} \right)^{\frac{1}{\alpha}} \right)^{\frac{\alpha}{\beta_L}} \tag{107}$$

where the second equality uses (106).

Q.0.2 Characterizing the Variable Profit Function

Each firm can produce their core variety $j = 0$, at no additional costs beyond the labour and capital necessary to produce each unit quantity. However, each marginal variety beyond their core variety, $j = 1, 2, ...$ requires that the firm pay a fixed cost f one period *before* they can begin selling it. Before we determine the product selection rule, consider a exogenously firm producing J varieties. Since productivity declines in the index j, a firm producing J varieties will produces varieties $j = 0, 1, ..., J - 1$. Given this set of varieties, they choose their quantities to maximize their profits, i.e.:

$$\max_{\mathbf{Y}_i} \sum_{j=0}^{J-1} \lambda\eta + \beta\lambda \left(\frac{Y_i}{V} \right)^{\frac{\sigma-1}{\sigma}} - \frac{w_i}{\left(K_i^{1-\beta_L} A_i^0 \right)^{\frac{1}{\beta_L}}} \left(\sum_{j=0}^{J-1} \left(\frac{Y_i^j}{\delta^j} \right)^{\frac{1}{\alpha}} \right)^{\frac{\alpha}{\beta_L}} \tag{108}$$

After evaluating each of the firm's J first order conditions, we obtain the following closed for expression for variable profits, as a function of J, the firm's predetermined capital stock K_i, their core product productivity level A_i^0, and firm specific wages w_i.

[70] We simply limit out attention to one dynamic input and one static input here to save on notation, but the approach easily generalizes to multiple static inputs inlucding materials.

$$\Pi(J, K_i, A_i^0, w_i) = \gamma_0 J + \gamma_1 \left(\frac{\left(K_i^{1-\beta_L} A_i^0\right)^{\frac{1}{\beta_L}}}{w_i} \right)^{\frac{(\sigma-1)\beta_L}{\sigma+\beta_L-\sigma\beta_L}} \left(\sum_{j=0}^{J-1} \left(\delta^j\right)^{\frac{\sigma-1}{\sigma+\alpha-\alpha\sigma}} \right)^{\frac{\sigma+\alpha-\alpha\sigma}{\sigma+\beta_L-\sigma\beta_L}} \tag{109}$$

where:

$$\gamma_0 \equiv \lambda\eta$$

$$\gamma_1 \equiv (\beta\lambda)^{\frac{\sigma}{\sigma+\beta_L-\sigma\beta_L}} V^{\frac{1-\sigma}{\sigma+\beta_L-\sigma\beta_L}} \left(\frac{\sigma-1}{\sigma}\beta_L \right)^{\frac{(\sigma-1)\beta_L}{\sigma+\beta_L-\sigma\beta_L}} \left(\frac{\beta_L+\sigma-\sigma\beta_L}{\sigma} \right)$$

Note that $\alpha + \sigma - \alpha\sigma > 1$ since $\sigma > 1$. Therefore $\frac{\sigma-1}{\alpha+\sigma-\alpha\sigma} > 0$ and since $\delta \in (0,1)$, then $\delta^{\frac{\sigma-1}{\alpha+\sigma-\alpha\sigma}} \in (0,1)$. From the standard formula for geometric series:

$$\Pi(J, K_i, A_i^0, w_i) = \gamma_0 J + \gamma_1 \left(\frac{\left(K_i^{1-\beta_L} A_i^0\right)^{\frac{1}{\beta_L}}}{w_i} \right)^{\frac{(\sigma-1)\beta_L}{\sigma+\beta_L-\sigma\beta_L}} \left(\frac{1-\left(\delta^{\frac{\sigma-1}{\alpha+\sigma-\alpha\sigma}}\right)^J}{1-\delta^{\frac{\sigma-1}{\alpha+\sigma-\alpha\sigma}}} \right)^{\frac{\sigma+\alpha-\alpha\sigma}{\sigma+\beta_L-\sigma\beta_L}}$$

or, by defining:

$$\gamma_2 \equiv \gamma_1 \left(\frac{1}{1-\delta^{\frac{\sigma-1}{\alpha+\sigma-\alpha\sigma}}} \right)^{\frac{\sigma+\alpha-\alpha\sigma}{\sigma+\beta_L-\sigma\beta_L}}$$

$$C_0(K_i, A_i^0, w_i) \equiv \left(\frac{\left(K_i^{1-\beta_L} A_i^0\right)^{\frac{1}{\beta_L}}}{w_i} \right)^{\frac{(\sigma-1)\beta_L}{\sigma+\beta_L-\sigma\beta_L}}$$

$$\Pi(J, K_i, A_i^0, w_i) = \gamma_0 J + \gamma_2 C_0(K_i, A_i^0, w_i) \left(1-\left(\delta^{\frac{\sigma-1}{\alpha+\sigma-\alpha\sigma}}\right)^J \right)^{\frac{\sigma+\alpha-\alpha\sigma}{\sigma+\beta_L-\sigma\beta_L}}$$

To guarantee that each firm produces a finite number of products J, we need $\frac{\partial\Pi(J,A_i^0,K_i,w_i)}{\partial J} > 0$ and $\frac{\partial^2\Pi(J,A_i^0,K_i,w_i)}{\partial J^2} < 0$. Note that it is straightforward to verify that the first condition always holds, since:

$$\frac{\partial \Pi(J, A_i^0, K_i, w_i)}{\partial J} = \gamma_0 - \ln \delta \gamma_2 C_0(K_i, A_i^0, w_i) \frac{\sigma - 1}{\sigma + \beta_L - \sigma\beta_L} \left(1 - \left(\delta^{\frac{\sigma-1}{\alpha+\sigma-\alpha\sigma}}\right)^J\right)^{\frac{\sigma+\alpha-\alpha\sigma}{\sigma+\beta_L-\sigma\beta_L}-1} \delta^{\frac{J(\sigma-1)}{\alpha+\sigma-\alpha\sigma}} > 0 \tag{110}$$

