# Within-Firm Productivity Dispersion: Estimates and Implications

Scott Orr

*University of British Columbia*

This paper develops a flexible recipe for identifying unobserved input allocations, as well as quantity- and revenue-based total factor productivity (TFP) across product lines, for multiproduct producers, using demand- and supply-side information. Applying variants of this recipe to a panel of plants manufacturing machinery in India from 2000–2007 yields sizable within-plant TFP differences. Removing an average plant's lowest-performing product increases the unweighted average of plant-level revenue-based productivity by 10%–65%. A 1 standard deviation decline in product-level revenue-based TFP generates around a 6 percentage point increase in the probability of dropping that product.

This paper is based on the first chapter of my PhD thesis and was previously circulated as "Productivity Dispersion, Import Competition, and Specialization in Multi-product Plants." I wish to thank Chad Syverson and four anonymous referees, who provided very helpful feedback on previous drafts of this paper. I also wish to thank my dissertation committee, Daniel Trefler, Victor Aguirregabiria, and Aloysius Siow, for their advice on this project, as well as Stephen Ayerst, Dominick Bartelme, Bernardo Blum, Kunal Dasgupta, Jan De Loecker, Kevin Devereux, Daniel Ershov, Garth Frazer, Leandro Freylejer, Nicolas Gendron-Carrier, Daniel Goetz, Ashique Habib, Keith Head, Ignatius Horstmann, Amit Khandelwal, Nicholas Li, Kevin Lim, Gaelan MacKenzie, Mathieu Marcoux, Pamela Medina, Juan Morales, Peter Morrow, Jean-William P. Laliberté, Swapnika Rachapalli, Eduardo Souza-Rodrigues, Sharon Traiberman, Eric Verhoogen, Frederic Warzynski, David Weinstein, Erhao Xie, and Genet Zinabou; a number of participants at the 2017 CEA (Canadian Economics Association) and EARIE (European Association for Research in Industrial Economics) conferences, the 2018 Danish International Economics Workshop, and the 2019 IIOC (International Industrial Organization Conference), Rocky Mountain Empirical Trade, and WEAI (Western Economic Association International) conferences; and seminar participants at the Bank of Canada, Columbia, McMaster, Simon Fraser University, University of British Columbia Sauder, University of Chicago Booth, and the University of Toronto. Any remaining errors are my own. Financial support from the Social Sciences and Humanities Research Council (SSHRC) is gratefully acknowledged. Replication files are available in a zip file. This paper was edited by Chad Syverson.

Electronically published September 30, 2022

*Journal of Political Economy*, volume 130, number 11, November 2022.

https://doi.org/10.1086/720465

## I. Introduction

Individual producers vary widely in terms of productivity. If these productivity differences are driven by heterogeneity in skill sets across firms, it stands to reason that some firms will be better at producing some products and worse at producing others. Whether such within-firm productivity dispersion exists is an important empirical question. However, to date there has been very little work documenting whether within-firm heterogeneity is quantitatively important.

The lack of empirical evidence on within-firm productivity differences is unlikely to be due to lack of importance or interest. For example, within-firm productivity dispersion is often invoked in the finance literature to understand why large, diversified conglomerates appear to trade at a discount (Maksimovic and Phillips 2002; Schoar 2002).[1] Similarly, a series of papers in the international trade literature (Eckel and Neary 2010; Bernard, Redding, and Schott 2011; Mayer, Melitz, and Ottaviano 2014, 2021) emphasize that within-firm heterogeneity can have important implications for productivity growth, as competition shocks lead firms to shed low-productivity varieties, increasing productivity.

Moreover, the magnitude of within-firm productivity dispersion is important for thinking about how firms might choose to optimally position their products. In particular, while much of the empirical industrial organization literature has often emphasized the importance of demand shifting product characteristics as a key strategic variable of the firm (e.g., Crawford 2012), another important dimension for thinking about where a firm will locate in product characteristic and price space is the magnitude of within- and across-firm cost heterogeneity, which may be driven by idiosyncratic productivity differences across product lines.[2]

Unfortunately, empirical evidence on the magnitude of within-firm heterogeneity is primarily lacking because of data limitations. The standard approach to measuring productivity uses production functions, which require an estimate of input use by output. However, the vast majority of data sets on manufacturing units record only input use at the firm or plant

[1] Maksimovic and Phillips (2002) argue that this "conglomerate discount" arises as a result of comparative advantage differences across industry segments within the conglomerate; highly specialized firms tend to be really good at one industry, while diversified firms have less variability in productivity across industry segments, leading single-segment firms to be more productive than conglomerates of a similar size. Schoar (2002), on the other hand, shows that the conglomerate discount is primarily a transitory phenomenon, as firm-level productivity decreases only after a plant operating in a new industry segment is acquired. She argues that these effects are due to managerial inputs being focused on integrating the new acquisition at the expense of incumbent industry lines.

[2] This distinction between cost-based advantages and product-differentiation-based advantages is also often emphasized in the strategic management literature, e.g., Porter (1980).

level, rather than at the firm-product level, making the production function approach infeasible.[3]

To address these shortcomings, this paper provides three key contributions. First, I develop a flexible recipe for identifying unobserved input allocations and total factor productivity (TFP) across product lines for multiproduct producers, using demand-side information.[4] In particular, I show that standard duality results that follow from profit maximization conditions generate a mapping from observable output prices and quantities to the unobservable input allocations for a wide class of models that are commonly used in applied work. Second, I apply this approach to a panel of plants manufacturing machinery in India, in order to compare the magnitude of within-producer heterogeneity to that of across-producer heterogeneity. I find sizable within-plant heterogeneity, accounting for around one-third of plant-product-level productivity differences. These within-plant productivity differences are found to have important quantitative implications: after estimating a within-plant core-competency efficiency ladder as in Eckel and Neary (2010) and Mayer, Melitz, and Ottaviano (2014), I find that removing an average plant's lowest-productivity variety increases the unweighted average of plant-level log revenue–based TFP by 10%–65%. Third, I show that plants are much more likely to drop low-productivity varieties, compared to high-productivity varieties. These results provide evidence that within-producer productivity differences are being taken into account by the producers themselves. Since firms appear to care about these within-firm productivity differences, abstracting from this form of heterogeneity is unlikely to be innocuous, as has been done in the vast majority of previous work on productivity estimation.

This paper can be broken down into roughly two parts. In the first part, which comprises section II, I outline how information in standard production data sets can be used to identify within-firm productivity differences. The approach exploits the fact that firm-product output prices and quantities contain information on within-firm input use. To illustrate

[3] There are some key exceptions, including Lamorgese, Linarello, and Warzynski (2015) and Garcia-Marin and Voigtländer (2017, 2019), who are able to estimate within-plant productivity dispersion in Chile using a novel data set containing information on cost shares by product line within a plant. Another important and related line of research includes Ichniowski, Shaw, and Prennushi (1997) and Ichniowski and Shaw (1999), who examine the effect of various management practices on production-line-specific productivity within Japanese and American steel mills, using product-line-specific "uptime," a novel productivity measure available for their specific study that is proportional to production delays.

[4] Note that the approach assumes that the researcher has access to quantity data, and as a result the measure of TFP identified in this paper is quantity TFP, or TFPQ, as described by Foster, Haltiwanger, and Syverson (2008). I often simply refer to the object as "TFP," unless a distinction must be made between TFPQ and revenue-based TFP, which I also consider when aggregating to the level of a firm or plant.

the basic idea behind the approach, consider a firm producing two outputs using labor. Suppose a researcher observes the level of output across the two production lines, as well as the total quantity of labor purchased by the firm. The key difficulty is that differences in output across production lines can be due to either differences in labor input or differences in unit labor requirements, thereby frustrating a researcher's ability to estimate within-firm TFP differences.

This problem can be solved by combining profit maximization conditions with some restrictions on the shape of the production technology. For example, suppose that a researcher is confident that both production lines are characterized by constant returns to scale and that the market for each output is perfectly competitive.[5] Since firms are price takers, the marginal cost of each product will have to equal its price. Given constant returns to scale, this implies that the price of each good will equal the wage over the unit labor requirements. This means that the ratio of output prices within the plant will be proportional to the ratio of unit labor requirements. Therefore, price variation provides information on relative productivity dispersion within a plant. One can combine this information with the structure of the production technologies, as well as an aggregate resource constraint, to generate four equations that uniquely pin down the four unknowns: the two unobserved unit labor requirements and the two unobserved labor allocations.[6]

While the above example imposes strong assumptions (single-input technologies, perfect competition, constant returns to scale), profit maximization separately identifies input allocations from within-firm TFP differences for a much wider class of pricing models and production technologies. These include models of oligopolistic competition and collusion, which account for cross-product cannibalization effects, as well as more general multiple-input production technologies with increasing or decreasing returns to scale. Unsurprisingly, this increased generality comes at a cost; one must know the shape of a firm's demand function to disentangle the information contained in firm-product output prices on input allocations from the information prices convey on differences in market

[5] While the scale of the firm is theoretically indeterminate with constant returns to scale and perfect competition, for this exercise one can consider a firm choosing an arbitrary scale of production consistent with zero profits or consider an isolated local market with a fixed quantity of inelastically supplied labor, where local wages adjust so that a representative firm earns zero profits by hiring all available labor.

[6] More formally, since labor is the only input and production is characterized by constant returns to scale, we can write $Y^j = A^j L^j$, where $j = 1, 2$ indexes products within the plant, $A^j$ is the unit labor requirement for output $j$, and $L^j$ is the quantity of labor used to produce $j$. This provides two equations. As long as labor is completely attributable to each production line, this provides a third restriction, $L^1 + L^2 = L$, where $L$ is the total quantity of observed labor. Profit maximization and perfect competition imply that $P^j = w/A^j$ for $j = 1, 2$, which implies that $P^1/P^2 = A^2/A^1$. This is the fourth condition necessary to uniquely determine $(A^1, A^2, L^1, L^2)$ using information on $(Y^1, Y^2, P^1, P^2, L)$.

power across production lines. However, as long as one observes outputs and prices by product line, one can use standard demand estimation tools to do this, as was previously exploited by Valmari (2016) in a setting where monopolistic producers face isoelastic demand. More generally, as long as one can recover product-level marginal costs from output and price information, which Berry and Haile (2014) show is possible under very general conditions, then one can use marginal cost information to recover firm-product input allocations, which can then be used to recover within-firm TFP differences.[7]

These new identification results suggest a straightforward, two-step recipe for estimating within-firm productivity dispersion in data sets that contain information on output prices, quantities, and aggregate inputs.[8] First, obtain an estimate of input use by product line by estimating demand and imposing a conduct assumption to recover marginal costs. Second, use the estimates of plant-product input use to estimate a firm-product-level production function. Since the approach allows for nonparametric specifications of demand and the production function and does not explicitly restrict how productivity evolves over time, this recipe can easily be adapted to many theoretical and empirical settings, while also allowing researchers to apply a variety of different estimators for demand and production functions that are appropriate for the particular application.[9]

In the second key part of the paper, in sections III–V, I apply a variant of this methodology to a panel of plants manufacturing machinery in India, using data from the 2000–2007 Annual Survey of Industries (ASI). The magnitude of within-plant productivity differences is important to understand in this setting, since multiproduct plants dominate manufacturing in India, while product turnover accounts for more than a quarter of manufacturing net sales growth over the 2002–8 time period (Boehm, Dhingra, and Morrow 2018). Whether this product turnover is also an

[7] While marginal costs are directly proportional to TFP differences if there are constant returns to scale, note that this is no longer true if there are increasing or decreasing returns to scale. As a result, uncovering within-firm TFP differences can convey additional information beyond what can be uncovered from simply examining marginal cost differences.

[8] Examples of this type of data include the US Census of Manufactures used in Foster, Haltiwanger, and Syverson (2008), the Belgian PRODCOM (*production communautaire*) data used in Dhyne et al. (2017), the Canadian manufacturing data used in Baldwin and Gu (2009), and the Indian Prowess data used in Goldberg et al. (2010b) and De Loecker et al. (2016), as well as the Indian ASI (Annual Survey of Industries) data used in this paper.

[9] While the approach does not explicitly require that researchers take a stance on the law of motion for productivity over time, it does require that firms have full information on their technology levels when choosing inputs. While this rules out proxy-variable approaches to production function estimation (e.g., Olley and Pakes 1996; Levinsohn and Petrin 2003; Ackerberg, Caves, and Frazer 2015; Gandhi, Navarro, and Rivers 2020), other popular estimators, such as dynamic panel (Blundell and Bond 2000) or cost-share-based (Syverson 2004; Foster, Haltiwanger, and Syverson 2008; Backus 2020) approaches, can easily be applied.

important source of productivity growth cannot be known unless the magnitude of within-firm productivity is estimated.

Applying the approach developed in this paper to the machinery, equipment, and parts industry, I find that productivity varies widely across production lines within the same plant, with within-plant variation in TFPQ accounting for 36% of the variance in variety-level TFPQ differences. In general, a plant performing well in one product line does not imply that it is equally good at producing other products; for example, multiproduct plants producing a variety in the 90th percentile of the TFPQ distribution for a particular product code will also produce other products that are, on average, ranked only around the 60th–70th percentile of their respective product codes.

I then use my estimates of within-producer productivity differences to consider whether the extensive margin of product choice can drive plant-level revenue-based productivity differences. For this purpose, I estimate a within-plant efficiency ladder by ranking each plant's products according to revenue-based TFP ("revenue efficiency"), which incorporates variation in both cost heterogeneity (TFPQ) and product appeal at the product level.[10] On average, removing a plant's lowest-performing product can generate 10% to just under 65% increases in the unweighted average of plant-level revenue efficiency, with the largest gains accruing to plants that produce only two products.

This paper concludes by providing estimates of the probability of dropping a product conditional on variety-level revenue efficiency, TFPQ, and product appeal. These results establish whether plants are actually aware of these sources of heterogeneity when choosing their production sets, which is a necessary condition for the various mechanisms related to the "core-competence" models of multiproduct producers, such as Eckel and Neary (2010) and Mayer, Melitz, and Ottaviano (2014), to be valid. I find that plants are more likely to drop low-revenue-efficiency products, with a 1 standard deviation decline in revenue efficiency being associated with around a 6.5 percentage point increase in the probability of ceasing production of that product. After further decomposing revenue efficiency into common product-level shocks, as well a product's relative TFPQ and product appeal ranking within a product code, I find that plants are more responsive to demand-based product appeal variation than they are to TFPQ variation. A 1 standard deviation decrease in appeal, holding the productivity of all other products fixed, generates an increase in dropping probabilities of around 13 percentage points. On the other hand, a similar shock to within-plant TFPQ generates an increase in product dropping of

[10] The notion of product appeal used in this paper is based on a demand residual, as in Khandelwal (2010), Amiti and Khandelwal (2013), and Hottman, Redding, and Weinstein (2016).

around 7 percentage points. These findings provide support for the core-competence model of multiproduct producers, as plants appear to be becoming "leaner and meaner" as they drop varieties (Eckel and Neary 2010), while clarifying that this mechanism is primarily driven by demand-side appeal differences, rather than cost-side productivity differences.

*Relationship to related literature.*—This paper contributes to a growing literature on how to estimate the productivity of multiproduct producers. Many approaches in this literature use information on multiple outputs to recover an estimate of firm-level productivity only. This is accomplished by using a firm-level output index to deflate total revenues, as in Eslava et al. (2004) and Smeets and Warzynski (2013), making use of an explicit within-firm aggregation model, as in Balat, Brambilla, and Sasaki (2016) and Eslava and Haltiwanger (2018), or by using a firm-level measure of technology differences for multiple-output technologies, as in Caves, Christensen, and Diewert (1982), or, more recently, Maican and Orth (2021) and Malikov and Lien (2021). While certainly useful for many questions, these approaches cannot estimate within-firm efficiency differences.[11]

An alternative approach estimates within-firm input allocations that apply to a firm-product-specific production technology. Many papers simply use revenue shares for this purpose, including Foster, Haltiwanger, and Syverson (2008), Atalay (2014), and Collard-Wexler and De Loecker (2015). A direct implication of the approach developed in this paper is that revenue shares are the correct way to allocate inputs, as long as there are no within-producer differences in markups and the production technology is homogeneous and identical across production lines, up to variety-specific TFP differences.[12]

De Loecker et al. (2016), Valmari (2016), Itoga (2019), and Gong and Sickles (2021) also provide models that can be used to uncover unobserved input allocations. The approach developed in De Loecker et al. (2016) allows a researcher to estimate input allocations without taking a stance on demand or the form of competition but also requires that there be no within-firm TFP dispersion, which by assumption rules out the within-firm efficiency differences that are the core object of interest in this paper. Valmari (2016), using a model with Cobb-Douglas production technologies and monopolistic producers facing isoelastic demand, shows that one can use output price and quantity information to estimate input allocations when there is unobserved TFP variation within a firm,

[11] It is worth noting that while the index-number approach is often applied to uncover firm-level measures of productivity, the general approach to uncovering input shares and within-firm productivity dispersion used in this paper shares some similarities with the index-number approach used in Caves, Christensen, and Diewert (1982), as both approaches rely on duality results to provide estimates of various unobservables.

[12] Collard-Wexler and De Loecker (2015) also remark that revenue shares are appropriate in their setting as long as markups are constant within firm, although their approach assumes no within-firm TFP dispersion.

assuming that productivity shocks are realized before all inputs are chosen, as well as zero adjustment costs for reallocating dynamic inputs across product lines. Itoga (2019) follows a similar strategy while instead allowing for oligopolistic competition with nested CES (constant elasticity of substitution) demand. Gong and Sickles (2021) impose exogenous output prices while also considering some different timing assumptions in a firm's input allocation decision and also allowing for productivity shocks that are unobserved by the firm when choosing inputs.

This paper provides a flexible recipe for identifying the unobserved input allocations, allowing for parametric and nonparametric specifications of the production technology as well as a wide class of parametric and nonparametric demand systems involving monopolistic competition, oligopoly, or collusive conduct. For this purpose, the assumptions of costless transferability of dynamic inputs, as well as the timing of the productivity shocks described in Valmari (2016), are crucial. However, this paper improves on past work in this area in three key ways. First, this paper shows that as long as the production function is homogeneous of degree $\phi > 0$ and output elasticities do not vary across product lines, then output prices and quantities fully reveal input allocations, without requiring knowledge of the production function parameters. This allows the researcher to directly include multiproduct firms in production function estimation, which cannot be accommodated by the framework developed in De Loecker et al. (2016). Moreover, since input use is uncovered independently of production function parameters, the researcher can avoid complications generated by the simultaneous estimation of input allocations and production function parameters, as is done in Valmari (2016) and Gong and Sickles (2021). Finally, by considering identification in a general nonparametric setting that allows for a number of different modes of competition, this paper provides a simple recipe for identifying unobserved input allocations, the input allocation rule, which applies to a wide class of models that are commonly used in empirical work. This will allow other researchers to apply these results to their preferred models of interest, rather than limiting attention to the particular models and functional forms considered in previous papers.

An alternative approach to estimating firm-product efficiency is developed by Dhyne et al. (2017, 2021). Rather than estimating input allocations, they estimate a firm-level production technology, the transformation function, to identify firm-product efficiency differences. One advantage of their approach is that it can apply to settings where there is joint production; that is, a firm's technology cannot be represented by firm-product production functions, as some inputs may be used in multiple product lines simultaneously. This paper, on the other hand, provides a comprehensive framework for uncovering within-firm productivity dispersion in nonjoint production settings, where firm-product production functions are well defined. An advantage of this framework is that it generates well-defined measures of productivity dispersion at the product level, even if firms produce different output sets.[13] As a result, the framework developed in this setting allows for direct estimation of a core-competence efficiency ladder, as in Eckel and Neary (2010), Mayer, Melitz, and Ottaviano (2014), and Arkolakis, Ganapati, and Muendler (2021), which quantifies the degree to which product-level efficiency falls as firms add new products and therefore move production away from their most efficient production line.

The empirical results in this paper are closely related to those of Garcia-Marin and Voigtländer (2017), who also estimate within-firm TFP dispersion, although in a special setting where they directly observe cost shares by product line, which they use to allocate inputs. While their paper shares an interest in testing various implications of the core-competence model of multiproduct firms, their focus is primarily on intensive-margin reallocations, rather than the extensive margin—for example, product dropping. By focusing on the extensive margin of adjustment, my paper provides some of the first evidence that plants actually care about these sources of heterogeneity, finding that plants are much more likely to cease production of low-efficiency products, in terms of both physical efficiency (TFPQ) and general demand-based product appeal. Such evidence is lacking in previous literature on this topic, even though it is actually required by many models of multiproduct producers, such as those described in Bernard, Redding, and Schott (2010, 2011), Eckel and Neary (2010), and Mayer, Melitz, and Ottaviano (2014).

This paper also contributes to a growing literature on separately identifying the different sources of firm heterogeneity. I provide a unified framework for identifying a number of important margins of firm heterogeneity, including across- and within-firm variation in TFP, quality, markups, and marginal costs. Previous approaches have largely used techniques that can uncover only a subset of these different margins, including Foster, Haltiwanger, and Syverson (2008) and Jaumandreu and Yin (2016), who focus on separately identifying demand- versus supply-side heterogeneity; De Loecker and Warzynski (2012) and De Loecker et al. (2016), who focus on separately identifying TFP and markups; Forlani et al. (2016) and Blum et al. (2021), who identify markup, demand, and productivity differences; Goldberg et al. (2010a, 2010b), who focus on identifying within-firm improvements driven by product adding and dropping; and Hottman, Redding, and Weinstein (2016), who provide a framework for separately

[13] This advantage is not shared by the approach developed by Dhyne et al. (2017, 2021), since their approach relies on a log-linear transformation function; in this setting, a single-product producer cannot be operating on the same log-linear transformation function as a two-product producer, as one of the function's arguments would go to negative infinity. As a result, producers with different output sets operate fundamentally different technologies, and therefore the relevant residuals, which would form the basis of productivity comparisons, are not directly comparable.

identifying markups, product appeal, and marginal costs across and within firms.

## II. Theoretical Framework

### A. Model Assumptions

During each period $t$, a set of differentiated products, $\Omega_t$, are sold on the market by $N_t$ firms. Each product is produced by a particular firm $i$, with $\mathbb{Y}_{it} \subset \Omega_t$ denoting the set of products produced by firm $i = 1, 2, \ldots, N_t$, and $J_{it} = |\mathbb{Y}_{it}|$ denoting the number of products produced by the firm.[14] Each product (or variety) $j \in \mathbb{Y}_{it}$ is produced using the following production technology:

$$Y_{it}^j = \exp(\omega_{it}^j)F(\vec{X}_{it}^j), \tag{1}$$

where $Y_{it}^j$ is total output of variety $j \in \mathbb{Y}_{it}$, $\vec{X}_{it}^j$ is the vector of inputs used in the production of $j$, and $\omega_{it}^j$ is log TFP, or productivity, of variety $j$.

To obtain an input allocation rule that depends only on demand-side data (prices, quantities, and demand elasticities), I require that the production function satisfy the following restrictions.

Assumption 1. $F(\cdot)$ is continuous and differentiable, equal to zero if any of its arguments are equal to zero, strictly increasing in all arguments, quasi-concave, and homogeneous of degree $\phi > 0$.

Assumption 2. The production technology differs across product lines within a firm only as a result of differences in $\omega_{it}^j$; that is, $F(\cdot)$ does not depend on $j \in \mathbb{Y}_{it}$.

Assumption 1 largely imposes standard regularity conditions on the production function, while still allowing for nonparametric specifications of the production technology. Assumption 2 requires some further discussion, as it differs from related work in important ways. First, note that this assumption allows the TFP shifters, $\omega_{it}^j$, to differ by product line in a completely flexible way—in particular, firms can be good at producing some products while being worse at producing others. This allows a researcher to investigate questions related to the degree of efficiency differences across product lines, which cannot be done using the framework described by De Loecker et al. (2016), which requires that $\omega_{it}^j = \omega_{it}$. On the other hand, this restriction does require ruling out productivity differences that are not Hicks neutral, such as factor-specific productivity differences.

Relatedly, requiring that the shape of the production technology, $F(\cdot)$, not differ across production lines may appear more restrictive than the setting considered in De Loecker et al. (2016), who state that their

[14] In what follows, the operator $|\cdot|$ always refers to the number of elements in a given set.

approach allows the technology to be product specific. Note, however, that assumption 2 guarantees that the share of inputs allocated to each production line does not depend on the identity of the input, or input proportionality, which is also required by De Loecker et al. (2016).[15] As a result, violating assumption 2 violates the input proportionality restrictions that are simply assumed by De Loecker et al. (2016), and therefore it is not more restrictive than their setting. More importantly, in practice most production data sets have only enough observations to feasibly estimate production function parameters at the level of an industry or sector, rather than at the product level. In these situations, the appropriateness of assumption 2 can be examined simply by checking the prevalence of firms producing in multiple industries.[16]

While the production technology can use an arbitrary number of inputs, I require that each firm use at least one static input, such as materials, that can be purchased from the market each time period according to some known price schedule; the next assumption states this more formally.

Assumption 3. $F(\cdot)$ takes as an input at least one element from the set $\mathbb{M}$, where $\mathbb{M}$ is a set of static inputs that can be purchased for one-period use from the market according to some known price schedule $W_{it}^M = \mathbf{W}^M(\sum_{j\in\mathbb{Y}_{it}} M_{it}^j, A_{it}^M)$ for each $M \in \mathbb{M}$, where $A_{it}^M$ is a vector of input price shifters.

Note that assumption 3 allows (but does not require) input buyers to have price-setting power, as in the monoposony/oligopsony literature (e.g., Robinson 1933; Card et al. 2018; Berger, Herkenhoff, and Mongey 2021).[17] If producers have no input market power, then one can simply write $W_{it}^M = \mathbf{W}^M(A_{it}^M)$; that is, wages do not depend on the total quantity of inputs purchased. Note that the core requirement embodied in assumption 3 is that there exists at least one input that does not have dynamic implications.

[15] The version of input proportionality assumed in De Loecker et al. (2016), which requires that value shares not vary with the identity of the input, or value proportionality, is slightly different from the form of input proportionality implied by assumption 2, which generates quantity proportionality, where input quantity shares do not depend on the identity of each input. Note, however, that if the input prices for input $X$ being used in product line $j$ by firm $i$ can be written as $W_{it}^{Xj} = W_t^X \times W_{it}^j$, which De Loecker et al. (2016) assume when using same input price control function for each input, then quantity proportionality implies value proportionality. More importantly, violations of quantity proportionality imply violations of value proportionality if input prices satisfy $W_{it}^{Xj} = W_t^X \times W_{it}^j$. To see this, suppose that a two-product firm has a higher capital-labor ratio in product line 1 than in product line 2, thereby violating quantity proportionality. Value proportionality will hold only if the ratio of capital to labor prices is lower in product line 1 than in product line 2. This is impossible, since $W_{it}^{K1}/W_{it}^{L1} = W_t^K/W_t^L = W_{it}^{K2}/W_{it}^{L2}$.

[16] In my empirical application, I also test for within-industry variation in the production function using single-product plants and cannot reject the null hypothesis of equal, identical production functions across 2-digit codes.

[17] See also Morlacco (2018), Brooks et al. (2021), and Rubens (2021) for recent work examining input market power using production function estimation.

Production may also use dynamic inputs, such as capital, which are accumulated through an investment process. Each dynamic input $K \in \mathbb{K}$, evolves over time according to a law of motion $K_{it} = l^K(K_{i,t-1}, I^K_{i,t-1}, I^K_{it})$, where $K_{it}$ is the stock of dynamic input $K$ used in firm $i$ at time $t$, and $I^K_{it}$ is firm $i$'s investment in dynamic input $K$ at time $t$. Note that I include both contemporaneous investment $(I^K_{it})$ and lagged investment $(I^K_{i,t-1})$ in the law of motion for dynamic inputs, so that the notation is general enough to encompass both predetermined inputs, in which case $l^K(K_{i,t-1}, I^K_{i,t-1}, I^K_{it}) = l^K(K_{i,t-1}, I^K_{i,t-1})$, and models of adjustment costs where dynamic inputs can adjust to contemporary shocks (e.g., Bond and Söderbom 2005), in which case $l^K(K_{i,t-1}, I^K_{i,t-1}, I^K_{it}) = l^K(K_{i,t-1}, I^K_{it})$.[18] Investment, as well as upkeep of the current stock of dynamic input $K$, costs the firm $d^K(K_{it}, I^K_{it})$. This setup allows dynamic inputs to face adjustment costs at the firm level. I require, however, that there not be any adjustment costs within a firm. In particular, I assume the following.

Assumption 4. All inputs can be costlessly transferred across production lines.

Note that assumption 4 is immediately satisfied for the definition of static inputs described by assumption 3. For dynamic inputs, assumption 4 means that each firm $i$ can reallocate their stock of dynamic inputs $K \in \mathbb{K}$ into any of their production lines at zero extra cost, with $K^j_{it}$ denoting the quantity of dynamic input $K \in \mathbb{K}$ going into production line $j \in \mathbb{Y}_{it}$. While this a relatively strong assumption, as it rules out unobserved adjustment costs of capital within a firm, it is necessary to allow for dynamic adjustment costs at the level of a firm. In particular, if there are unobserved dynamic adjustment costs for capital by product line, the optimal allocation of capital in each period will generally depend on the stock of capital used in each product line in the previous period. This adds another $|\mathbb{Y}_{it}| = J_{it}$ unobservables that would have to be pinned down to solve the input allocation problem, which is unlikely to be feasible, particularly if the magnitude of adjustment costs varies across firms. On the other hand, note that assumption 4 is entirely consistent with dynamic adjustment costs (e.g., time to build) for capital across firms, since costless transferability of capital means that the capital allocations are no longer a state variable in the firm's problem.[19]

[18] In my empirical application, I assume that capital is predetermined, while I allow labor to adjust to productivity shocks in the current period while still involving some dynamic implications due to adjustment (e.g., hiring and firing) costs.

[19] Gong and Sickles (2021) consider an important alternative to assumption 4, although this assumption weakens the degree to which one can allow for adjustment costs at the level of a firm. In particular, they assume that capital stock used at time $t$ in each product line $j$ is decided at time $t - 1$. However, the product-line-specific capital stock in this setting is bought at a constant marginal cost and fully depreciates each time period. The setting considered in this paper allows capital to follow an arbitrary accumulation process at the level of a firm, such as processes involving partial depreciation of capital over time, as is standard

Given these assumptions, it will be useful to distinguish between input allocations, which describe the allocation of inputs across uses within a firm conditional on the aggregate resources firm $i$ commands, and aggregate input vectors, which correspond to the total quantity of inputs used by the firm $i$ in period $t$. Formally, I denote input allocations by the input matrix $\mathbb{X}_{it}$, with typical element $(j, X)$ equal to $X_{it}^{j}$, which denotes the total quantity of input $X \in (\mathbb{K}, \mathbb{M})$ allocated to production line $j \in \mathbb{Y}_{it}$. On the other hand, I denote aggregate input vectors by $\vec{X}_{it} \equiv (\vec{K}_{it}, \vec{M}_{it})$, where $\vec{K}_{it}$ is the vector of dynamic input stocks owned by firm $i$ at time $t$ and $\vec{M}_{it}$ is the vector of total static inputs purchased by firm $i$.

I make the following assumption to ensure that input allocations are well defined.

Assumption 5. Aggregate inputs, $\vec{X}_{it}$, are completely attributable to production lines: $X_{it}^{j} = S_{it}^{jX} X_{it}$, where $S_{it}^{jX} \in [0, 1]$ and $\Sigma_{j=1}^{J_{it}} S_{it}^{jX} = 1$, for any input $X$.

Note that assumption 5 rules out public inputs, such as an input that can be used in more than one production process at once. This assumption may limit the role of economies of scope due to cost savings generated by public inputs, as described in Baumol, Panzar, and Willig (1982). While my baseline approach, which allows for general production technologies, cannot account for this, I show in section II.C.2 that this restriction can be relaxed for some types of public inputs if the technology is Cobb-Douglas. Moreover, assumption 5 is compatible with some economies of scope as long as they are embodied in TFP differences, rather than differences in the shape of the production technology, as was pointed out by De Loecker et al. (2016).

I also make the following restriction on the relationship between inputs and outputs.

Assumption 6. The input sets $(\mathbb{K}, \mathbb{M})$ do not contain any products produced within the same firm; that is, $\mathbb{Y}_{it} \not\subset (\mathbb{K}, \mathbb{M})$.

Assumption 6 does not allow plants to produce and sell products that are also used as inputs into one of their other outputs, such as a vertically integrated T-shirt producer who also produces and sells cotton fabric. The methods outlined in this paper do not easily generalize to this case, as the input allocation rules in this setting will critically vary with the input-output structure of each production process.

The next set of assumptions describes the industry structure. I assume that each product sold on the market faces a downward-sloping demand function, $Q_{it}^{j}(\vec{P}_t, \vec{\eta}_t)$, where $\vec{P}_t$ is the vector of prices charged for each product and $\vec{\eta}_t$ is the vector of product characteristics for each product. Note that the demand function depends on the entire vector of prices charged

in many other settings; e.g., Olley and Pakes (1996). See also sec. II.C.3 for an alternative approach that allows the researcher to relax this restriction for some inputs.

on the market, $\vec{P}_t$, allowing for fairly general patterns of cross-product substitutability. Letting $\vec{Q}_{it}(\vec{P}_t, \vec{\eta}_t)$ denote the vector of demand functions, I also require that the overall demand system satisfy the following restriction, described in more detail in Berry, Gandhi, and Haile (2013) and Berry and Haile (2014).

Assumption 7. The demand system $\vec{Q}_{it}(\vec{P}_t, \vec{\eta}_t)$ exhibits connected substitutes in prices.

Assumption 7 is a fairly weak restriction that is satisfied by most of the demand systems used in applied work. Roughly speaking, this restriction requires that all goods be weak substitutes for each other (demand for each good $j$ is nondecreasing in the price of all goods $k \neq j$) and that some goods are strict substitutes for one another (demand for good $j$ is strictly increasing in the price of some subset goods $m \neq j$).

Finally, I assume that firms optimally choose inputs and outputs; assumption 8 states this more formally.[20]

Assumption 8. Each firm chooses their aggregate input vectors $\vec{X}_{it}$, the allocation of each input across production lines $\mathbb{X}_{it}$, a vector of output prices $\vec{P}_{it}$, and a vector of investment levels $\vec{I}_{it}$ to maximize the present discounted value of their profits, given the laws of motion for the dynamic inputs, the aggregate demand system, their output sets $\mathbb{Y}_{it}$, the set of products produced by other firms $\mathbb{Y}_{-i,t} \equiv \Omega_t \setminus \mathbb{Y}_{it}$, the vector of firm-level TFP terms $\vec{\omega}_{it}$, the vector of firm-level input price shifters $\vec{A}_{it}$, the vector of prices charged by all other firm $\vec{P}_{-i,t}$, the vector of lagged stocks of dynamic inputs $\vec{K}_{i,t-1}$, and the market-level vector of product characteristics $\vec{\eta}_t$.

The Bellman equation associated with assumption 8 is

$$
\begin{aligned}
V_t(\chi_{it}) = & \max_{\vec{I}_{it}, \mathbb{X}_{it}, \vec{X}_{it}, \vec{P}_{it}} \sum_{j \in \mathbb{Y}_{it}} P^j_{it} Q^j_{it}(\vec{P}_t, \vec{\eta}_t) - \sum_{M \in \mathbb{M}} \sum_{j \in \mathbb{Y}_{it}} \mathbf{W}^M \left( \sum_{j \in \mathbb{Y}_{it}} M^j_{it}, A^M_{it} \right) M^j_{it} \\
& - \sum_{K \in \mathbb{K}} d^K(K_{it}, I^K_{it}) + \beta \mathbb{E}\{ V_{t+1}(\chi_{i,t+1}) | \chi_{it} \},
\end{aligned}
$$

subject to (2)

$$
\begin{aligned}
& \exp(\omega^j_{it}) F(\vec{X}^j_{it}) \geq Q^j_{it}(\vec{P}_t, \vec{\eta}_t) \quad \forall\, j \in \mathbb{Y}_{it}, \\
& \sum_{j \in \mathbb{Y}_{it}} X^j_{it} = X_{it} \quad \forall\, X \in \mathbb{K}, \\
& K_{it} = l^K(K_{i,t-1}, I^K_{i,t-1}, I^K_{it}) \quad \forall\, K \in \mathbb{K},
\end{aligned}
$$

where $\chi_{it} \equiv (\mathbb{Y}_{it}, \mathbb{Y}_{-i,t}, \vec{\omega}_{it}, \vec{A}_{it}, \vec{P}_{-i,t}, \vec{K}_{i,t-1}, \vec{\eta}_t)$.

[20] While this formulation of the firm's problem assumes Bertrand competition, i.e., Nash equilibrium with price setting, I show in app. B that one can obtain the same results under Cournot competition, or Nash equilibrium with quantity setting.

While equation (2) is a fairly standard profit maximization problem, there are some features of the problem that are worth emphasizing. First, note that this formulation of the firm's problem takes each firm's output sets, $\mathbb{Y}_{it}$, as given. This means that the model I use to uncover the unobserved input allocations does not discipline how firms choose their product sets over time. In particular, while equation (2) implies that a firm optimally allocates inputs across product lines for a given output set $\mathbb{Y}_{it}$ to maximize the present discounted value of their profits, this problem does not require that output sets be chosen optimally in any sense. Rather, the key requirement is that firms choose their output sets in a way that does not interact with the input choice problem.[21] This means that questions related to how firms choose their product sets, such as which products they wish to drop, are testable in this framework.[22] I return to this problem in section V.C, where I examine which products are most likely to be dropped by an Indian plant.

Second, the firm conditions on the full vector of TFP terms, $\vec{\omega}_{it}$, when choosing their input allocations. This information requirement is necessary for input allocations to be fully revealed by demand-side information. While the empirical framework considered below will allow these productivity shocks to be unobserved by the econometrician, necessitating a way to deal with endogeneity bias, note that this assumption rules productivity shocks that are unobserved by both the firm and the econometrician, as is common in the "proxy-variable" literature (Olley and Pakes 1996; Levinsohn and Petrin 2003; Ackerberg, Caves, and Frazer 2015; Gandhi, Navarro, and Rivers 2020).[23] Finally, even though there are dynamic adjustment costs due to the inclusion of the $d^K(\cdot)$ functions,

[21] For example, a model where firms choose their output sets optimally, as in Mayer, Melitz, and Ottaviano (2014), can be accommodated as long as the fixed costs of producing each good are paid in nominal (dollar or rupee) units, rather than input units (e.g., labor units). In the latter case, the firm's input allocation problem would have be modified to account for these extra fixed costs that interact with the firm's labor choice, which is inconsistent with assumption 8.

[22] Relatedly, note that I have not fully specified the law of motion for all state variables in $\chi_{it}$. This is intentional, as the subsequent theorem on recovering the allocation of inputs does not depend on the law of motion for all state variables. Rather, theorem 1 will be valid for any process governing the evolution of $\chi_{it}$ as long as each plant's per-period payoffs take the form described by eq. (2). This includes flexible specifications of the law of motion for productivity, which can depend on endogenous firm choices such as R&D, as in Doraszelski and Jaumandreu (2013).

[23] The core reason for this is that these extra shocks generates an extra set of $|\mathbb{Y}_{it}|$ unknowns. Each of these shocks would end up being arguments in the input allocation rule derived below, which would severely complicate estimation. That said, this assumption is not arbitrary, nor is it without economic content. Rather, the key assumption here is that producers know their productivity levels; i.e., I am considering a full-information rather than an imperfect-information model. This is a standard approach in empirical industrial organization that focuses on modeling price competition from the demand side (e.g., Berry 1994; Berry, Levinsohn, and Pakes 1995), which also forms the basis of the input allocation rule derived below.

since demand depends entirely on a vector of current prices, there will be no dynamic pricing in this model, except insofar as the set of dynamic inputs affect marginal costs.[24]

### B. Identifying Unobserved Inputs Using Demand-Side Data

I now show that assumptions 1–8 imply a mapping from observable prices and quantities to the unobserved firm-product input allocations. First, note that since dynamic inputs are costlessly transferable across uses within the firm (assumption 3) and $F(\cdot)$ is strictly increasing in all of its inputs (assumption 1), each firm will choose an input allocation that minimizes total static input costs conditional on its stock of firm-level dynamic inputs $\vec{K}_{it}$ and some desired set of output levels $\vec{Y}_{it}$. More formally, any solution to equation (2) will involve an input allocation $\mathbb{X}_{it}$ that minimizes static input costs, subject to some desired output levels $Y^j_{it}$ for each $j \in Y_{it}$, and a given stock of dynamic inputs. The Lagrangian for this conditional cost minimization problem is given by[25]

$$L = -\sum_{M\in\mathbb{M}}\sum_{j\in\mathbb{Y}_{it}}\mathbf{W}^M\left(\sum_{j\in\mathbb{Y}_{it}}M^j_{it}, A^M_{it}\right)M^j_{it} + \sum_{j\in\mathbb{Y}_{it}}\lambda^j_{it}\left(\exp\left(\omega^j_{it}\right)F(\vec{X}^j_{it}) - Y^j_{it}\right) + \sum_{K\in\mathbb{K}}\varsigma^K_{it}\left(K_{it} - \sum_{j\in\mathbb{Y}_{it}}K^j_{it}\right), \tag{3}$$

where $\lambda^j_{it}$ is the Lagrangian multiplier for the production constraint and $\varsigma^K_{it}$ is the Lagrangian multiplier for resource constraint for dynamic input $K \in \mathbb{K}$.