To obtain the second condition, we need to assume that:

$$\delta < \left(\frac{\sigma + \beta_L - \sigma\beta_L}{\alpha + \sigma - \alpha\sigma}\right)^{\frac{\alpha+\sigma-\alpha\sigma}{\sigma-1}} \tag{111}$$

In which case it is easily verified that:

$$\begin{aligned}
\frac{\partial^2 \Pi(J, A_i^0, K_i, w_i)}{\partial J^2} &= (\ln \omega)^2 \gamma_2 C_0 \frac{\sigma - 1}{\sigma + \beta_L - \sigma\beta_L} \frac{(\sigma - 1)}{\alpha + \sigma - \alpha\sigma} \\
&\times \left(1 - \left(\delta^{\frac{\sigma-1}{\alpha+\sigma-\alpha\sigma}}\right)^J\right)^{-1} \frac{\delta^{\frac{J(\sigma-1)}{\alpha+\sigma-\alpha\sigma}}}{(\sigma + \beta_L - \sigma\beta_L)\left(1 - \delta^{\frac{J(\sigma-1)}{\alpha+\sigma-\alpha\sigma}}\right)} \\
&\times \left(\delta^{\frac{J(\sigma-1)}{\alpha+\sigma-\alpha\sigma}} (\alpha - \alpha\sigma + \sigma\beta_L - \beta_L) - (\sigma + \beta_L - \sigma\beta_L)\left(1 - \delta^{\frac{J(\sigma-1)}{\alpha+\sigma-\alpha\sigma}}\right)\right) < 0
\end{aligned}$$

Which establishes that the variable profit function is concave in J.

Q.0.3 Product Selection Rule

Recall that we previously made the assumption that a firm will choose to add products at time $t-1$. To account for this, we know index all variables with t. Since a firm needs to decide to add products one period before selling that product, and $\Pi(.)$ is concave in J_{it}, then a single product firm will only choose to add a new product in the following period if:[71]

$$\mathbb{E}\left(\Pi(2, A_{i,t}^0, K_{i,t}, w_{i,t}) - \Pi(1, A_{i,t}^0, K_{i,t}, w_{i,t}) | \mathbb{I}_{i,t-1}\right) \geq f, \tag{112}$$

i.e. the marginal gain in profit from adding a new product is less than the marginal fixed cost of adding a new product. Since it is easily verified that:

$$\frac{\partial^2 \Pi(J, A_{it}^0, K_i, w_i)}{\partial A_{it}^0 \partial J} > 0, \tag{113}$$

then $\Pi(2, A_{it}^0, K_{it}, w_{it}) - \Pi_i(1, A_{it}^0, K_{it}, w_{it})$ is strictly increasing in A_{it}^0. As long as A_{it} follows a Markov process

$$A_{it}^0 = \mathbb{E}\left(A_{it}^0 | \mathbb{I}_{i,t-1}\right) + \xi_{i,t} = g\left(A_{i,t-1}^0\right) + \xi_{i,t} \tag{114}$$

[71] For simplicity, we consider myopic firms that one look forward one period when deciding to add products; however, it is straightforward to extend the logic here to the case of infinitely lived, forward looking firms, at the cost of some extra notation.

such that $g\left(A^0_{i,t-1}\right)$ is strictly increasing in $A^0_{i,t-1}$, then $\Pi(2, g\left(A^0_{i,t-1}\right) + \xi_{i,t}, K_{it}, w_{it}) - \Pi_i(1, g\left(A_{i,t-1}\right) + \xi_{i,t}, K_{it}, w_{it})$ is strictly increasing in $A^0_{i,t-1}$ for every realization of ξ_{it}, K_{it}, w_{it}, and therefore, as long as the distribution of w_{it} conditional on $\mathbb{I}_{i,t-1}$ does not depend on $A^0_{i,t-1}$, then $\mathbb{E}\left(\Pi(2, A^0_{i,t}, K_{i,t}, w_{i,t}) - \Pi(1, A^0_{i,t}, K_{i,t}, w_{i,t}) | \mathbb{I}_{i,t-1}\right)$ is strictly increasing in $A^0_{i,t-1}$. This means that the firm will remain single product in the subsequent period as long as $A^0_{i,t-1} \leq \overline{A}$, where $\overline{A}$ solves

$$\mathbb{E}\left(\Pi(2, g\left(\overline{A}\right) + \xi_{i,t}, K_{i,t}, w_{i,t}) - \Pi(1, g\left(\overline{A}\right) + \xi_{i,t}, K_{i,t}, w_{i,t}) | \mathbb{I}_{i,t-1}\right) = f, \tag{115}$$

Note that the relevant variable in the firm's information set at time $t-1$ for predicting future period profits are lagged wages, the lagged capital stock, and previous period's investment (which allows them to perfectly forecast K_{it}, leading to a cutoff productivity level that will be a function of the lagged state variable of the firms, and investment, i.e.

$$J_{it} > 1 \quad \text{if} \quad A_{i,r-1} > \overline{A}_{i,t-1} = \overline{A}_{t-1}(k_{i,t-1}, i_{i,t-1}, w_{i,t-1}), \quad \text{otherwise} \quad J_{it} = 1 \tag{116}$$

which has the same structure as the cutoff rule (83).