Letting $\nu^X_{it} = W^X_{it} + (\partial\mathbf{W}^X(\Sigma_{j\in\mathbb{Y}_{it}}X^j_{it}, A^X_{it})/\partial(\Sigma_{j\in\mathbb{Y}_{it}}X^j_{it}))\Sigma_{j\in\mathbb{Y}_{it}}X^j_{it}$ if $X \in \mathbb{M}$ and letting $\nu^X_{it} = \varsigma^X_{it}$ if $X \in \mathbb{K}$, the first-order necessary condition for any input $X^j_{it}$ can be written as

$$-\nu^X_{it} + \lambda^j_{it}\exp(\omega^j_{it})\frac{\partial F(\vec{X}^j_{it})}{\partial X} = 0. \tag{4}$$

To obtain a simple input allocation formula that depends only on demand-side information, I make use of the following lemma.

Lemma 1. If assumptions 1–6 hold, then there exists a solution to the firm's conditional cost minimization problem satisfying $X^j_{it} = S^j_{it}X_{it}\ \forall X \in (\mathbb{K}, \mathbb{M})$, where $S^j_{it} \in [0, 1]$ and $\Sigma_{j\in\mathbb{Y}_{it}}S^j_{it} = 1$.

*Proof.* See appendix A.

[24] Put differently, the pricing problem is entirely static, conditional on the set of dynamic inputs chosen at time $t$.

[25] A formal statement of the conditional cost minimization problem is stated in app. A.

This result is a straightforward implication of homogeneous production technologies. Since homogeneous production technologies generate isoquants that have constant slopes along any ray from the origin, this means that a cost-minimizing firm will choose constant input ratios across production lines within the firm; that is, $X_{it}^j/Z_{it}^j = X_{it}^k/Z_{it}^k$ for each $X, Z \in (\mathbb{K}, \mathbb{M})$ and each $j, k \in \mathbb{Y}_{it}$. This immediately implies that the input shares within a particular production line $j$ do not depend on the identity of the input, and therefore $X_{it}^j/X_{it} = S_{it}^j \ \ \forall\ X \in (\mathbb{K}, \mathbb{M})$.

As a result, assumptions 1–6 mean that the vector of inputs allocated to each production line can be written as $\vec{X}_{it}^j = S_{it}^j \vec{X}_{it}$. Substituting this into equation (4) and then using the fact that all of the partial derivatives of a homogeneous of degree $\phi > 0$ function are homogeneous of degree $\phi - 1$, one obtains

$$\nu_{it}^X = \lambda_{it}^j \exp(\omega_{it}^j)\left(S_{it}^j\right)^{\phi-1} \frac{\partial F(\vec{X}_{it})}{\partial X}. \tag{5}$$

Divide this expression by $Y_{it}^j = \exp(\omega_{it}^j) F(S_{it}^j \vec{X}_{it}) = \exp(\omega_{it}^j)(S_{it}^j)^\phi F(\vec{X}_{it})$, which yields, after some minor manipulations,

$$S_{it}^j = \frac{\partial F(\vec{X}_{it})}{\partial X} \frac{\lambda_{it}^j Y_{it}^j}{F(\vec{X}_{it}) \nu_{it}^X}. \tag{6}$$

One can then sum equation (6) over all $j \in \mathbb{Y}_{it}$, yielding $(\partial F(\vec{X}_{it})/\partial X)([\sum_{j\in\mathbb{Y}_{it}} \lambda_{it}^j Y_{it}^j]/[F(\vec{X}_{it})\nu_{it}^X])$. Dividing equation (6) by this expression yields

$$S_{it}^j = \frac{\lambda_{it}^j Y_{it}^j}{\sum_{k\in\mathbb{Y}_{it}} \lambda_{it}^k Y_{it}^k} = \frac{\mathrm{MC}_{it}^j Y_{it}^j}{\sum_{k\in\mathbb{Y}_{it}} \mathrm{MC}_{it}^k Y_{it}^k}, \tag{7}$$

where the second equality follows from the envelope theorem $\lambda_{it}^j = \partial C(\vec{K}_{it}, \vec{Y}_{it}, \vec{\omega}_{it}, \vec{A}_{it})/\partial Y_{it}^j \equiv \mathrm{MC}_{it}^j$, where $C(\vec{K}_{it}, \vec{Y}_{it}, \vec{\omega}_{it}, \vec{A}_{it})$ is the cost function for static inputs, conditional on the level of dynamic inputs $\vec{K}_{it}$ and a desired level of output $\vec{Y}_{it}$.[26]

Equation (7) provides a simple input allocation rule, where the unobserved input allocations are given by output times marginal cost shares. While marginal costs are almost always unobservable in any firm-level data set, the input allocation rule becomes empirically useful after one notes that many models of interfirm competition imply a direct mapping from observable demand-side variables to unobservable product-level conditional marginal costs. To see this, note that once the input allocation problem has been solved for any potential level of dynamic inputs $\vec{K}_{it}$ and desired output levels $\vec{Y}_{it}$, one can determine the static conditional cost function and substitute this into equation (2), yielding the simplified firm's problem:

[26] More formally, $C(\vec{K}_{it}, \vec{Y}_{it}, \vec{\omega}_{it}, \vec{A}_{it})$ is the objective function associated with the solution to problem (CM) in app. A.

$$V_t(\chi_{it}) = \max_{\vec{P}_{it}, \vec{I}_{it}, \vec{K}_{it}} \sum_{j \in \mathbb{Y}_{it}} P_{it}^j Q_{it}^j \left(\vec{P}_t, \vec{\eta}_t\right) - C\left(\vec{K}_{it}, \vec{Q}_{it}\left(\vec{P}_t, \vec{\eta}_t\right), \vec{\omega}_{it}, \vec{A}_{it}\right)$$
$$- \sum_{K \in \mathbb{K}} d^K(K_{it}, I_{it}^K) + \beta \mathbb{E}\{V_{t+1}(\chi_{i,t+1}) | \chi_{it}\}, \tag{8}$$

subject to

$$K_{it} = l^K(K_{i,t-1}, I_{i,t-1}^K, I_{it}^K) \quad \forall\, K \in \mathbb{K},$$

where I have used the fact that quantity produced will equal quantity sold in equilibrium, that is, $Y_{it}^j = Q_{it}^j(\vec{P}_t, \vec{\eta}_t)$.

Taking the first-order condition for any $P_{it}^j$ in equation (8) yields

$$Q_{it}^j + \sum_{k \in \mathbb{Y}_{it}} \frac{\partial Q_{it}^k}{\partial P_{it}^j} (P_{it}^k - \mathrm{MC}_{it}^k) = 0, \tag{9}$$

where the $\partial Q_{it}^k / \partial P_{it}^j$ terms for $k \neq j$ are included to account for within-firm cannibalization effects, that is, the effect of pricing in firm $j$ on other product lines.

One can then stack the $J_t \equiv |\Omega_t|$ first-order conditions defined by equation (9), which, in matrix notation, defines the following system of equations:

$$\vec{Q}_t + \Delta_t \left(\vec{P}_t - \overrightarrow{\mathrm{MC}}_t\right) = 0, \tag{10}$$

where $\Delta_t = \mathbb{O}_t \circ \partial_t$, with $\partial_t$ corresponding to a $J_t \times J_t$ matrix of demand derivatives, with typical element $(j, k)$ equal to $\partial Q_{it}^j / \partial P_{mt}^k$ and $\mathbb{O}_t$ being the ownership matrix, with element $(j, k)$ equal to 1 if products $j$ and $k$ are both produced by the same firm $(i = m)$ and 0 otherwise.

Note that one can use equation (10) to solve for the equilibrium marginal costs as a function of quantities produced, prices, and demand derivatives, by premultiplying by $\Delta_t^{-1}$, yielding

$$\overrightarrow{\mathrm{MC}}_t = g\left(\vec{Q}_t, \vec{P}_t, \partial_t, \mathbb{O}_t\right) = \Delta_t^{-1} \vec{Q}_t + \vec{P}_t. \tag{11}$$

For the marginal cost inversion described by equation (10) to be valid, $\Delta_t$ must be invertible. Berry and Haile (2014) show that a sufficient condition for $\Delta_t$ to be invertible is that assumption 7 holds, a relatively weak restriction that is satisfied by many of the demand systems used in applied work.[27]

[27] While assumption 7 is a sufficient condition for $\Delta_t^{-1}$ to exist, it is not necessary. In particular, demand systems where some goods are complements, which will tend to violate assumption 7, will often still generate a mapping from prices to marginal costs. This means that eq. (7) can still be used to uncover unobserved input allocations in many settings with complementary goods. See Song, Nicholson, and Lucarelli (2017), Thomassen et al. (2017), Bokhari and Mariuzzo (2018), Ershov et al. (2021), Iaria and Wang (2021), and Wang (2021) for examples of demand systems with complementarity that still generate marginal cost inversions.

An important point worth emphasizing is that as long as assumption 7 holds, $\Delta_t$ is invertible for any ownership matrix $\mathbb{O}_t$. This means that this marginal cost inversion is also possible under imperfect competition with some forms of collusion, as long as the econometrician knows which firms are colluding and which are not. In particular, as noted by Nevo (1998), collusion between firms can be thought of as joint-profit maximization, that is, legally separate firms choosing their prices as if they were one multiproduct firm.[28] As long as the set of colluding firms fully internalize their pricing decisions across products, then this simply corresponds to an alternative ownership matrix, $\mathbb{O}'_t$, to that observed in the data. For example, full market collusion simply corresponds to the case where $\mathbb{O}'_t$ is a $J_t \times J_t$ matrix of 1s, implying that all prices are chosen to internalize all possible cannibalization effects.

Combining equations (10) and (7) yields the following theorem.

THEOREM 1. As long as the cost-minimizing input allocation is unique, then assumptions 1–8 imply that the share of any input $X \in (\mathbb{K}, \mathbb{M})$ going into production line $j \in \mathbb{Y}_{it}$ satisfies $S^j_{it} = [g^j_{it}(\vec{Q}_t, \vec{P}_t, \partial_t, \mathbb{O}_t) Y^j_{it}]/[\Sigma_{k \in \mathbb{Y}_{it}} g^k_{it}(\vec{Q}_t, \vec{P}_t, \partial_t, \mathbb{O}_t) Y^k_{it}]$, where each $g^j_{it}(\cdot)$ is a known function of prices, quantities, demand derivatives, and the ownership matrix.

*Proof.* The proof follows from the discussion in text and equations (7) and (11), which will always hold if there is a unique solution to the firm's input allocation problem, as per lemma 1. Note that uniqueness of the input allocation problem is guaranteed for standard homogenous production functions such as Cobb-Douglas and CES, as it is straightforward to verify that equation (4) implies that $X^j_{it} = S^j_{it} X_{it} \;\; \forall\, j \in \mathbb{Y}_{it}$. QED

### C. *Discussion and Extensions*

The key significance of theorem 1 is that under the maintained assumptions, demand-side information (output prices and quantities) can be used to infer the allocations of inputs across production lines, if demand derivatives and ownership structures are known. This immediately suggests a simple strategy for estimating within-firm TFP dispersion. First, use price and quantity data to estimate the shape of the demand function. Having estimated the demand function, the researcher can then obtain estimates of demand derivatives at the product level and then apply theorem 1 to obtain the allocations of inputs across production lines within a firm. This provides the researcher with estimates of input use at the firm-product level, which can then be used to estimate a firm-product level of production function.

[28] Note that this form of collusion requires that firms cooperate only in output markets, i.e., through pricing first-order conditions, not through input markets; i.e., firms cannot engage in labor or capital sharing.

Note that since assumptions 1–8 do not require that the researcher specify exactly how the various unobservables at the firm level, $(\vec{\omega}_{it}, \vec{\eta}_{it})$, evolve over time, this general recipe is flexible enough to allow for a wide variety of estimation strategies for both the demand and production functions. In particular, researchers are free to formally model the distribution of $(\vec{\omega}_{it}, \vec{\eta}_{it})$ in various ways to identify the production and demand parameters, through either maximum likelihood or instrument-based GMM (generalized method of moments) methods.[29]

Under some further restrictions, input allocations can be recovered without having to estimate parameters governing the demand system. In particular, the following corollary follows immediately from theorem 1.

COROLLARY 1. If assumptions 1–8 hold, the cost-minimizing input allocation is unique, and a firm charges the same markup $\mu^j_{it} \equiv P^j_{it}/\mathrm{MC}^j_{it} = \mu_{it}$ on all products, then input allocations are revealed by revenue shares.

*Proof.* By assumption, $\mathrm{MC}^j_{it} = P^j_{it}/\mu_{it}$. Applying this to the input allocation rule (7) yields

$$S^j_{it} = \frac{\left(P^j_{it}/\mu_{it}\right) Y^j_{it}}{\sum_{k\in\mathbb{Y}_{it}} (P^k_{it}/\mu_{it}) Y^k_{it}} = \frac{P^j_{it} Y^j_{it}}{\sum_{k\in\mathbb{Y}_{it}} P^k_{it} Y^k_{it}}.$$

QED

Note that the assumptions of corollary 1 are true for two well-known cases. First, if output markets are perfectly competitive, firms do not charge a markup and therefore $\mu^j_{it} = 1 \ \forall\ j$. Moreover, if variety-level demand for each product follows the same CES demand function, then a firm will charge the same markup on all products, as in Feenstra and Ma (2007). Note that this result does not require monopolistic competition or a continuum of firms; rather, even if firms engage in Nash-Bertrand oligopoly pricing, which implies different markups across firms, each firm will choose the same markup across all products produced within the same firm, as a result of their internalization of cannibalization effects across their production lines.[30]

Whenever firms charge different markups across production lines, within-plant price variation also contains information on differences in market power across production lines, which has to be accounted for

[29] Note, however, that the estimation approach must be consistent with firms having full information on $\vec{\omega}_{it}$ when choosing inputs. This rules out proxy-variable-based production function approaches, as I noted when discussing assumption 8. However, this approach still allows for a number of alternative strategies based on various instrumental variables, dynamic panel-type approaches, or even cost-share-based approaches to identifying the production function.

[30] It is worth noting that variety-level CES is actually a slightly stronger requirement than is necessary for this result; see Hottman, Redding, and Weinstein (2016), who consider more general nesting structures which generate constant within-firm markups.

when estimating unobserved input allocations. In such models, one must first estimate the approximate shape of the demand system to recover product-specific marginal costs for the input allocation rule (7).

### 1. General Functional Forms

To examine the role the functional form restrictions in assumptions 1 and 2 play, suppose first that $Y_{it}^{j} = \exp(\omega_{it}^{j})F^{j}(\vec{X}_{it}^{j})$, where $F^{j}(\cdot)$ is still differentiable but is neither homogeneous nor quasi-concave.[31] One can then modify equation (4) to allow the production function to vary with $j$ and then divide this expression by $Y_{it}^{j} = \exp(\omega_{it}^{j})F^{j}(\vec{X}_{it}^{j})$ and rearrange, yielding

$$X_{it}^{j} = \frac{\theta_{X}^{j}\left(\vec{X}_{it}^{j}\right)\lambda_{it}^{j}Y_{it}^{j}}{\nu_{it}^{X}}, \tag{12}$$

where $\theta_{X}^{j}(\vec{X}_{it}^{j}) \equiv (\partial F^{j}(\vec{X}_{it}^{j})/\partial X)(X_{it}^{j}/F^{j}(\vec{X}_{it}^{j}))$ is the *output elasticity* for input $X \in (\mathbb{K}, \mathbb{M})$. Summing equation (12) over all $j \in \mathbb{Y}_{it}$ and dividing equation (12) by this new expression yields, after the envelope theorem is applied,

$$X_{it}^{j} = \frac{\theta_{X}^{j}\left(\vec{X}_{it}^{j}\right)\mathrm{MC}_{it}^{j}Y_{it}^{j}}{\sum_{k\in\mathbb{Y}_{it}}\theta_{X}^{k}\left(\vec{X}_{it}^{k}\right)\mathrm{MC}_{it}^{k}Y_{it}^{k}}\sum_{j\in\mathbb{Y}_{it}}X_{it}^{j}. \tag{13}$$

Note that equilibrium input shares, if they exist, will form a fixed point of the system of equations described by equation (13). More importantly, this mapping does not depend on the unobservable TFP terms $\omega_{it}^{j}$, implying that input shares may be separately identified from TFP if $F^{j}$ is known, even if assumptions 1 and 2 do not hold.[32]

While assumptions 1 and 2 are unnecessary if $F^{j}(\cdot)$ is known, in general the production function must also be estimated. This can lead to difficulties in applying equation (13) in practice, since, if one wishes to use multiproduct firms to estimate the production function, one would have to already know the production technology to determine their input allocations. To deal with this, one must either use only single-product firms to estimate the technology and then solve for the input allocations using equation (13), as in De Loecker et al. (2016), or use an estimator that

[31] I consider the case of nondifferentiable production technologies in app. C.

[32] Note that a fixed point is guaranteed to exist in eq. (13) for the case of heterogeneous Cobb-Douglas production functions; i.e., $F^{j}(\vec{X}_{it}^{j}) = \prod_{X\in(\mathbb{K},\mathbb{M})}(X_{it}^{j})^{\beta_{X}^{j}}$. Since output elasticities are constant for each input with Cobb-Douglas, i.e., $\theta_{X}^{j}(\vec{X}_{it}^{j}) = \beta_{X}^{j}$, eq. (13) simplifies to

$$X_{it}^{j} = \frac{\beta_{X}^{j}g_{it}^{j}\left(\vec{Q}_{t}, \vec{P}_{t}, \partial_{t}, \mathbb{O}_{t}\right)Y_{it}^{j}}{\sum_{k\in\mathbb{Y}_{it}}\beta_{X}^{k}g_{it}^{k}\left(\vec{Q}_{t}, \vec{P}_{t}, \partial_{t}, \mathbb{O}_{t}\right)Y_{it}^{k}}\sum_{j\in\mathbb{Y}_{it}}X_{it}^{j}. \tag{14}$$

estimates input allocations and production function parameters simultaneously, as in Valmari (2016) or Gong and Sickles (2021). Both approaches have limitations. The former approach, using single-product firms only, may be subject to selection bias absent a correction that models the process governing selection into producing multiple products, is potentially less efficient than using all observations in the estimation algorithm, and cannot allow single- and multiproduct plants to have fundamentally different technologies.

On the other hand, the latter approach, which estimates input allocations and production function parameters simultaneously, may suffer from identification problems, as it is no longer clear what sources of variation can be used to separately identify output elasticities from input allocations. For example, following a GMM-based approach as in Valmari (2016) will require formulating moment conditions that are inherently nonlinear in the parameters to be estimated. Determining primitive conditions under which these moments would uniquely pin down the parameters of interest is quite difficult, as described by Newey and McFadden (1994), who note that, in practice, identification is often simply assumed in nonlinear GMM settings. Moreover, since input shares depend on parameters in other production lines, this means that estimation would have to be done simultaneously for all products belonging to a connected set, that is, the set of firm-level product sets that overlap in at least one product. This can generate a severe dimensionality problem, as the set of nonlinear parameters that have to be estimated simultaneously grows multiplicatively in the number of products in a connected set and the number of inputs, which becomes intractable quite quickly.[33] On the other hand, invoking assumption 2 can allow the researcher to abstract from these extra complications, obtaining an estimate of input use that does not depend on unknown production function parameters, allowing the researcher to invoke standard arguments for identification of the production function parameters used in past research on productivity in a firm-level setting.

### 2. Public or Joint Inputs

It is also worth considering whether assumption 5 is appropriate, that is, whether all inputs are attributable to each production line, so that $X_{it}^{j} = S_{it}^{jX} X_{it}$, where $\sum_{j\in\mathbb{Y}_{it}} S_{it}^{jX} = 1 \;\; \forall \; X$. While this assumption is made by basically all the literature that allocates inputs to deal with multiproduct

[33] For example, if we consider a connected set of around half the available 2-digit HS (harmonized system) codes in a production setting with three inputs (labor, capital, and materials), this generates a problem with just under 150 nonlinear parameters that have to be estimated simultaneously.

firms, it is restrictive in the sense that it rules out public or joint inputs, which may affect many production lines simultaneously, as in Baumol, Panzar, and Willig (1982), generating economies of scope. I now show that a variant of the identification result carries through for the special case of Cobb-Douglas technologies with public or joint inputs.

Suppose that the use of some input $X_{it}$ within a firm can be divided into a public, or common, component, $X_{it}^{\mathrm{C}}$, and a rivalrous component, $X_{it}^{\mathrm{R}}$. Rivalrous inputs $X_{it}^{\mathrm{R}}$ can be allocated to only a single production line, as in assumption 5, so that $X_{it}^{j\mathrm{R}} = S_{it}^{jX\mathrm{R}} X_{it}^{\mathrm{R}}$, with $\Sigma_{j\in\mathbb{Y}_{it}} S_{it}^{jX\mathrm{R}} = 1$. Common inputs are allocated to every production line automatically. Hence, the quantity of effective inputs allocated to each production lines is given by $X_{it}^{j} = X_{it}^{\mathrm{C}} + X_{it}^{j\mathrm{R}}$.

Suppose further that the public component of $X_{it}$ is a constant fraction of total inputs owned by the firm—that is, $X_{it}^{\mathrm{C}} = \kappa^X X_{it}$ and $X_{it}^{\mathrm{R}} = (1 - \kappa^X)X_{it}$—and the production technology is Cobb-Douglas, so that $F(\vec{X}_{it}^{j}) = \prod_{X\in(\mathbb{K},\mathbb{M})}(X_{it}^{j})^{\beta_X}$. Firms will then allocate their private inputs $X_{it}^{\mathrm{R}}$ across production lines to minimize static production costs conditional on aggregate dynamic inputs. It is straightforward to show, following the derivation in the text, that this slight modification implies that effective inputs will satisfy the following:

$$\begin{aligned} X_{it}^{j} &= \frac{g_{it}^{j}(\cdot)\,Y_{it}^{j}}{\sum_{k\in\mathbb{Y}_{it}} g_{it}^{k}(\cdot)\,Y_{it}^{k}} \sum_{j\in\mathbb{Y}_{it}} X_{it}^{j} \\ &= \frac{g_{it}^{j}(\cdot)\,Y_{it}^{j}}{\sum_{k\in\mathbb{Y}_{it}} g_{it}^{k}(\cdot)\,Y_{it}^{k}} [1 + (J_{it} - 1)\kappa^X] X_{it}, \end{aligned} \tag{15}$$

where $J_{it} \equiv |\mathbb{Y}_{it}|$ is the number of products produced by firm $i$.[34]

While the fractions of public inputs, $\kappa^X$, are unobservable, and hence the level of input usage will not be identified in this framework, note that under Cobb-Douglas, equation (15) implies that the unobservable component of these input allocations, $(J_{it} - 1)\kappa^X$, is observationally equivalent to a TFP shifter received by multiproduct firms. To see this, substitute equation (15) into the production function, yielding

$$Y_{it}^{j} = \exp(\omega_{it}^{j} + \mathrm{SC}_{it}) \prod_{X\in(\mathbb{K},\mathbb{M})} \left(\hat{X}_{it}^{j}\right)^{\beta_X}, \tag{16}$$

where $\hat{X}_{it}^{j}$ is the level of input usage obtained from theorem 1, and $\mathrm{SC}_{it}$ is the economies-of-scope shifter, given by $\mathrm{SC}_{it} \equiv \Sigma_{X\in(\mathbb{K},\mathbb{M})}\beta_X \ln(1 + (J_{it} - 1)\kappa^X)$.

[34] Implicitly, this derivation assumes that rivalrous inputs are chosen to be nonnegative, and this constraint is not binding; i.e., the firm does not wish to move common inputs from one production line to another.

Hence, when public inputs take this form, they are observationally equivalent to TFP shifters that depend on the number of products produced by a firm. As a result, one can deal with the complications introduced by public inputs by controlling for the number of products in the production function estimation routine.[35]

Note, however, that under these assumptions the production function residual, $\hat{\omega}_{it}^{j}$, will be composed of both a "pure" TFP component and an economy-of-scope shifter, $SC_{it}$; that is, $\hat{\omega}_{it}^{j} = \omega_{it}^{j} + SC_{it}$. This means that multiproduct firms may have higher measured TFP either because of selection—that is, firms with high $\omega_{it}^{j}$ terms are more likely to produce many products, as emphasized by Bernard, Redding, and Schott (2010) and Mayer, Melitz, and Ottaviano (2014)—or because of economies of scope—that is, multiproduct production scales up input effectiveness as a result of public inputs. Lacking a clean source of variation to distinguish between these two stories, I do not attempt to distinguish economies of scope from selection in my empirical application, although I do use these results to inform my identification strategy, which controls for scope effects (which could occur via selection or public inputs) through the number of product dummies.

### 3. Violating Costless Transferability

Assumption 4, which requires that inputs be perfectly transferable across product lines, is unlikely to hold if some inputs face dynamic adjustment costs that are product specific. For example, some product lines may require investment in specialized machinery, or workers may be required to learn a particular set of specialized skills to produce a particular product. In such settings, the lagged allocation of capital or labor will tend to affect the future allocations of these inputs, complicating the task of uncovering the unobserved input allocations by generating further unobservable state variables.

Note, however, that for some questions a researcher may be content to simply determine the allocation of static inputs, leaving the allocation of these dynamic inputs to form part of the productivity residual. In particular, if some dynamic inputs are highly specialized, so that transferring these inputs across products is largely infeasible, it may make more sense to use productivity measures that include the effect of these specialized inputs. In this setting, the unobserved dynamic input allocations for specialized inputs will simply act as components of a firm's product-line-specific technology that they can invest in and improve over time. More

[35] More generally, one could allow the fraction of public inputs, $\kappa^{X}$, to depend on the sets of outputs produced by the firm, in which case one would have to include product set fixed effects.

formally, suppose that $Y_{it}^{j} = \exp(\omega_{it}^{j})\prod_{X\in(\mathbb{K},\mathbb{M})}(X_{it}^{j})^{\beta_X}$, with $0 < \beta_X < 1 \ \forall \ X$. Further suppose that there exists some subset of specialized, dynamic inputs $\mathbb{S} \subset \mathbb{K}$ that face product-line-specific adjustment costs, so that assumption 4 does not hold. One can then define product-line-specific productivity inclusive of the specialized dynamic inputs as $\omega_{it}^{j\mathbb{S}} = \omega_{it}^{j} + \Sigma_{X\in\mathbb{S}}\beta^{X}\ln(X_{it}^{j})$. Substituting this expression into the production function yields the modified technology:

$$Y_{it}^{j} = \exp(\omega_{it}^{j\mathbb{S}}) \prod_{X\in(\mathbb{M},\mathbb{K}\setminus\mathbb{S})} (X_{it}^{j})^{\beta_X}. \tag{17}$$

Note that the production function $\prod_{X\in(\mathbb{M},\mathbb{K}\setminus\mathbb{S})}(X_{it}^{j})^{\beta_X}$ satisfies assumptions 1 and 2. As a result, one can immediately apply theorem 1 and uncover the remaining input allocations, using demand-side data. As long as the remaining production function parameters can then be identified, the researcher will then be able to estimate $\omega_{it}^{j\mathbb{S}}$, which can be interpreted as product-line-specific productivity that includes the effect of specialized investments in hard-to-transfer machinery or workers.[36]

While this provides one alternative path forward for researchers who do not find assumption 4 appropriate, note that allowing unobserved inputs to form part of the production function residual generates further identification challenges. Specifically, since the firm directly controls the quantity of inputs used in a given production line by accumulating dynamic inputs, the residual itself is an endogenous variable, which will tend to react to the same shocks as the remaining inputs in the production function. I consider this problem in more detail in appendix D and show that if one wishes to use all firms in the production function estimation algorithm, one must use identification strategies different from those commonly used in the production function literature.

## III. Data

The primary data set used in this paper comes from the 2000–2007 Indian ASI, provided by the Indian Ministry of Statistics.[37] The sample frame for the survey is all manufacturing plants in India that employ more than 10 workers. Plants with more than 100 workers ("census" plants) are

[36] This result relies on the Cobb-Douglas functional form. If the production function were to be homogenous of degree $\phi > 0$ but not Cobb-Douglas, then one cannot simply treat unobserved dynamic inputs as analogous to a Hicks-neutral productivity shifter, since fixing the quantity of one input generates a technology that is no longer homogeneous in the remaining inputs.

[37] Years in the ASI are recorded from April 1 to March 31. While the Ministry of Statistics refers to years by the end year, I refer to years by the start year, since the majority of production time takes place in that year. I follow this convention when matching the data to other data sets (e.g., trade data).

surveyed every year, while smaller plants are randomly sampled each year. The data contain consistent plant-level identifiers across years, allowing me to construct plant-level panels.[38] As described in Martin, Nataraj, and Harrison (2017), the panel data are of fairly high quality and cover a much larger subset of Indian producers than other comparable data sets for the country, such as Prowess.

I focus on a single industry in my empirical application: machinery, equipment, and parts, the details of which are described in appendix E. I focus on this industry for two reasons. First, I wish to focus on an industry where it is appropriate to think of inputs as being directly allocated to different production lines, as per assumption 5. Unfortunately, this is not the case for two of India's largest industries, sugar and textiles, as many plants in these industries produce groups of products that are by-products. In particular, most refined-sugar producers also produce molasses, which is generated by the refining process, while many cotton producers also produce cotton waste, a by-product of cotton production that is often resold for further production purposes. Unfortunately, this means that theorem 1 is unlikely to apply for these plants. Machinery manufacturing, on the other hand, does not commonly generate by-products that are also sold on the market by the same plant, making theorem 1 more likely to apply.

Second, to satisfy assumption 6, I wish to focus on industries where the plant-level output sets are not driven by vertical integration concerns. If some of the outputs produced and sold by a plant are also inputs in a vertically integrated production line, then theorem 1 will not apply. Unfortunately, this appears to be the case for many plants. After constructing an input-output table, the details of which I describe in appendix E1, I find that large portions of the observed revenue in industries such as steel, food, and synthetic textiles are produced by multiproduct plants that are potentially vertically integrated, in the sense that one of their outputs is likely an input for another one of their production lines. While there are some plants in the machinery, equipment, and parts industry that produce product sets that may indicate vertical integration, as I document in appendix E1, this is a smaller problem than in other industries.

The key variables used in this study are described in tables 1 and 2. Each plant lists the revenues and quantities produced and sold for up to 10 different products produced within the plant.[39] Associated with

[38] Since the unit of observation is a plant-item, rather than a firm-item, I consider separate plants as separate firms in my empirical analysis. For plants to approximate firms, I implicitly assume that the firms that operate multiple plants decentralize the pricing decision to local managers or that within-firm, across-plant cannibalization effects are small enough to be safely ignored.

[39] Differences between quantity sold and quantity produced reflect inventory adjustments. As I describe in app. E3, adapting the model described in sec. II to account for inventory adjustments changes little in practice.

TABLE 1
PLANT-PRODUCT-YEAR SUMMARY STATISTICS: MACHINERY, EQUIPMENT, AND PARTS (69,516 Observations)

| Variable | Mean | Standard Deviation | Minimum | Maximum | Median |
|---|---|---|---|---|---|
| Log Revenue ($r_{it}^{j}$) | 15.87 | 2.59 | 1.39 | 25.14 | 15.88 |
| Log Quantity Sold ($q_{it}^{j}$) | 7.65 | 4.17 | −6.91 | 24.17 | 7.46 |
| Log Prices ($p_{it}^{j}$) | 8.23 | 3.6 | −3.05 | 21.43 | 8.26 |
| Log Quantity Produced ($y_{it}^{j}$) | 7.78 | 4.18 | −5.52 | 24.19 | 7.64 |
| Multiproduct | .76 | .43 | 0 | 1 | 1 |
| Single Industry | .5 | .5 | 0 | 1 | 0 |
| Vertical Integration | .2 | .4 | 0 | 1 | 0 |

NOTE.—Multiproduct, Single Industry, and Vertical Integration are all dummy variables. Single Industry refers to products produced by plants that produce only products belonging to ASICC codes 74–78. Vertical Integration refers to plants that produce output sets that I classify as potentially vertically integrated, using information on input use by single-product plants (see app. E1 for more details on the classification of potentially vertically integrated plants). Revenue is measured in nominal rupees, while quantity sold and quantity produced are measured in 5-digit ASICC code–specific units (kilograms, meters, number of units, etc.). Prices are measured as unit values for quantity sold, i.e., the ratio of nominal revenues in rupees, divided by quantity produced.

each product entry is a 5-digit ASI commodity classification, or ASICC, code, with just over 1,000 unique item codes belonging to the machinery, parts, and equipment industry.[40] Each ASICC code is associated with a particular unit of quantity, such as kilograms, tonnes, or units sold, which allows one to use the information on revenues and quantities to construct a within-product-code consistent unit price, which I take to be item-level prices.[41] While each item produced by a plant is assigned a product code, approximately 17% of plant-year observations report multiple entries for the same product code. These are, according to the ASI documentation, not to be regarded as duplicates, indicating that plants also report separate product lines within a 5-digit ASICC code as well. In my empirical analysis, I consider each entry as a separate variety of the same general product class, rather than aggregating to the 5-digit ASICC code level.[42]

I take the inputs of the production function to be labor $L_{it}$, capital $K_{it}$, and materials $M_{it}$. I measure labor input $L_{it}$ by the number of man-days worked and capital $K_{it}$ with the perpetual inventory method. Since I observe information on the price and quantity of various inputs at the 5-digit ASICC code level, to measure materials inputs $M_{it}$ in a manner that is not

[40] These product codes correspond to all product codes that belong to the 2-digit ASICC categories 74–78. See app. E1 for more details and examples of the 5-digit codes.

[41] Note that some product codes do not have quantity information. These account for just over 12% of product-year observations in the data set and are dropped.

[42] In practice, most multiproduct firms report a single variety of each product code. See fig. 3, in app. E, for histograms of product counts by plant measured according to the number of unique entries in the ASI, vs. the number of unique product codes. The vast majority of plants report less than 10 products using either approach to counting unique product lines.

TABLE 2
Plant-Year Summary Statistics: Machinery, Equipment, and Parts (20,702 Observations)

| Variable | Mean | Standard Deviation | Minimum | Maximum | Median |
|---|---|---|---|---|---|
| Log Labor ($l_{it}$) | 8.92 | 1.44 | 3.26 | 14.54 | 8.65 |
| Log Capital Stock ($k_{it}$) | 15.15 | 2.25 | −.09 | 23.48 | 14.94 |
| Log Materials ($m_{it}$) | 6.91 | 3.29 | −3.91 | 22.99 | 6.39 |
| No. of varieties ($J_{it}$) | 1.55 | 1.31 | 1 | 28 | 1 |
| Multiproduct | .26 | .44 | 0 | 1 | 0 |

Note.—Summary statistics are reported only for plants that produce only products belonging to ASICC codes 74–78. Labor is measured in number of man-days worked; Capital is measured deflated rupees, using the capital deflator used in Allcott, Collard-Wexler, and O'Connell (2016); and Materials refers to measured materials expenditures (in current rupees) deflated by a plant-specific Cobb-Douglas price index. See app. E3 for more details on variable construction.

subject to the input price bias discussed by De Loecker and Goldberg (2014) and De Loecker et al. (2016), I deflate materials expenditures by a plant-level Cobb-Douglas input price index. See appendix E3 for more details.

## IV. Estimation

To estimate within-plant heterogeneity, I work with a simple parametric model that satisfies the assumptions of theorem 1. In particular, I assume that the plant-product-level production technology is Cobb-Douglas:

$$Y_{it}^{j} = \exp(\omega_{it}^{j})\left(L_{it}^{j}\right)^{\beta_L}\left(K_{it}^{j}\right)^{\beta_K}\left(M_{it}^{j}\right)^{\beta_M}. \tag{18}$$

Since only aggregate inputs ($L_{it}, K_{it}, M_{it}$) are recorded in the ASI, before estimating equation (18), I must first estimate $(L_{it}^{j}, K_{it}^{j}, M_{it}^{j})$ using demand-side data. For this purpose, I need to specify a demand system and a market structure.

Let $\Lambda_t^{g}$ denote the set of varieties belonging to product nest $g$, where each product nest $g$ corresponds to a 5-digit ASICC code, while a variety $j$ corresponds to a particular plant-product entry in the ASI.[43] I denote the 5-digit code a particular variety $j$ belongs to as $g(j)$, where $g(j)$ should be read a function corresponding to the mapping between varieties and product codes.[44] Demand for product $j \in \Lambda_t^{g(j)} \in \Omega_t^{h(j)} \subset \Omega_t$—where $\Lambda_t^{g(j)}$ and $\Omega_t^{h(j)}$ are the sets of varieties belonging to 5- and 3-digit ASICC codes $\tilde{g} = g(j)$ and $\tilde{h} = h(j)$, respectively—is then given by[45]

[43] Recall that some plants report multiple entries for the same product code, which I treat as separate varieties.

[44] I occasionally make reference to a particular variety's 3-digit and 2-digit product code as well—for this purpose, I use $h(j)$ for 3-digit ASICC codes and $d(j)$ for 2-digit codes.

[45] In all the subsequent notation, I let lowercase letters correspond to natural logs, so $p_{it}^{j} \equiv \ln(P_{it}^{j})$.

$$Q_{it}^{j}\left(\vec{P}_t, \vec{\eta}_t\right) = \frac{I_t^{h(j)}}{P_{it}^{j}} \frac{\exp\left(\delta_{it}^{j}/\sigma\right)\left(\sum_{k\in\Lambda_t^{g(j)}} \exp\left(\delta_{mt}^{k}/\sigma\right)\right)^{\sigma-1}}{\sum_{\Lambda_t^{l}\in\Omega_t^{h(j)}}\left(\sum_{k\in\Lambda_t^{l}} \exp\left(\delta_{mt}^{k}/\sigma\right)\right)^{\sigma}}, \tag{19}$$

where $\alpha > 0$ and $\sigma \in (0, 1]$ are demand parameters, $\delta_{it}^{j} \equiv \eta_{it}^{j} - \alpha p_{it}^{j}$ is the mean utility of product $j$, which depends on both the product's log price $p_{it}^{j}$ and its appeal $\eta_{it}^{j}$, and $I_t^{h}$ is total expenditure by buyers in the market $h$ at time $t$. I consider different 3-digit ASICC codes to be different markets, in the sense that they correspond to different choice sets $\Omega_t^{h}$.[46] Each product set $\Omega_t^{h}$ contains an outside option whose mean utility $\delta_t^{0_h}$ is normalized to 0, which should be regarded as a set of varieties that are relevant to the consumer's choice set that are not actually recorded in the ASI or import flows. As I show in appendix F1, this aggregate-demand system can be derived from a discrete-choice model incorporating continuous quantity choice described in Björnerstedt and Verboven (2016).[47]

One advantage of this demand system is that it is invertible, thereby allowing one to use the market share inversion approach developed in Berry (1994) to estimate its parameters. More importantly, Björnerstedt and Verboven (2016) have shown that the relevant market share inversion for this model is one based on revenue shares rather than quantity shares. This is important, as the market share inversion approach can be sensitive to measurement error in the shares, because of the nonlinearity of the inversion function. Moreover, mismeasurement in unit quantities—and therefore quantity shares—is a common concern when working with plant-level survey data, as responders may not have fully accurate records of quantities or may accidentally report quantities in the wrong units. Relying on an inversion based on revenue shares addresses these concerns, as the nonlinear inversion is instead based on revenues, which are likely better measured in accounting statements and are not likely to be subject to unit errors.[48]

My preferred market structure in this setting is Nash-Bertrand price competition, which will account for within-plant cross-price effects. Since the vast majority of multiproduct plants produce multiple product codes and nests in the above model are defined at the level of a product code, this implies that most plants will charge different markups on different products. As a result, revenue shares are not appropriate in this setting, unless I find that $\sigma = 1$ or I assume that plants are behaving "as if" there was monopolistic competition—that is, treating the price index

[46] See app. E2 for a list of 3-digit ASICC codes.

[47] Note that this demand model is also equivalent to a nested CES demand function, as originally shown by Verboven (1996).

[48] Note, however, that measurement error in quantities can still generate inconsistent estimates through measurement error in unit values or prices. I deal with this issue by relying on an instrument for prices that is unlikely to be correlated with plant-product measurement error in unit quantities.

terms $(\sum_{k\in\Lambda_t^l}\exp(\delta_{mt}^k/\sigma))^\sigma$ as constants.[49] Note, however, that estimates of $(\alpha, \sigma)$ are the only extra information I need to apply the input allocation rule, as I show in appendix F2. In the next section, I outline my sequential estimation strategy, which involves first estimating $(\alpha, \sigma)$ and then applying theorem 1 to estimate $(L_{it}^j, K_{it}^j, M_{it}^j)$, which I then use to estimate the production function (18).[50]

## A. *Demand Estimation*

I apply the demand inversion technique developed in Berry (1994) to equation (19) to estimate the parameters governing the demand function. After following a derivation similar to that in Berry (1994) for the nested logit, where I replace quantity shares with the corresponding revenue shares, I obtain the following estimating equation:

$$\mathrm{rs}_{it}^j - \mathrm{rs}_t^{0_{h(j)}} = (1-\sigma)\mathrm{rs}_{it}^{j|g(j)} - \alpha p_{it}^j + \eta_{it}^j, \tag{20}$$

where $\mathrm{rs}_{it}^j \equiv \ln(R_{it}^j/I_t^{h(j)})$ is the log of product $j$'s revenue share in terms of total revenue generated by 3-digit sector $h(j)$, $\mathrm{rs}_t^{0h(j)}$ is the natural log of the revenue share of the outside option in market $h(j)$, whose mean utility $\delta_t^{0h}$ is normalized to 1, and $\mathrm{rs}_{it}^{j|g(j)} \equiv \ln(R_{it}^j/\sum_{k\in\Lambda_t^{g(j)}} R_{mt}^k)$ is the natural log of the revenue share of variety $j$ within 5-digit ASICC code $g(j)$.[51] Note that I consider imported goods to be part of each 5-digit ASICC nest and hence include total imported revenue within $\Lambda_t^g$.[52]

Equation (20) can be estimated using linear instrumental variables methods, with unobserved product appeal, $\eta_{it}^j$, functioning as the structural residual. This unobserved source of product-level heterogeneity captures product-specific demand shocks as well as differences in quality

[49] Note that within-plant markups are not constant for my specification of demand with Bertrand-Nash pricing, while they are in Hottman, Redding, and Weinstein (2016), because my upper-level nests are 5-digit product codes, while Hottman, Redding, and Weinstein (2016) specifies the upper nest as a firm identifier. By setting the upper-level nest as a product code, I am allowing within-product-code substitution to be larger than across-product-code substitution, which is appropriate for my setting, where choice sets include broad classes of inputs, such as generators and transformers. Note that this nesting pattern causes a plant to charge different markups by product line, since the larger across-product substitution effects lead firms to internalize different degrees of market power across different product codes. This causes product-level markups to vary because of differences in the within-product-code market share of each good. On the other hand, the firm-nesting model considered in Hottman, Redding, and Weinstein (2016), has symmetric substitution patterns across varieties in the same firm, which leads to the same markup being charged on all products.

[50] Since I use a sequential estimation algorithm, I use a plant-level block bootstrap procedure to construct my standard errors when necessary. See app. G for further details.

[51] See app. F2 for further details on the construction of market size and the outside option.

[52] See app. E4 for details on the mapping between 5-digit ASICC codes and 4-digit HS codes used to determine total imports by ASICC code.

across producers. This means that consistent estimation requires instruments that shift production costs while not directly affecting product appeal. For this purpose, I use information on input prices to construct the following instruments:

$$Z_{it}^{g(j)} = Z_t^g = \sum_{k \in \mathbb{I}^g} \gamma^{kg} \times \ln\left(\frac{\sum_{m \in \mathbb{F}_t^{kg}} W_{mt}^k}{\left|\mathbb{F}_t^{kg}\right|}\right) - \frac{1}{T}\sum_{t=1}^{T}\sum_{k \in \mathbb{I}^g} \gamma^{kg} \times \ln\left(\frac{\sum_{m \in \mathbb{F}_t^{kg}} W_{mt}^k}{\left|\mathbb{F}_t^{kg}\right|}\right), \quad (21)$$

$$Z_{it}^{-jg} = \frac{\left(\sum_{k \in \mathbb{Y}_{it}^M} Z_{it}^{g(k)}\right) - Z_{it}^{g(j)}}{\left|\mathbb{Y}_{it}^M\right| - 1}, \quad (22)$$

where $W_{it}^k$ denotes the price of input code $k$ paid by firm $i$, $\mathbb{I}^g$ denotes the set of 5-digit input product codes that I observe being used by single-product producers of product code $g$, $\mathbb{F}_t^{kg}$ denotes the set of plants observed in the ASI at time $t$ who purchase an input with product code $k \in \mathbb{I}^g$ who do not sell any outputs in the machinery, equipment, and parts sector, $\gamma^{kg}$ is the overall cost share of input $k \in \mathbb{I}^g$ in the production of product code $g$ by single-product firms, and $\mathbb{Y}_{it}^M \subset \mathbb{Y}_{it}$ is the set of varieties belonging to the machinery sector sold by plant $i$ at time $t$.

The first instrument, equation (21), which varies across time only within a 5-digit product code, leverages average input price growth experienced by plants in other output markets that use similar inputs.[53] For the instrument to be valid, input price variation should be driven by demand and supply shocks in other industries that are orthogonal to machinery demand shocks or general changes in machinery quality by product code.[54] This requires that changes in average input prices not be driven by machinery demand, which will obviously not be satisfied in input markets where the machinery industry is the primary downstream consumer. To deal with this concern, I exclude any input codes $k$ from $k \in \mathbb{I}^g$ if more than 30% of the revenue I observe going into purchases of $k$ comes from machinery, equipment, and parts producers.[55] Note that equation (21) explicitly avoids using the direct plant-level variation in input prices to identify the demand elasticities, since plant-level input prices likely incorporate

[53] I trim the 95th and 5th percentiles of these prices by product code to limit the influence of outliers when constructing average prices.

[54] Note that requiring that the cost shifters not be correlated with general changes in machinery quality by product code requires that firms not quality-upgrade/downgrade in response to the cost shocks. A sufficient condition for this to hold is that product characteristics, including unobserved product quality, are fixed at product birth, similar to the exogenous-product-characteristics assumption used in Berry, Levinsohn, and Pakes (1995).

[55] Since I am excluding some inputs from the construction of the input price instrument, this means that the input weights, $\gamma^{kg}$, do not necessarily sum to 1 for each output code $g$. However, since the instrument is already demeaned within each product code across time, differences in the size of the admissible set of products will not affect the level of the instrument.

TABLE 3
Demand Estimates (64,917 Observations)

| | OLS | IV | $p_{it}^{j}$ | $\text{rs}_{it}^{j\|g(j)}$ |
|---|---|---|---|---|
| Estimates: | | | | |
| $p_{it}^{j}$ | .007 | −.220 | | |
| | (.002) | (.116) | | |
| $\text{rs}_{it}^{j\|g(j)}$ | .946 | .621 | | |
| | (.004) | (.306) | | |
| First stage: | | | | |
| $Z_{t}^{g(j)}$ | | | .284 | .136 |
| | | | (.097) | (.041) |
| $Z_{it}^{-jg}$ | | | .322 | −.186 |
| | | | (.198) | (.108) |

Note.—Firm age controls, as well as dummies for product code, year, state, census status, rural locations, organization, ownership type, and the number of products, are included in all regressions. Sanderson-Windmeijer $F$-statistics are 16.35 and 10.07 for $p_{it}^{jg}$ and for $\text{rs}_{it}^{j|g}$, respectively. Standard errors adjusted for two-way clustering by plant and product code are in parentheses. IV = instrumental variables.

information on plant-specific output quality, as argued by Kugler and Verhoogen (2012) and De Loecker et al. (2016).

The second instrument, equation (22), provides a second source of identifying variation by recognizing that these cost shocks will affect each multiproduct plant differently through variation in their output sets. In particular, this instrument is based on the average value of $Z_{t}^{g}$ taken by other products produced within the same firm. Note that this instrument will be correlated with prices and revenue shares either through pricing that internalizes within-plant cannibalization effects or through cost shocks that affect common inputs that are used in multiple production lines within a plant.[56]

The results of this estimation strategy can be found in table 3.[57] As expected, OLS (ordinary least squares) estimation of equation (20) generates a price coefficient of the wrong sign, since prices tend to be positively correlated with product appeal. The instrumental variables strategy appears to fix this bias, generating point estimates that imply an average own-price elasticity of approximately −1.6. I also report the first-stage estimates in table 3, with these coefficients generally taking the sign one would intuitively expect (i.e., increased input prices are associated with

[56] Products outside the machinery sector are not included in this average, since $Z_{t}^{g}$ is unlikely to be correlated with machinery prices for these products.

[57] As an alternative to including plant fixed effects, the regressions include plant age controls as well as a set of dummies for state, rural location, census status, organization, and ownership type, as well as the number of products by plant. I have explored using plant fixed effects as well, but these results suffered from serious issues related to instrument strength, with the first-stage $F$-statistics falling below 1. This is likely because this instrument primarily harnesses variation in cost growth across product codes, rather than within-plant cost growth.

output price increases).[58] The Sanderson-Windmeijer first-stage $F$-statistics are found to be 16.35 and 10.07 for $p_{it}^{jg}$ and $\text{rs}_{it}^{j|g}$, respectively, consistent with the instruments having sufficient explanatory power to identify the demand parameters.

### *B. Production Function Estimation*

After estimating demand, I use the estimated demand parameters $(\hat{\alpha}, \hat{\sigma})$ to obtain an estimate of the within-firm input allocations, using theorem 1.[59] I then estimate the plant-product-level production function (18), using a nonlinear GMM estimation procedure. Specifically, to obtain an estimating equation, I assume that productivity follows an exogenous AR(1) process given by[60]

$$\omega_{it}^{j} = \rho_0^{g(j)} + \rho^{d(j)}\omega_{i,t-1}^{j} + \xi_{it}^{j}, \tag{23}$$

where $\rho_0^{g(j)}$ is the mean TFP of varieties belonging to product code $\tilde{g} = g(j)$ and $\rho^{d(j)}$ is a TFP persistence parameter that differs by 2-digit industry $\tilde{d} = d(j)$.[61]

Taking logs of equation (18) and substituting in the law of motion (eq. [23]) yields $y_{it}^{j} = \rho_0^{g(j)} + \beta_L l_{it}^{j} + \beta_K k_{it}^{j} + \beta_M m_{it}^{j} + \rho^{d(j)}\omega_{i,t-1}^{j} + \xi_{it}^{j}$. Quasi differencing, or "$\rho$-differencing," this expression for each 2-digit industry $d$ equation yields

$$\begin{aligned} y_{it}^{j} = {} & \rho_0^{g(j)} + \rho^{d(j)} y_{i,t-1}^{j} + \beta_L\left(\hat{l}_{it}^{j} - \rho^{d(j)}\hat{l}_{i,t-1}^{j}\right) \\ & + \beta_K\left(\hat{k}_{it}^{j} - \rho^{d(j)}\hat{k}_{i,t-1}^{j}\right) + \beta_M\left(\hat{m}_{it}^{j} - \rho^{d(j)}\hat{m}_{i,t-1}^{j}\right) + \xi_{it}^{j} + \varepsilon_{it}^{j}, \end{aligned} \tag{24}$$

where I have substituted estimated input use by product line $\hat{x}_{it}^{j}$ for realized input use $x_{it}^{j}$. As a result, there are two unobservables in this expression: $\xi_{it}^{j}$, the innovation to productivity that is unknown to the plant at time $t - 1$; and $\varepsilon_{it}^{j}$, an error term accounting for the fact that input allocations are estimated in the first-stage demand regression and therefore $\varepsilon_{it}^{j} \neq 0$ in finite samples.[62]

[58] The fact that the input price instrument is positively correlated with the within-product-code revenue shares is perhaps surprising, since one would expect input price increases to decrease market shares. Note, however, that since this instrument does not vary within a product code, this is likely due to increased exit of competitors, which would tend to increase $\text{rs}_{it}^{j|g}$ within a product code.

[59] See app. F2 for more details on the appropriate mapping, given the assumed demand system of eq. (19).

[60] Note that I assume linearity of the productivity process simply for empirical tractability; see app. H4 for an example of how to generate a similar estimating equation for the simple nonlinear law of motion $\omega_{it}^{j} = \rho_0 + \rho_1\omega_{i,t-1}^{j} + \rho_2(\omega_{i,t-1}^{j})^2 + \xi_{it}^{j}$.

[61] Specifications that allow the persistence parameter $\rho$ to vary by 3- and 5-digit product code can be found in app. H3.

[62] This approach abstracts from selection when considering the law of motion, as is common in the dynamic panel approach to production function estimation. While further research

Since the above model is nonlinear in parameters, because of the interactions between $\rho^{d(j)}$ and the production function parameters, equation (24) is estimated with a nonlinear GMM estimation procedure based on a series of moment conditions of the form

$$\mathbb{E}\left[(\xi_{it}^{j} + \epsilon_{it}^{j}) \times \vec{Z}_{it}^{j}\right] = \vec{0}, \tag{25}$$

where $\vec{Z}_{it}^{j}$ is a vector of instruments.[63] Estimation is based on an unbalanced panel of plant-products.[64]

Note that the structure of this estimating equation is very similar to the popular proxy-variable approach to estimation developed in Ackerberg, Caves, and Frazer (2015), with $y_{it}^{j}$ essentially taking the place of their first-stage predicted values. While both approaches leverage very similar sources of variation, note that applying a proxy-variable approach would violate assumption 8, which requires that plants know their entire vector of productivities, $\vec{\omega}_{it}$. As a result, productivity shocks that are unobserved by the plant when choosing inputs, which form the basis of the first-stage estimating equations in the proxy-variable literature, cannot be accommodated, since this breaks the link between observed prices, outputs, and the unobserved input allocations.[65]

I now turn to the problem of choosing appropriate instruments. Since $\xi_{it}^{j}$ is realized in period $t$, lagged inputs $l_{i,t-1}$, $k_{i,t-1}$, and $m_{i,t-1}$ will function

into whether this may generate significant bias would be useful, past research indicates that as long as one does not generate further selection bias by conditioning only on balanced panels, rather than the unbalanced panel, correcting for attrition bias tends to have a negligible effect on the estimated production function coefficients (Olley and Pakes 1996; De Loecker et al. 2016).

[63] As long as the first-stage estimator is consistent, asymptotically the error term in eq. (25) simply becomes $\xi_{it}^{j}$, implying that as long as the instruments are uncorrelated with innovations to productivity $\xi_{it}^{j}$, a two-step GMM estimator will be consistent for the production function parameters.

[64] Note that after solving for the input allocations using each plant's full product sets, I drop plant-products with multiple entries for the same product code in a given plant-year ("multivariety" observations) during estimation, since one cannot precisely determine the appropriate panel variable for these observations.

[65] To clarify this point, note that the input allocation rule is derived by using the plant's first-order conditions determining input use at time $t$, which conditions on the firm's current information set $I_{it}$. Assumption 8 requires that each plant choose the share of inputs being allocated to each product line after all productivity shocks generating $\vec{\omega}_{it}$ have been revealed. As a result, $Y_{it}^{j}$ is equal to each firm's expected output level at time $t$, conditional on $I_{it}$ and the quantity of inputs they choose to allocate to product line $j$. On the other hand, if $Y_{it}^{j} = \mathbb{E}(Y_{it}^{j}|I_{it})\epsilon_{it}^{j}$, where $\epsilon_{it}^{j}$ is a shock that is realized at the end of time $t$ but unknown to the plant (and econometrician) when choosing inputs, then $\mathbb{E}(Y_{it}^{j}|I_{it}) = Y_{it}^{j}/\epsilon_{it}^{j}$ would be the relevant argument for output in the input allocation rule, since first-order conditions for input use would be taken conditional on $I_{it}$. Note, however, that this introduces a whole new set of unobservables to the input allocation rule, making the approach infeasible without some other set of assumptions to pin down these extra shocks.

as valid instruments in this setting.[66] I further follow Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg, Caves, and Frazer (2015) and assume that capital is predetermined because of time to build, which allows $k_{it}$ to be added to the instrument set.[67] Finally, since I wish to identify a separate persistence parameter $\rho^{d(j)}$ for each 2-digit industry, I also add $y^j_{i,t-1}$ multiplied by an indicator for a product's 2-digit industry to the instrument set.

While it is fairly standard to also use the lagged value of static inputs in both proxy-variable and dynamic-panel approaches to production function estimation, note that using only this type of variation to identify static input elasticities, such as $\beta_M$, is problematic.[68] In particular, it is not entirely obvious why the lagged value of a static input should be correlated with differences in static input usage, except through productivity or demand changes, as has recently been pointed out by Gandhi, Navarro, and Rivers (2020). While the fact that lagged inputs may be weak instruments has been widely acknowledged in the dynamic-panel literature (see Blundell and Bond 2000), note that this problem is most likely to be pronounced with static inputs, since there are no adjustment costs for these inputs that make lags relevant for determining future levels.

To deal with this concern, I also use the current and lagged values of the input price instruments described in section IV.A, $Z^{g(j)}_t$ and $Z^{g(j)}_{t-1}$, to identify $\beta_M$. As long as these instruments are valid for demand, meaning that variation in the input price instrument is driven by demand and supply shocks in other output markets, which affect the input prices of goods used to produce product code $g$, then they will function as valid instruments for estimating equation (24) as well.[69] Since this type of variation is not commonly used in the production function literature, I consider a series of specifications that either use $Z^{g(j)}_t$ and $Z^{g(j)}_{t-1}$, or lagged materials, to identify the production function parameters.

Finally, I also include a series of 5-digit product code fixed effects, to deal with differences in units across product codes, as well as fixed effects for the current and lagged number of products, to account for potential

[66] Note that estimated lagged input use by product line would also function as a valid instrument but generates attenuation bias in a finite sample. Moreover, note that if instrument relevance is driven by the existence of adjustment costs (Bond and Söderbom 2005), since adjustment costs vary at the plant rather than the plant-product level in this model, instrument validity is more likely to hold with respect to plant-level inputs than for plant-product-level ones.

[67] Note, however, that contemporary levels of plant-product capital use are not valid instruments in this setting; this is because assumption 4 implies that input shares can react to contemporary shocks, even though the total stock of capital is predetermined.

[68] I regard labor as a dynamic input because of dynamic adjustment costs related to hiring and firing.

[69] Note that I also require that plants not respond to these cost shocks by directly changing their productivity, which eq. (23) already rules outs.

economies-of-scope effects.[70] Since I have more instruments than parameters, I use the standard NL2SLS (nonlinear two-stage least squares) weighting matrix $W = (Z'Z)^{-1}$ when constructing the GMM criterion function.

OLS as well as GMM results based on the instrument sets described in the previous subsection can be found in table 4. The most striking feature worth emphasizing is that the particular set of instruments used to identify the production function can affect whether there is evidence for increasing or constant returns to scale. In particular, in columns 2 and 3, where I include lagged materials in the instrument set, I can reject the constant returns to scale at the standard levels of statistical significance. On the other hand, in column 4, where I drop lagged materials as an instrument, instead relying on the input price instruments $Z_t^{g(j)}$ and $Z_{t-1}^{g(j)}$ to identify the materials elasticity, I can no longer reject constant returns to scale at standard significance levels.[71]

Perhaps more strikingly, increasing returns appear to be driven by the materials output elasticity, with columns 2 and 3 implying a materials elasticity that is three times that obtained by OLS.[72] Since the large output elasticities and returns to scale in columns 2 and 3 may cast some doubt on the validity of lagged materials as an instrument, I rely on the results from column 4 to generate the estimates of TFPQ used in the next section of the paper.[73]

Note that the above estimates assume that the production functions are the same for all product lines. To verify that this is an appropriate assumption for my sample, I also consider specifications that allow the production function to differ by 2-digit ASICC code and conduct a Wald test

[70] In practice, I demean all variables in the estimating equation within product code to difference out the product-code fixed effects. Note that during bootstrapping I occasionally encountered some collinearity issues with the full set of fixed effects for the current and lagged number of products. To avoid these numerical issues, I drop the fixed effects for more than five products and instead have the fifth fixed effect equal 1 for any plant selling five or more products.

[71] Note that one well-known source of increasing returns is shared inputs, which my model generally rules out except for the special form of shared inputs explored in sec. II.C.2. As a result, in non–joint-production settings considered in this paper, one would generally expect to find evidence closer to constant or decreasing returns to scale, as in col. 4 rather than cols. 2 and 3, which have fairly large increasing returns to scale.

[72] In app. H2, I report Hansen's overidentification test for each specification after using a two-step estimator with an optimal weighting matrix, finding that there is evidence that the second specification is misspecified. On the other hand, I do not reject the overidentification test for the specification in cols. 3 and 4.

[73] Moreover, note that cols. 2 and 3 imply the opposite direction for the bias of OLS for materials and capital, as one often expects OLS to generate an upward-biased estimate of the more flexible inputs and downward-biased estimates of the least flexible inputs (Levinsohn and Petrin 2003). The estimates in col. 4, however, are consistent with GMM solving the standard bias of OLS. While this does provide one rationale for choosing the estimates in col. 4, I still report results on within-plant heterogeneity, using the production function estimates obtained with these alternative instruments (see app. J3), and find them quantitatively similar to the baseline results reported in the next section.

TABLE 4
Production Function Estimates (3,620 Observations)

| | OLS | GMM | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $\beta_L$ | .542 | .331 | .321 | .626 |
| | (.057) | (.192) | (.191) | (.261) |
| $\beta_K$ | .223 | .101 | .097 | .236 |
| | (.040) | (.082) | (.081) | (.099) |
| $\beta_M$ | .262 | .790 | .806 | .217 |
| | (.021) | (.191) | (.186) | (.352) |
| $\rho^{74}$ | | .757 | .747 | .842 |
| | | (.222) | (.197) | (.199) |
| $\rho^{75}$ | | .657 | .661 | .670 |
| | | (.082) | (.078) | (.068) |
| $\rho^{76}$ | | .651 | .653 | .623 |
| | | (.098) | (.104) | (.079) |
| $\rho^{77}$ | | .420 | .422 | .541 |
| | | (.062) | (.060) | (.087) |
| $\rho^{78}$ | | .194 | .181 | .569 |
| | | (.319) | (.265) | (.651) |
| $\beta_L + \beta_K + \beta_M$ | 1.027 | 1.222 | 1.224 | 1.078 |
| | (.035) | (.084) | (.080) | (.113) |
| $(Z_t^{g(j)}, Z_{t-1}^{g(j)})$ | . . . | Yes | No | Yes |
| $m_{i,t-1}$ | . . . | Yes | Yes | No |

Note.—OLS and GMM estimates of the production function (18) using an unbalanced panel of plant-product observations. All GMM specifications include $l_{i,t-1}$, $k_{it}$, $k_{i,t-1}$, and $y_{i,t-1}^j$ interacted with a 2-digit ASICC dummy, as instruments. Columns 2 and 3 also include $m_{i,t-1}$ as an instrument, while cols. 2 and 4 include $(Z_t^{g(j)}, Z_{t-1}^{g(j)})$, as defined by eq. (21). All specifications include a series of indicator variables for product code $g(j)$, as well as the total number of products produced by the plant. GMM specifications based on quasi differencing also include indicators for the lagged number of products produced by the plant, interacted with $\rho^{d(j)}$. Plant-level block bootstrapped standard errors are in parentheses (1,000 replications).

for whether the production function differs across 2-digit codes. To get around the identification issues highlighted in section II.C.1, I consider a sample of single-product plants for this purpose, as well as a panel of single-industry plants, which produce only products within the same 2-digit ASICC code, in which case I can still invoke theorem 1 to allocate inputs without knowledge of the production function parameters. The Wald test statistics and corresponding $p$-values can be found in table 5.[74] I find that I cannot reject the null of identical production functions across 2-digit ASICC codes for all specifications. As a result, imposing assumption 2, as I do in table 4, is likely appropriate for the setting considered in this paper.[75]

[74] More details on the estimation procedure, as well as point estimates and standard errors for the production functions, can be found in app. H1.

[75] An important caveat worth noting here is that since the point estimates for the production function parameters in app. H1 are also imprecisely estimated, failure of the Wald test might also be due to a lack of statistical power, rather than a clear rejection of assumption 2. In app. J5, I provide an alternative way to examine whether assumption 2 is an important

TABLE 5
Wald Tests: $F^{D} = F^{D'}$ (2,780 Observations)

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Wald test statistic | 16.039 | 16.978 | 11.424 | 7.090 | 7.054 | 5.488 |
| $p$-value | [.189] | [.150] | [.493] | [.852] | [.854] | [.940] |
| Single product | Yes | Yes | Yes | No | No | No |
| Single industry | No | No | No | Yes | Yes | Yes |
| $(Z_t^{g(j)}, Z_{t-1}^{g(j)})$ | Yes | No | Yes | Yes | No | Yes |
| $m_{i,t-1}$ | Yes | Yes | No | Yes | Yes | No |

Note.—The above reports the Wald test statistics for a joint hypothesis test that $\beta_L^d = \beta_L^{d'}$, $\beta_K^d = \beta_K^{d'}$, and $\beta_M^d = \beta_M^{d'}$ for all $(d, d')$ combinations in {74, 75, 76, 77, 78}, with the corresponding $p$-value for this test in brackets. Production function parameters and standard errors are reported in app. H1. Columns 1–3 use only single-product plants, while cols. 4–6 use only plants that produce within a single 2-digit ASICC code. All specifications use the standard set of instruments described in table 4, as well as the input price instruments $(Z_t^{g(j)}, Z_{t-1}^{g(j)})$ in cols. 1, 3, 4, and 6. In cols. 1, 2, 4, and 5, I also use lagged materials as an instrument.

## V. Results

### A. *Quantifying Within-Plant Heterogeneity*

Having estimated the demand and production parameters for the machinery, equipment, and parts industry, I use this information to construct two measures of plant-product-specific heterogeneity; log TFPQ, defined as $\omega_{it}^j$ in equation (18), and plant-product-specific product appeal, defined as the residual from equation (20), $\eta_{it}^j$. I then use these terms to estimate the magnitude and economic implications of within-plant heterogeneity.[76] Note that since variety-level TFPQ and appeal levels are not necessarily comparable across product codes, because of differences in units, in what follows I normalize each of the variables to have the zero mean and unit variance for each product code, unless otherwise stated.

Table 6 provides a simple first pass at quantifying the degree of within-plant heterogeneity, by decomposing the variance of log TFPQ and appeal into across-plant and within-plant variation using a standard variance decomposition. This approach is based on the identity $x_{it}^j = [(1/J_{it})\Sigma_{k\in\mathbb{Y}_{it}} x_{it}^k] + [x_{it}^j - (1/J_{it})\Sigma_{k\in\mathbb{Y}_{it}} x_{it}^k]$, where the variance of the first term corresponds to the across-plant variance in $x$, while the second component corresponds to the within-plant variance.[77] While the largest component of the variance is accounted for by across-plant heterogeneity, as one might expect, the

restriction by using a cost-share approach for estimating production function parameters by 2-, 3-, and 5-digit ASICC codes. I find that this approach still leads to relatively similar estimates of within-plant heterogeneity, compared to the identical-production-function approach used in the main text.

[76] In app. J1, I also consider a similar decomposition where I reweight TFPQ and appeal according to their overall revenue, as well similar decomposition for marginal costs and output.

[77] The across and within terms are uncorrelated by construction, so they sum to the total variance of $x$.

TABLE 6
Multiproduct Plant Variance Decompositions
(11,933 Observations)

| Type | Across | Within | Total |
|---|---|---|---|
| TFPQ: | | | |
| Variance | .682 | .390 | 1.072 |
| Percentage (%) | 64 | 36 | 100 |
| Appeal | | | |
| Variance | .666 | .397 | 1.063 |
| Percentage (%) | 63 | 37 | 100 |

Note.—Only products produced by multiproduct plants are included in the sample. TFPQ corresponds to log quantity TFP, $\omega_{it}^{j}$. Appeal corresponds to the estimated residual from eq. (20). All variables have been standardized to have mean zero and unit variance by 5-digit product code. See text for definition of "Across" and "Within."

within-plant component is quite sizeable, accounting for just over one-third of the total variation in TFPQ and quality.

An alternative way to characterize the degree of within-plant heterogeneity is to ask the following question: Do plants produce products that lie in similar or different segments of the product-ranking hierarchies? For example, do plants producing high-quality goods produce only products that lie in the top percentile of the product appeal distribution, or does the fact that a plant produces a high-TFPQ product not predict the productivity of their other products? In the former case, across-plant heterogeneity is more important, while in the latter case, the within-plant component clearly matters.

To answer this question, I rank each variety on the basis of TFPQ or appeal within a given 5-digit product code and year, to determine each variety's decile rank. I then regress the decile rank of each variety on a series of dummies for the decile rank of each plant's top-performing variety (i.e., the variety with the highest decile rank, or their "core" product).[78] If there were no within-firm heterogeneity in TFPQ, so that each firm was equally efficient at producing all varieties, then by construction the coefficient on the decile rank dummy $j$ would equal $j$, since the top-ranked decile would be equal to every other decile. On the other hand, the degree to which the estimated coefficient on the decile rank dummy $j$ is below $j$ provides a measure of the degree of within-plant productivity dispersion.

The results of these regressions can be found in table 7, revealing sizable differences in efficiency across products. For example, a plant with a core variety in the top decile (90th–100th percentile) produces other products that lie, on average, somewhere between the 60th and 70th percentiles

[78] Note that I drop plants whose best-performing product is in bottom decile, since by construction all other products will also be in the bottom decile and therefore the max-decile dummy will perfectly predict the outcome for these plants.

TABLE 7
Decile Ranks: Core versus Other Products

| | TFPQ Decile (1) | Appeal Decile (2) |
|---|---|---|
| 2nd decile (10th–20th percentiles) | 1.2985 | 1.2872 |
| | (.0712) | (.0645) |
| Top variety in 3rd decile (20th–30th percentiles) | 1.8710 | 1.7820 |
| | (.1090) | (.1050) |
| Top variety in 4th decile (30th–40th percentiles) | 2.3741 | 2.1396 |
| | (.1163) | (.1315) |
| Top variety in 5th decile (40th–50th percentiles) | 2.9809 | 2.6886 |
| | (.1747) | (.1614) |
| Top variety in 6th decile (50th–60th percentiles) | 3.4419 | 3.2403 |
| | (.1616) | (.1659) |
| Top variety in 7th decile (60th–70th percentiles) | 4.0611 | 3.7018 |
| | (.1839) | (.2413) |
| Top variety in 8th decile (70th–80th percentiles) | 4.7582 | 4.5523 |
| | (.2051) | (.2344) |
| Top variety in 9th decile (80th–90th percentiles) | 5.9277 | 5.5000 |
| | (.2578) | (.2270) |
| Top variety in 10th decile (90th–100th percentiles) | 6.6091 | 6.8969 |
| | (.2560) | (.2191) |
| Observations | 2,122 | 2,143 |

Note.—OLS regressions of a series of indicators for the decile rank of a plant's top product, according to product-code-year TFPQ ranking (col. 1) or product-code-year product appeal ranking (col. 2), on the decile rank of each particular variety within a plant (excluding the top-ranked variety). The sample consists of multiproduct plants that produce only product codes with at least 10 observations per product-code-year. Plants whose best-ranked product is in the bottom decile are dropped. Plant-level block bootstrapped standard errors are in parentheses (1,000 replications).

of the quality ladder in other product codes. This is a fairly steep drop in performance, which appears to hold across the efficiency ladder. These estimates indicate that it is not innocuous to simply ignore product-line efficiency differences, as having a high-ranked variety in one product segment does not guarantee that a plant's other varieties will also be high performing.

### B. *Plant-Level Efficiency and Marginal Varieties*

The previous section documented fairly large within-plant heterogeneity, providing evidence that plants producing high-quality or high-TFPQ products tend to be less efficient at producing other varieties. This suggests that the core-competence model of multiproduct firms (Eckel and Neary 2010; Mayer, Melitz, and Ottaviano 2014; Arkolakis, Ganapati, and Muendler 2021), where firms are less productive at producing varieties outside their core competence, is a first-order feature of the data. In this section, I consider whether there is sufficient within-plant heterogeneity for the extensive margin of product choice to have quantitatively important effects on plant-level efficiency, by estimating a simple core-competence-style efficiency ladder.

I then use these estimates to examine how much plant-level efficiency would increase if one simply removed a plant's lowest-efficiency (or marginal) variety.

## 1. Measuring Plant-Level Productivity

To quantify how the extensive margin of product choice can affect plant-level efficiency, one has to take a stance on how to measure productivity at the plant level. However, this is not straightforward if one measures productivity using TFPQ, since TFPQ will implicitly vary across products because of differences in the quantity units across product codes; as a result, it is not clear how to aggregate across product-code-specific TFPQ to obtain a meaningful plant-level metric of productivity. While standardizing TFPQ to have mean zero and unit variance by product code, as I have so far, partly alleviates this concern, note that product differentiation within a product code can generate differences in measured TFPQ across products that have nothing to do with productivity differences. For example, large motors likely require more capital and labor hours than small motors—as a result, measured TFPQ for large-motor producers will tend to be smaller than that for small-motor producers.

Figure 1 suggests that these concerns may matter in practice. In particular, after standardizing both TFPQ and appeal to have mean zero and unit variance within a product code, I find that high-TFPQ products tend to have low appeal, similar to the findings in Forlani et al. (2016) and Jaumandreu and Yin (2016), who separately identify demand shifters and TFPQ for a number of different industries.[79] A simple explanation for this pattern is that producing high-appeal products, which may be larger or more durable,

[79] Note that the standard errors on this correlation, obtained using the bootstrap, are fairly large. If one simply takes the point estimates of the demand and production function parameters as given when constructing standard errors, both these correlations are statistically significant at the 1% level. However, accounting for uncertainty in the demand and production function parameters generates larger standard errors, as the direction and magnitude of the correlation between TFPQ and appeal are sensitive to the particular parameters used. For example, if one fixes the output elasticities, but recovers TFPQ using $y_{it}^j - [\tilde{\phi}/(\beta_L + \beta_K + \beta_M)](\beta_L l_{it}^j + \beta_K k_{it}^j + \beta_m m_{it}^j)$, where $\tilde{\phi}$ is a desired level of returns to scale, I find that the estimated slope between TFPQ and appeal decreases as $\tilde{\phi}$ increases for the given sample, with the slope of the line of best fit for all plant-product increasing from around $-0.394$ to $-0.067$ as one varies assumed returns to scale from $\tilde{\phi} = 1.3$ to $\tilde{\phi} = 0.7$. Similarly, the estimated slope between TFPQ and appeal tends to decrease as $\alpha$ (the parameter governing price effects in demand) increases, with the slope of the line of best fit increasing from $-0.565$ to $-0.036$ as one varies $\alpha$ from 0.8 to 0.05. This implies that one is most likely to obtain zero or positive correlations when returns to scale and $\alpha$ are small. In particular, of the 37 bootstrap simulations where the correlation between TFPQ and appeal becomes weakly positive for all plant-products, the average value of $\alpha$ is 0.093, while the average value of returns to scale is 0.937. Note, however, that accounting for uncertainty in estimated returns to scale and demand parameters in my bootstrap simulations still generates a statistically significant negative correlation for all plant products at the 10% level, although one cannot reject zero correlation between TFPQ and appeal for the subset of single-product plants.

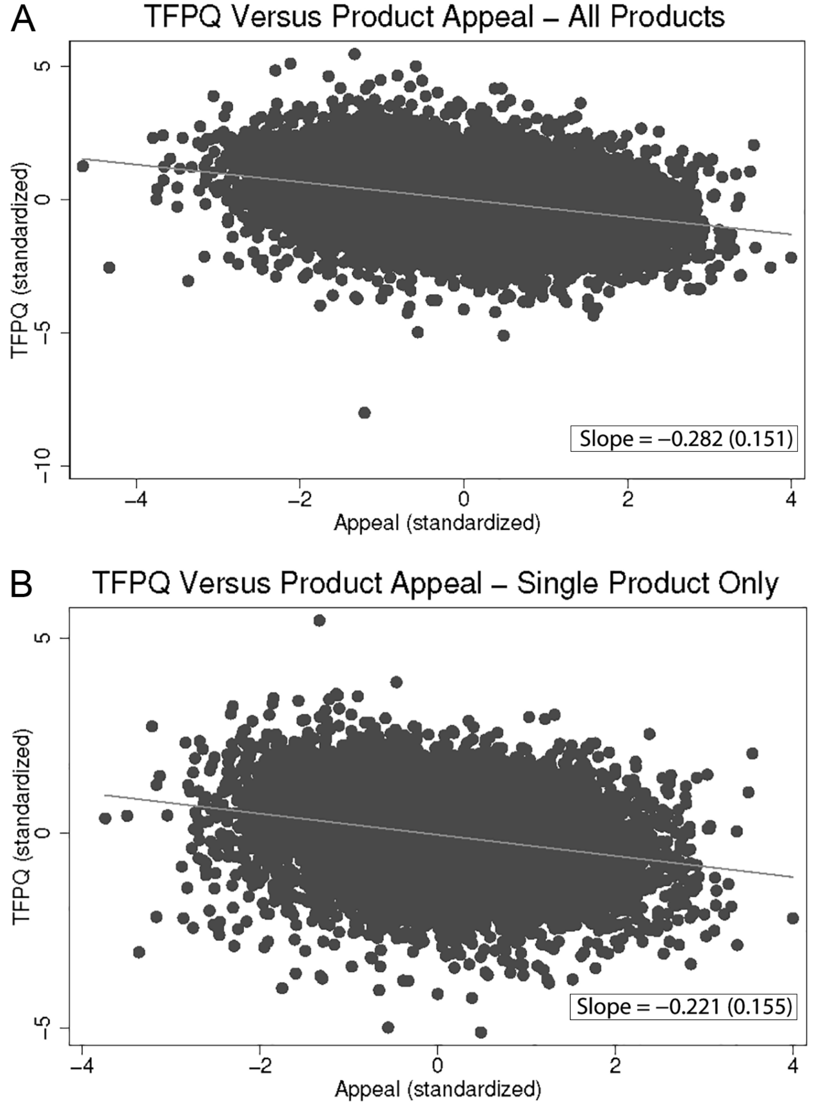


FIG. 1.—Negative correlation between TFPQ and product appeal. Appeal refers to $\eta^j_{it}$, the residual from equation (20), while TFPQ refers to $\omega^j_{it}$ in equation (18). Both variables have been standardized to have a mean of zero and unit variance within each product code. *A* includes all products, while *B* includes only single-product plants. Plant-level block bootstrapped standard errors are in parentheses (1,000 replications).

is costly, which will show up in the data as increased input use—for example, workers might need to work more diligently to produce higher-quality motors. Another explanation for this negative correlation is that firms will choose to produce low-appeal goods only if they are sufficiently cheap—as a result, selection of products may generate a TFPQ cutoff that is decreasing in idiosyncratic appeal, generating a negative relationship between average appeal and TFPQ.

Regardless of which mechanism generates this negative relationship, the core problem is that different varieties within a product code almost certainly vary in terms of their product characteristics, such as size, quality, and durability, and these product characteristics are unobserved in the ASI. This makes TFPQ, which is a purely physical comparison of productivity differences in quantity units, a potentially misleading measure of efficiency.[80] Note, however, that product appeal, or the demand residual $\eta_{it}^{j}$, captures these unobserved differences, since it by construction provides a measure of the portion of variety-level demand that is unexplained by price. A natural way to deal with this comparability problem, then, would be to use some measure of appeal-adjusted TFPQ at the plant-product level, which we might construct by multiplying the measured TPFQ units by estimated appeal. This would provide a measure of productivity that would be in product appeal units and would therefore be directly comparable across product codes.

Interestingly, the approach to productivity measurement described in Klette and Griliches (1996) and De Loecker (2011), based on examining the residual from the revenue production function, directly provides such a measure of appeal-adjusted TFPQ. As described in further detail in appendix I, the demand function (eq. [19]) and the production function (eq. [18]) imply that log revenues at the product level are related to input use as follows:

$$\ln(R_{it}^{j}) = \frac{\alpha/\sigma}{1+(\alpha/\sigma)}\left(\beta_l l_{it}^{j} + \beta_k k_{it}^{j} + \beta_m m_{it}^{j}\right) + \underbrace{\tilde{\omega}_{it}^{j} + \tilde{\eta}_{it}^{j}}_{\equiv E_{it}^{j}}, \tag{26}$$

where

$$\tilde{\eta}_{it}^{j} \equiv \frac{\eta_{it}^{j}}{\alpha+\sigma} + \frac{1}{1+(\alpha/\sigma)}\ln\left(I_t^{h(j)}\right) - \frac{1-\sigma}{1+(\alpha/\sigma)}\ln\left(\sum_{k\in\Lambda_t^{g(j)}}\exp\left(\frac{\eta_{it}^{k}-\alpha p_{it}^{k}}{\sigma}\right)\right)$$

$$- \frac{1}{1+(\alpha/\sigma)}\ln\left(\sum_{\Lambda_t^{l}\in\Omega_t^{h(j)}}\left(\sum_{k\in\Lambda_t^{l}}\exp\left(\frac{\eta_{it}^{k}-\alpha p_{it}^{k}}{\sigma}\right)\right)^{\sigma}\right), \tag{27}$$

$$\tilde{\omega}_{it}^{j} \equiv \frac{\alpha/\sigma}{1+(\alpha/\sigma)}\omega_{it}^{j}. \tag{28}$$

[80] See also Atkin, Khandelwal, and Osman (2019) for further evidence from the flatweave rug industry that TFPQ can perform poorly as a measure of productivity in settings within significant quality variation.

The residual from equation (26), which I have denoted $E_{it}^{j}$ for variety-level "efficiency," turns out a function quite similar to appeal-adjusted TFPQ as described in the previous paragraph. Moreover, since the residual is now measured in units of log revenues over inputs, it is directly comparable across plants and products. For this reason, I refer to $E_{it}^{j}$ as "revenue efficiency."

Note that I intentionally use the term "revenue efficiency" for my notion of revenue-based TFP, rather than "TFPR," as in Foster, Haltiwanger, and Syverson (2008), since the residual from a revenue production function is not the same thing as TFPR, a point that has been emphasized by Haltiwanger (2016) and Foster et al. (2016). In particular, TFPR is defined in Foster, Haltiwanger, and Syverson (2008) as TFPQ × Price, which by construction will vary with prices and therefore partly will measure differences in market power across firms. Note, however, that this is not necessarily the case for revenue efficiency; rather, differences in prices across products within the same product code are captured by the revenue production function itself. This can be seen by noting that the production function coefficients in equation (26) are each scaled down by $(\alpha/\sigma)/[1+(\alpha/\sigma)]$, since increases in input use will require that prices fall for quantity produced to still equal quantity demanded. Since the revenue production function is constructed by assuming that each firm operates on their inverse demand curve, price variation is partly captured by differences in input use, rather than being loaded on a residual term.

On the other hand, differences in market power across product codes or across time can still show up in revenue efficiency $E_{it}^{j}$, through the price index terms $\ln(\Sigma_{k\in\Lambda_t^g}\exp((\eta_{it}^k-\alpha p_{it}^k)/\sigma))$ and $\ln(\Sigma_{\Lambda_t^l\in\Omega_t^h}(\Sigma_{k\in\Lambda_t^l}\exp((\eta_{it}^k-\alpha p_{it}^k)/\sigma))^{\sigma})$. Note, however, that these terms are identical for all products within a given product code and year and therefore can be purged from my efficiency measures by demeaning $E_{it}^{j}$ within a given product code–year.

Another useful feature of working with the revenue production function is that it aggregates up to a plant-level revenue production function in an internally consistent way, providing a natural metric to quantify the effects of product dropping. As I show in appendix I, the model estimated in this paper implies the following plant-level log revenue efficiency measure:

$$
\begin{aligned}
E_{it} &\equiv \ln\left(\sum_{j\in\mathbb{Y}_{it}} R_{it}^{j}\right) - \frac{\alpha/\sigma}{1+(\alpha/\sigma)}(\beta_l l_{it}+\beta_k k_{it}+\beta_m m_{it}) \\
&= \underbrace{\frac{1}{|\mathbb{Y}_{it}|}\sum_{j\in\mathbb{Y}_{it}} E_{it}^{j}}_{\equiv E_{it}^{M}:\ \text{mean plant efficiency}} + \underbrace{\ln\left(\sum_{j\in\mathbb{Y}_{it}} \left(S_{it}^{j}\right)^{\phi(\alpha/\sigma)/[1+(\alpha/\sigma)]}\exp\left(E_{it}^{j}-\frac{1}{|\mathbb{Y}_{it}|}\sum_{j\in\mathbb{Y}_{it}} E_{it}^{j}\right)\right)}_{\equiv E_{it}^{S}:\ \text{specialization effects}}.
\end{aligned} \tag{29}
$$

Overall plant-level revenue efficiency depends on two key terms: mean plant efficiency, which is simply an unweighted average of plant-product-level log revenue efficiency $E_{it}^{j}$, and specialization effects, which will depend on the correlation between the plant's chosen input shares and within-plant demeaned log revenue efficiency. I show in appendix I2 that the vast majority of the variation of plant-level revenue efficiency is due to variation in the unweighted average of plant-product revenue efficiency, while specialization effects account for only around 3%–4% of the variance of plant-level revenue efficiency. This means that by focusing on variation in the mean plant-efficiency term, I can capture close to 96%–97% of the determinants of plant-level revenue efficiency. As a result, in the next subsection, I approximate the full effect of the extensive margin of product choice on plant-level efficiency by examining the effect of removing a product on mean plant-level product-code demeaned efficiency:

$$\hat{E}_{it}^{M} \equiv \frac{1}{|\mathbb{Y}_{it}|} \sum_{j \in \mathbb{Y}_{it}} \left( E_{it}^{j} - \frac{1}{|\Lambda_t^{g(j)}|} \sum_{k \in \Lambda_t^{g(j)}} E_{it}^{k} \right), \tag{30}$$

where I focus on product-code demeaned efficiency so that all efficiency gains will be based on within-product-code variation in revenue efficiency, rather than efficiency gains due to product-code-level differences in competition structure as captured by the price index terms.

## 2. Marginal Varieties in Core-Competence Models

In this section, I estimate a within-plant technology based on the core-competence models developed in Eckel and Neary (2010), Mayer, Melitz, and Ottaviano (2014), and Arkolakis, Ganapati, and Muendler (2021) to examine whether the removing marginal (lowest-efficiency) products leads to quantitatively important changes in plant-level revenue efficiency. These papers assume that each multiproduct producer faces a technology where, as new product lines are added to a producer's product set, these new varieties are less productive than their other product lines, capturing the idea that a firm is moving farther away from their "core competence." This class of technologies takes the following form when applied to my notion of within-plant log revenue efficiency:

$$\hat{E}_{it}^{r} - \hat{E}_{it}^{1} = -H(r), \tag{31}$$

where $H'(r) > 0$, and $r = 1, 2, \ldots, J_{it}$ indexes a particular variety within a multiproduct plant, after ordering all products from highest to lowest revenue efficiency.

For example, Mayer, Melitz, and Ottaviano (2014) consider a linear specification of the technology, $H(r) = \beta \times (r - 1)$, while Arkolakis, Ganapati,

TABLE 8
WITHIN-PLANT HETEROGENEITY IN CORE-COMPETENCE MODELS

| | Linear (1) | Log-Linear (2) | Nonparametric (3) |
|---|---|---|---|
| $r-1$ | −.4773 | | |
| | (.1261) | | |
| $\ln(r)$ | | −1.2378 | |
| | | (.3165) | |
| Indicator for $r=2$ | | | −.9602 |
| | | | (.2197) |
| Indicator for $r=3$ | | | −1.4527 |
| | | | (.3504) |
| Indicator for $r=4$ | | | −1.7336 |
| | | | (.4397) |
| Indicator for $r=5$ | | | −1.9281 |
| | | | (.5081) |
| Indicator for $r=6$ | | | −2.0349 |
| | | | (.5632) |
| Indicator for $r=7$ | | | −2.0428 |
| | | | (.6206) |
| Indicator for $r=8$ | | | −2.1260 |
| | | | (.6552) |
| Indicator for $r=9$ | | | −2.3322 |
| | | | (.6957) |
| Indicator for $r\geq 10$ | | | −2.7367 |
| | | | (.8775) |
| Observations | 7,997 | 7,997 | 8,019 |

NOTE.—OLS regressions of $\hat{E}^r_{it}-\hat{E}^1_{it}$ on listed independent variables. Products with rank $j=1$ are excluded. Columns 1 and 2 exclude 22 plant-products that are ranked larger than 10 because of the rarity of these particular rankings. Plant-level block bootstrapped standard errors are in parentheses (1,000 replications).

and Muendler (2021) consider a log-linear specification, $H(r)=\beta\times\ln(r)$.[81] Lacking direct estimates of within-plant productivity differences, both papers use the structure of their model and variation in export revenues across products and markets to estimate $\beta$. However, since I have direct estimates of within-plant productivity, I can obtain an estimate of these parameters by directly estimating equation (31) by OLS, which provides a useful summary statistic of the degree of within-plant heterogeneity. These results can be found in columns 1 and 2 of table 8.[82]

Both functional forms yield extremely large estimates of within-plant heterogeneity.[83] In particular, column 1 implies that a plant's worst product is approximately 45% less efficient than its second-worst product. Column 2, on the other hand, which allows for nonlinear product scope effects

[81] Eckel and Neary (2010) often leave the form of $H(r)$ unrestricted, although they are able to strengthen some of their results by considering a linearly declining technology.

[82] I do not include a constant in these regressions, so all deviations from the model and the data are captured by the error term.

[83] Note that the magnitude of these coefficients is important, not the direction, as by construction I have to obtain negative coefficients in these regressions.

because of the inclusion of the natural log of product rank, implies that a plant's second-best product is approximately 58% less efficient than the plant's core product, while a plant's third-best product is approximately 39% worse than its second-best product. In column 3, I nonparametrically account for declines in productivity, finding effects that are fairly similar in magnitude to the log-linear specification.

In figure 2, I plot the increase in plant-level efficiency, as measured by equation (30), generated by removing a plant's worst-performing product according to the core-competence technology described in table 8. The implied productivity effects of product dropping are extremely large, increasing plant-level revenue efficiency by around 22%, according to the linear specification of the technology, while generating log revenue efficiency gains of 10% to just under 65% when one relies on the log-linear or nonparametric specification of the technology.

Overall, these results imply that fairly large increases in plant-level revenue efficiency can be generated by simply removing a plant's least efficient variety. Note, however, that these revenue efficiency increases will be realized only if plants actually drop less efficient goods. In the next section,

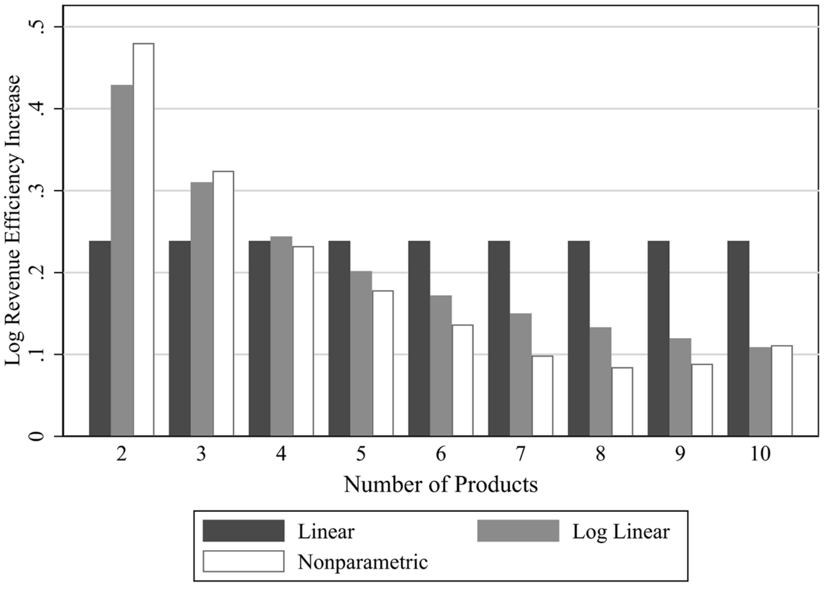


FIG. 2.—Plant-level revenue efficiency growth after removal of marginal varieties. The figure shows the implied log revenue efficiency growth from removing a plant's lowest-efficiency variety according to core-competence technology estimates in table 8. Log plant-level revenue efficiency is measured according to equation (30), with the $y$-axis denoting the change in plant-level log revenue efficiency after a plant's worst product is dropped.

I provide evidence that plants are most likely to quit producing their least efficient varieties.

### C. *Do Plants Care about Within-Plant Heterogeneity?*

While I have presented evidence that within-plant efficiency differences exist and are sizeable, I have not yet provided any evidence that manufacturing plants actually care about these objects. Since TFPQ and appeal are residual-based measures, they may simply be a "measure of our ignorance," as per Syverson (2011, 330), rather than real economic primitives that plants are taking into account when making decisions.

Since the approach developed in this paper does not impose any discipline on the manner in which plants choose their product sets over time, I now test whether variety-specific heterogeneity is related to a plant's decision to drop a product, that is, cease production of that variety in the following period. If plants also choose their product sets to maximize their profits, as in Maksimovic and Phillips (2002), Bernard, Redding, and Schott (2010, 2011), Eckel and Neary (2010), and Mayer, Melitz, and Ottaviano (2014), then plants should be more likely to drop products with lower TFPQ or appeal, all else equal.[84] I also decompose which sources of variety-specific heterogeneity are primarily driving product-dropping behavior to determine which dimensions of product positioning are most salient to Indian plants.

In particular, I first estimate a probit model for the probability of dropping a product in the subsequent period, given total plant-product-level revenue efficiency. For this purpose, I consider a product dropped whenever I observe a plant producing a particular product code $g$ at time $t$, but not at time $t + 1$.[85] Note that product dropping is quite common in my sample, as 43.1% of the observed plant-years have at least one product dropped in the subsequent period (excluding the last year each plant is

[84] Many of these papers pin down the equilibrium number of products by assuming the existence of product-specific fixed costs. One can add such fixed costs to the model developed in this paper without changing any of the features of the input allocation problem as long as those fixed costs are in nominal (rupee) units. For example, if product assortment choices are made a period in advance of production and efficiency shocks are persistent, then low-TPFQ or low-appeal products will be more likely to be dropped, since in expectation those varieties will be less likely to cover their fixed cost.

[85] Note that the procedure for classifying dropped products is more complicated when a plant produces multiple varieties of the same product code. For example, one may observe a particular plant producing two varieties of transformers in period $t$ but only one variety at time $t + 1$, To deal with this, I classify the varieties with the most different value of revenues, when compared across time periods, as the dropped products. More formally, if a plant drops $V$ varieties of the same product code $g$, I calculate the minimum pairwise difference between the revenues earned by each variety $j \in \Lambda_{it}^{g}$ and the revenues of each variety $j \in \Lambda_{i,t+1}^{g}$. The set of varieties with the $V$ largest values of this minimum pairwise difference are then classified as the dropped products.

TABLE 9
PRODUCT-DROPPING PROBITS (3,472 Observations)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| $E_{it}^{j}$ | −.0665 | | | | | | | |
| | (.0218) | | | | | | | |
| $\bar{P}_{t}^{g}$ | | −.0318 | | | | | | −.0254 |
| | | (.0260) | | | | | | (.0264) |
| $\bar{A}_{it}^{j}$ | | | −.0403 | | | | | −.0389 |
| | | | (.0195) | | | | | (.0195) |
| $\bar{C}_{it}^{j}$ | | | | .0029 | | | | −.0078 |
| | | | | (.0174) | | | | (.0207) |
| $\hat{P}_{t}^{g}$ | | | | | −.0574 | | | −.0896 |
| | | | | | (.0220) | | | (.0256) |
| $\hat{A}_{it}^{j}$ | | | | | | −.0797 | | −.1277 |
| | | | | | | (.0188) | | (.0289) |
| $\hat{C}_{it}^{j}$ | | | | | | | .0088 | −.0699 |
| | | | | | | | (.0177) | (.0375) |

NOTE.—Estimates of the marginal effects from a probit regression on an indicator variable for whether a plant drops a product in the following period. Definitions of all independent variables are described by eqq. (26), (32) and (33). $E_{it}^{j}$, and $P_{t}^{g}$ have been standardized to have mean zero and unit variance over the entire sample, while $C_{it}^{j}$ and $A_{it}^{j}$ are standardized within a product-code-year. To guarantee that a product-dropping variable is capable of being constructed for all observations, the sample consists of multiproduct plants that I observe in the current as well as the subsequent period. All regressions include year fixed effects. Marginal effects are evaluated at the sample average for the covariates. Plant-level block bootstrapped standard errors are in parentheses (1,000 replications).

observed).[86] In these regressions, I first consider the full value of revenue efficiency as in equation (26), so that all the determinants of revenue efficiency, including product-code-specific shocks such as changes in demand or level of competition, are embodied in this object. I then decompose product-level revenue efficiency into a number of separate components, to examine whether different sources of heterogeneity have different effects on dropping behavior, similar in spirit to the decomposition for plant exit effects examined in Foster, Haltiwanger, and Syverson (2008) for demand versus TFPQ shocks. These results are reported in table 9.

The baseline results in column 1 imply that plants do indeed care about product-level revenue efficiency. In particular, a product that is 1 standard deviation below the mean value of revenue efficiency is around 6.5 percentage points more likely to be dropped in the subsequent period than a product with the mean value of revenue efficiency. These are fairly large effects, generating marginal exit probabilities that are around four to six times larger than the effect of plant-level TFPR variation on plant-level exit documented in Foster, Haltiwanger, and Syverson (2008).

[86] Note that plants are about equally likely to add products as well—in particular, focusing on plant-years where I observe the plant in at least one earlier time period, I also find that 43.3% of plant-years have production sets with at least one new product.

In columns 2–8 of table 9, I dig a bit deeper into the components of revenue efficiency that are driving a plant's decision to drop a product. In particular, I rely on the following decomposition of revenue efficiency into three separate components:

$$E_{it}^{j} = \underbrace{\frac{1}{\left|\Lambda_t^{g(j)}\right|}\sum_{k\in\Lambda_t^{g(j)}} E_{it}^{k}}_{\equiv \text{ product code profitability } (P_t^g)} + \underbrace{\tilde{\omega}_{it}^{j} - \frac{1}{\left|\Lambda_t^{g(j)}\right|}\sum_{k\in\Lambda_t^{g(j)}} \tilde{\omega}_{it}^{k}}_{\equiv \text{ relative cost advantage } (C_{it}^j)} + \underbrace{\tilde{\eta}_{it}^{j} - \frac{1}{\left|\Lambda_t^{g(j)}\right|}\sum_{k\in\Lambda_t^{g(j)}} \tilde{\eta}_{it}^{k}}_{\equiv \text{ relative appeal advantage } (A_{it}^j)}. \quad (32)$$

The first component, which I call product-code profitability, is simply the average value of revenue efficiency by product code and time period.[87] As should be clear from equations (26)–(28), this variable partly reflects variation in the difficulty of producing different product codes but also variation in competition and overall demand and market size across product codes.[88] The second component, which I call a variety's relative cost advantage, essentially measures a particular product's ranking within a product code's current TFPQ ladder, while the third component, relative appeal advantage, measures a product's ranking within the current quality/appeal ladder.[89] By decomposing revenue efficiency into these three components, I can then examine whether product dropping is primarily driven by across-product-code variation in profitability or by differences in appeal or cost comparative advantage within product codes. This is useful, as it can speak to whether cost-side product positioning, where a firm may gain an advantage in the market by producing at lower costs and potentially undercutting their competitors, is more or less important than demand-based product positioning.

In order to get a fuller picture of what is driving dropping behavior, I then further decompose each component of revenue efficiency into its plant-level average and its within-plant relative ranking. More formally, for each revenue efficiency component $\mathcal{X} \in (P, C, A)$, I define

[87] Averages are done with respect to all observable values of revenue efficiency in my data set, meaning that I always include single-product varieties in these averages to obtain the correct ranking of products.

[88] While one could further break this variable down into an average TFPQ component and an average demand-shifter component, note that the level of these variables across product codes is very difficult to interpret because of differences in units. As a result, it is easier to make comparisons of the composite variable, as it is in revenue efficiency units for all product codes.

[89] Since relative appeal and cost advantages do not have any natural units, I standardize these variables within each product code and time period. This removes variation in the length of different TFPQ/appeal ladders by product code, so that a 1-unit increase in cost or appeal advantage is measured as a 1 standard deviation increase in the appeal/TFPQ ladder of that particular product.

$$\mathcal{X}_{it}^{j} = \underbrace{\frac{1}{|\mathbb{Y}_{it}|}\sum_{k\in\mathbb{Y}_{it}}\mathcal{X}_{it}^{k}}_{\text{across plant }(\bar{\mathcal{X}}_{it})} + \underbrace{\mathcal{X}_{it}^{jg} - \frac{1}{|\mathbb{Y}_{it}|}\sum_{k\in\mathbb{Y}_{it}}\mathcal{X}_{it}^{k}}_{\text{within plant }(\hat{\mathcal{X}}_{it}^{j})}. \tag{33}$$

This allows me to examine whether product dropping is driven by a particular product's ranking within a plant's revenue efficiency (or TFPQ/appeal) ladder or is a general response to overall low performance by the plant. Note that the former effect is consistent with a core-competence model, as product dropping leads plants to become "leaner and meaner" (Eckel and Neary 2010). On the other hand, if only plant-level average productivity matters for product dropping, it is unlikely that plants are considering within-plant heterogeneity when choosing their product sets.[90]

Applying the decomposition in equation (33) to each of the three right-hand-side variables in equation (32) leads to six components of revenue efficiency variation that could be driving product-dropping behavior. In columns 2–7, I run the product-dropping probits for each of these variables separately, and then in column 8, I control for all individual sources of heterogeneity.

Note that simply conditioning on cost advantages alone, as in columns 3 and 7, explains little to no dropping behavior. However, if one conditions on all sources of within-plant heterogeneity, as in column 8, then within-plant variation in cost advantages ($\hat{C}_{it}^{j}$) appears to explain a sizable component of product-dropping behavior, with a 1 standard deviation decrease in relative cost advantages, holding productivity of all other products fixed, generating around a 7 percentage point increase in the probability of dropping that product.[91] This effect is slightly smaller than the effect of within-plant product code profitability ($\hat{P}_{t}^{g}$) shocks, where a 1 standard deviation decrease in profitability leads to a 9 percentage point increase in the likelihood of dropping that product. Plants appear to be most sensitive to within-plant appeal shocks ($\hat{A}_{it}^{j}$), with a 1 standard deviation decrease in appeal advantages leading to around a 13 percentage point increase in the probability of dropping that product. The fact that plants appear to be more sensitive to demand-side appeal shocks than to cost-side TFPQ shocks is consistent with Hottman, Redding, and Weinstein (2016), who find that appeal shocks are more important than cost dispersion for explaining differences in firm size. Moreover, these results may also indicate that to first order, demand-side product characteristics, as captured

[90] Since I wish to compare within-plant effects to average plant effects, I drop single-product plants from these regressions, since they cannot have within-plant effects. Note, however, that my findings are robust to including single-product plants as well. See app. J2.

[91] This estimate is somewhat imprecisely estimated, with the corresponding $t$-statistic lying slightly below the threshold for significance at the 5% level, although it is well within the bounds for significance at the 10% level.

by variety appeal, are a more salient feature of product position than is a variety's location in TFPQ space.

The within-plant variation in the various components of product heterogeneity appears to generate much larger product-dropping effects, while the across-plant effects are generally much smaller and often statistically indistinguishable from zero.[92] This shows that plants appear to be taking the various dimensions of within-plant product heterogeneity into account when choosing their products sets. As a result, table 9 provides new evidence that the productivity gains from product dropping emphasized by the literature on core competence are operative in the data.[93]

### *D. Robustness*

The baseline results estimated the production function and input allocations using a Bertrand-Nash pricing assumption. In appendix J4, I show that the quantitative conclusions are largely unaffected by varying the assumed competition structure to monopolistic competition or full collusion. The within-plant component of TPFQ accounts for slightly more of the total variance under monopolistic competition (38%), while TFPQ obtained assuming collusion generates a within-plant component that is about the same size as that found in the baseline Bertrand-Nash case (36%). I also find similar increases in plant-level efficiency by removing a plant's worst product according to the core-competence model, accounting for increases of around 50%–65% when considering two-product plants and increases of around 10% once one considers plants with seven or more products. I also find that plants are similarly sensitive to within-plant heterogeneity, as the product-dropping probits provide quantitatively similar marginal effects. While more research is needed into the robustness of various input allocation rules for different questions, it is useful to know that the monopolistic competition specification, which generates a revenue-share input allocation rule that is straightforward to construct and often used in the literature, delivers estimates similar to those one would obtain by estimating a full model of Bertrand-Nash competition or collusion in this particular application.

[92] In fact, only firm-level average appeal advantages appear to have a statistically significant effect on product dropping. This may indicate that plant-level demand shocks are more important than supply shocks. For example, plant-level reputation may be an important component of product-level demand, which is likely to spill over to each of the products a plant produces.

[93] Appendix J2 reports the corresponding estimates for the specifications in cols. 1 and 8 of table 9 using logit, OLS, and OLS with product code and plant fixed effects, as well as the baseline probit specification, where I also include single-product plants in the estimation sample. These effects are quantitatively similar to the baseline results, although the sensitivity to full $E^j_{it}$ shocks appears 60% larger in the linear probability model with product code and plant fixed effects.

In appendix J4, I also consider the robustness of my conclusions concerning within-plant heterogeneity according to the extension developed in section II.C.3, which does not require assuming costless transferability for all inputs. Treating capital's contribution to production as an unobserved input generates a slightly smaller role for within-plant TFPQ variation (34%), while allowing both capital and labor to be unobserved components of TFPQ generates a within-plant contribution accounting for between 28% and 37% of total TFPQ variation, depending on the magnitude of the estimated output elasticity for materials. The product-dropping sensitivities remain quite close to the baseline marginal effects when one allows only capital to not be subject to assumption 4. On the other hand, the marginal effects of within-plant cost and appeal heterogeneity are, respectively, around 25% and 40% smaller than the baseline effects when both labor and capital do not satisfy costless transferability for my preferred specification, which uses single-product plants to estimate the materials output elasticity. More importantly, relaxing assumption 4 by including the contribution of specific capital or labor to TFPQ can generate much larger increases in plant-level revenue efficiency when marginal varieties are removed according to the simple core-competence efficiency ladder. In particular, two-product plants can experience efficiency gains of around 0.66–0.72 log points (or efficiency increases of around 94%–105%) when I remove their worst product, which is about 50% larger than the baseline effects. Since these specifications treat specialized inputs as another source of within-plant heterogeneity, this is perhaps unsurprising, as this simply means that allowing for further sources of product-level heterogeneity generates a steeper within-plant efficiency ladder.[94] Note, however, that if one is comfortable allowing labor to be subject to costless transferability, thus allowing the use of theorem 1 to determine the labor allocations, the gains from product dropping are quantitatively quite similar to those obtained in the baseline.

Finally, while I found evidence that identical output elasticities by product line are likely appropriate for the machinery sector considered in this paper, I also explore in appendix J5 whether relaxing the assumption of identical production functions across product lines affects my conclusions concerning the magnitude of within-plant heterogeneity. For this purpose, I rely on the cost-share approach to obtain output elasticities, as in Syverson (2004), Foster, Haltiwanger, and Syverson (2008), and Backus (2020), relying on average output elasticities by 2-, 3-, and 5-digit code obtained using single-product plants.[95] I find that this approach generates slightly less

[94] In particular, since these unobserved dynamic inputs are likely to be positively correlated with plant-product revenue efficiency, the total variance of $E_{it}^{j} + \beta_X x_{it}^{j}$ is likely to be larger than the variance of $E_{it}^{j}$ alone.

[95] While this approach requires some assumptions that are not necessary in the production function estimation framework considered in this paper—namely, that all inputs are

within-plant TFPQ variation, accounting for closer to one-quarter of the total TFPQ variation, rather than a third. On the other hand, the product-dropping regressions are quantitatively similar, generating product-dropping sensitivities to revenue efficiency, as well as within-plant relative cost and appeal advantages, that are similar in magnitude to those in the main text.

While these different specifications do lead to some small differences in the quantitative implications of within-plant heterogeneity, note that all specifications point toward the same general picture: within-plant heterogeneity is sizeable, and plants appear to care about within-plant heterogeneity, as they are more likely to drop their lowest-performing products.

## VI. Conclusion

This paper has developed a flexible recipe for estimating product-level TFP for multiproduct firms in data sets where the allocation of inputs across production lines is unknown, using a combination of demand- and supply-side information. The only additional empirical information required to implement the approach developed in this paper is firm-product-level prices and quantities. Since this information is becoming more widely available in many firm- and plant-level data sets, the techniques developed here should be able to help researchers interested in productivity differences across firms to tackle a wider class of problems.

Applying the approach to a panel of manufacturing plants in India from 2000–2007 provides sizeable estimates of within-plant heterogeneity in efficiency. Productivity differences across product lines are quantitatively important, potentially accounting for plant-level efficiency gains of around 50%–65% if a two-product plant were to cease producing its lower-productivity good. Such productivity gains are often realized in practice, as plants are found to be more likely to drop their lowest-performing products.

The results in this paper suggest important areas for future research. For example, while this paper has documented a number of ways in which product dropping can matter quantitatively, more precise estimates of the gains from product dropping may be obtained by carefully modeling the process through which firms choose their product sets. Progress on this question could be made by applying the techniques developed in this paper to a model of product-level entry and exit with multiproduct producers. For example, one might use such a model to obtain estimates of (fixed) product development costs, which could then be used to quantify the effect of competition shocks on product variety and welfare. The techniques and

static and competitively supplied with constant returns to scale—a key advantage of this framework is that since it as based on cost shares (which must always be between 0 and 1), it is straightforward to obtain output elasticities at very fine levels of aggregation.

empirical results described in this paper will hopefully provide a preliminary first step toward such a model for future work.

## References


Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators." *Econometrica* 83 (6): 2411–51.

Allcott, Hunt, Allan Collard-Wexler, and Stephen D. O'Connell. 2016. "How Do Electricity Shortages Affect Industry? Evidence from India." *A.E.R.* 106 (3): 587–624.

Amiti, Mary, and Amit K. Khandelwal. 2013. "Import Competition and Quality Upgrading." *Rev. Econ. and Statis.* 95 (2): 476–90.

Arkolakis, Costas, Sharat Ganapati, and Marc-Andreas Muendler. 2021. "The Extensive Margin of Exporting Products: A Firm-Level Analysis." *American Econ. J. Macroeconomics* 13 (4): 182–245.

Atalay, Enghin. 2014. "Materials Prices and Productivity." *J. European Econ. Assoc.* 12 (3): 575–611.

Atkin, David, Amit K. Khandelwal, and Adam Osman. 2019. "Measuring Productivity: Lessons from Tailored Surveys and Productivity Benchmarking." *AEA Papers and Proc.* 109:444–49.

Backus, Matthew. 2020. "Why Is Productivity Correlated with Competition?" *Econometrica* 88 (6): 2415–44.

Balat, Jorge, Irene Brambilla, and Yuya Sasaki. 2016. "Heterogeneous Firms: Skilled-Labor Productivity and the Destination of Exports." Working paper.

Baldwin, John, and Wulon Gu. 2009. "The Impact of Trade on Plant Scale, Production-Run Length and Diversification." In *Producer Dynamics: New Evidence from Micro Data*, edited by Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts, 557–92. Chicago: Univ. Chicago Press (for NBER).

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

Berger, David, Kyle Herkenhoff, and Simon Mongey. 2022. "Labor Market Power." *A.E.R.* 112 (4): 1147–93.

Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott. 2010. "Multiple-Product Firms and Product Switching." *A.E.R.* 100 (1): 70–97.

———. 2011. "Multiproduct Firms and Trade Liberalization." *Q.J.E.* 126 (3): 1271–318.

Berry, Steven T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *RAND J. Econ.* 25 (2): 242–62.

Berry, Steven, Amit Gandhi, and Philip Haile. 2013. "Connected Substitutes and Invertibility of Demand." *Econometrica* 81 (5): 2087–111.

Berry, Steven T., and Philip A. Haile. 2014. "Identification in Differentiated Products Markets Using Market Level Data." *Econometrica* 82 (5): 1749–97.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63 (4): 841–90.

Björnerstedt, Jonas, and Frank Verboven. 2016. "Does Merger Simulation Work? Evidence from the Swedish Analgesics Market." *American Econ. J. Appl. Econ.* 8 (3): 125–64.

Blum, Bernardo S., Sebastian Claro, Ignatius Horstmann, and David A. Rivers. 2021. "The ABCs of Firm Heterogeneity When Firms Sort into Markets: The Case of Exporters." Working paper.

Blundell, Richard, and Stephen Bond. 2000. "GMM Estimation with Persistent Panel Data: An Application to Production Functions." *Econometric Rev.* 19 (3): 321–40.

Boehm, Johannes, Swati Dhingra, and John Morrow. 2018. "Product Diversification in Indian Manufacturing." In *Developments in Global Sourcing*, edited by Wilhelm Kohler and Erdal Yalcin, 337–50. Cambridge, MA: MIT Press.

Bokhari, Farasat A. S., and Franco Mariuzzo. 2018. "Demand Estimation and Merger Simulations for Drugs: Logits v. AIDS." *Internat. J. Indus. Org.* 61:653–85.

Bond, Stephen, and Måns Söderbom. 2005. "Adjustment Costs and the Identification of Cobb Douglas Production Functions." IFS Working Paper no. WP05/04, Inst. Fiscal Studies, London.

Brooks, Wyatt J., Joseph P. Kaboski, Yao Amber Li, and Wei Qian. 2021. "Exploitation of Labor? Classical Monopsony Power and Labor's Share." *J. Development Econ.* 150:102627.

Card, David, Ana Rute Cardoso, Joerg Heining, and Patrick Kline. 2018. "Firms and Labor Market Inequality: Evidence and Some Theory." *J. Labor Econ.* 36 (S1): S13–S70.

Caves, Douglas W., Laurits R. Christensen, and W. Erwin Diewert. 1982. "The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity." *Econometrica* 50 (6): 1393–414.

Collard-Wexler, Allan, and Jan De Loecker. 2015. "Reallocation and Technology: Evidence from the US Steel Industry." *A.E.R.* 105 (1): 131–71.

Crawford, Gregory S. 2012. "Endogenous Product Choice: A Progress Report." *Internat. J. Indus. Org.* 30 (3): 315–20.

De Loecker, Jan. 2011. "Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity." *Econometrica* 79 (5): 1407–51.

De Loecker, Jan, and Pinelopi K. Goldberg. 2014. "Firm Performance in a Global Market." *Ann. Rev. Econ.* 6:201–27.

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

De Loecker, Jan, and Frederic Warzynski. 2012. "Markups and Firm-Level Export Status." *A.E.R.* 102 (6): 2437–71.

Dhyne, Emmanuel, Amil Petrin, Valerie Smeets, and Frederic Warzynski. 2017. "Multi Product Firms, Import Competition, and the Evolution of Firm-Product Technical Efficiencies." Working Paper no. 23637 (July), NBER, Cambridge, MA.

———. 2021. "Theory for Extending Single-Product Production Function Estimation to Multi-product Settings." Working paper.

Doraszelski, Ulrich, and Jordi Jaumandreu. 2013. "R&D and Productivity: Estimating Endogenous Productivity." *Rev. Econ. Studies* 80 (4): 1338–83.

Eckel, Carsten, and J. Peter Neary. 2010. "Multi-product Firms and Flexible Manufacturing in the Global Economy." *Rev. Econ. Studies* 77 (1): 188–217.

Ershov, Daniel, Jean-William Laliberté, Mathieu Marcoux, and Scott Orr. 2021. "Estimating Complementarity with Large Choice Sets: An Application to Mergers." Working paper. https://doi.org/10.2139/ssrn.3802097.

Eslava, Marcela, and John Haltiwanger. 2018. "The Life-Cycle Growth of Plants: The Role of Productivity, Demand and Distortions." Working paper. https://doi.org/10.2139/ssrn.3177289.

Eslava, Marcela, John Haltiwanger, Adriana Kugler, and Maurice Kugler. 2004. "The Effects of Structural Reforms on Productivity and Profitability Enhancing Reallocation: Evidence from Colombia." *J. Development Econ.* 75 (2): 333–71.

Feenstra, Robert, and Hong Ma. 2007. "Optimal Choice of Product Scope for Multiproduct Firms under Monopolistic Competition." Working Paper no. 13703 (December), NBER, Cambridge, MA.

Forlani, Emanuele, Ralf Martin, Giordano Mion, and Mirabelle Muûls. 2016. "Unraveling Firms: Demand, Productivity and Markups Heterogeneity." CEPR Discussion Paper no. DP11058, Centre Econ. Policy Res., London.

Foster, Lucia, Cheryl Grim, John Haltiwanger, and Zoltan Wolf. 2016. "Firm-Level Dispersion in Productivity: Is the Devil in the Details?" *A.E.R.* 106 (5): 95–98.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *A.E.R.* 98 (1): 394–425.

Gandhi, Amit, Salvador Navarro, and David A. Rivers. 2020. "On the Identification of Gross Output Production Functions." *J.P.E.* 128 (8): 2973–3016.

Garcia-Marin, Alvaro, and Nico Voigtländer. 2017. "Product-Level Efficiency and Core Competence in Multi-product Plants." 2018 Meeting Papers no. 737, Soc. Econ. Dynamics, Minneapolis. https://economicdynamics.org/meetpapers/2018/paper_737.pdf.

———. 2019. "Exporting and Plant-Level Efficiency Gains: It's in the Measure." *J.P.E.* 127 (4): 1777–825.

Goldberg, Pinelopi Koujianou, Amit Kumar Khandelwal, Nina Pavcnik, and Petia Topalova. 2010a. "Imported Intermediate Inputs and Domestic Product Growth: Evidence from India." *Q.J.E.* 125 (4): 1727–67.

———. 2010b. "Multiproduct Firms and Product Turnover in the Developing World: Evidence from India." *Rev. Econ. and Statis.* 92 (4): 1042–49.

Gong, Binlei, and Robin C. Sickles. 2021. "Resource Allocation in Multi-divisional Multi-product Firms." *J. Productivity Analysis* 55 (2): 47–70.

Haltiwanger, John. 2016. "Firm Dynamics and Productivity: TFPQ, TFPR, and Demand-Side Factors." *Economía* 17 (1): 3–26.

Hottman, Colin J., Stephen J. Redding, and David E. Weinstein. 2016. "Quantifying the Sources of Firm Heterogeneity." *Q.J.E.* 131 (3): 1291–364.

Iaria, Alessandro, and Ao Wang. 2021. "Identification and Estimation of Demand for Bundles." Working paper. https://doi.org/10.2139/ssrn.3458543.

Ichniowski, Casey, and Kathryn Shaw. 1999. "The Effects of Human Resource Management Systems on Economic Performance: An International Comparison of U.S. and Japanese Plants." *Management Sci.* 45 (5): 704–21.

Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi. 1997. "The Effects of Human Resource Management Practices on Productivity: A Study of Steel Finishing Lines." *A.E.R.* 87 (3): 291–313.

Itoga, Takaaki. 2019. "Within-Firm Reallocation and the Impacts of Trade under Factor Market Imperfection." In "Essays on Multi-product Firms," 16–55. PhD diss., Pennsylvania State Univ.

Jaumandreu, Jordi, and Heng Yin. 2016. "Cost and Product Advantages: A Firm-Level Model for the Chinese Exports and Industry Growth." Working paper.

Khandelwal, Amit. 2010. "The Long and Short (of) Quality Ladders." *Rev. Econ. Studies* 77 (4): 1450–76.

Klette, Tor Jakob, and Zvi Griliches. 1996. "The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous." *J. Appl. Econometrics* 11 (4): 343–61.

Kugler, Maurice, and Eric Verhoogen. 2012. "Prices, Plant Size, and Product Quality." *Rev. Econ. Studies* 79 (1): 307–39.

Lamorgese, Andrea R., Andrea Linarello, and Frederic Warzynski. 2015. "Free Trade Agreements and Firm-Product Markups in Chilean Manufacturing." Working paper.

Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *Rev. Econ. Studies* 70 (2): 317–41.

Maican, Florin, and Matilda Orth. 2021. "Determinants of Economies of Scope in Retail." *Internat. J. Indus. Org.* 75:102710.

Maksimovic, Vojislav, and Gordon Phillips. 2002. "Do Conglomerate Firms Allocate Resources Inefficiently across Industries? Theory and Evidence." *J. Finance* 57 (2): 721–67.

Malikov, Emir, and Gudbrand Lien. 2021. "Proxy Variable Estimation of Multiproduct Production Functions." *American J. Agricultural Econ.* 103 (5): 1878–902.

Martin, Leslie A., Shanthi Nataraj, and Ann E. Harrison. 2017. "In with the Big, Out with the Small: Removing Small-Scale Reservations in India." *A.E.R.* 107 (2): 354–86.

Mayer, Thierry, Marc J. Melitz, and Gianmarco I. P. Ottaviano. 2014. "Market Size, Competition, and the Product Mix of Exporters." *A.E.R.* 104 (2): 495–536.

———. 2021. "Product Mix and Firm Productivity Responses to Trade Competition." *Rev. Econ. and Statis.* 103 (5): 874–91.

Morlacco, Monica. 2018. "Market Power in Input Markets: Theory and Evidence from French Manufacturing." Working paper.

Nevo, Aviv. 1998. "Identification of the Oligopoly Solution Concept in a Differentiated-Products Industry." *Econ. Letters* 59 (3): 391–95.

Newey, Whitney K., and Daniel McFadden. 1994. "Large Sample Estimation and Hypothesis Testing." In *Handbook of Econometrics*, vol. 4, edited by Robert F. Engle and Daniel L. McFadden, 2111–245. Amsterdam: North-Holland.

Olley, G. Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64 (6): 1263–97.

Porter, Michael E. 1980. *Competitive Strategy: Techniques for Analyzing Industries and Competitors.* New York: Free Press.

Robinson, Joan. 1933. *The Economics of Imperfect Competition.* London: Palgrave Macmillan.

Rubens, Michael. 2021. "Market Structure, Oligopsony Power, and Productivity." Working paper. https://doi.org/10.2139/ssrn.3800254.

Schoar, Antoinette. 2002. "Effects of Corporate Diversification on Productivity." *J. Finance* 57 (6): 2379–403.

Smeets, Valerie, and Frederic Warzynski. 2013. "Estimating Productivity with Multiproduct Firms, Pricing Heterogeneity and the Role of International Trade." *J. Internat. Econ.* 90 (2): 237–44.

Song, Minjae, Sean Nicholson, and Claudio Lucarelli. 2017. "Mergers with Interfirm Bundling: A Case of Pharmaceutical Cocktails." *RAND J. Econ.* 48 (3): 810–34.

Syverson, Chad. 2004. "Market Structure and Productivity: A Concrete Example." *J.P.E.* 112 (6): 1181–222.

———. 2011. "What Determines Productivity?" *J. Econ. Literature* 49 (2): 326–65.

Thomassen, Øyvind, Howard Smith, Stephan Seiler, and Pasquale Schiraldi. 2017. "Multi-category Competition and Market Power: A Model of Supermarket Pricing." *A.E.R.* 107 (8): 2308–51.

Valmari, Nelli. 2016. "Estimating Production Functions of Multiproduct Firms." ETLA Working Paper no 37, Res. Inst. Finnish Econ., Helsinki. http://pub.etla.fi/ETLA-Working-Papers-37.pdf.

Verboven, Frank. 1996. "The Nested Logit Model and Representative Consumer Theory." *Econ. Letters* 50 (1): 57–63.

Wang, Ao. 2021. "A BLP Demand Model of Product-Level Market Shares with Complementarity." Working paper. https://doi.org/10.2139/ssrn.3785148.