# The Comparative Advantage of Firms

# Johannes Boehm

Sciences Po, Centre for Economic Performance, London School of Economics, and Centre for Economic Policy Research

# Swati Dhingra

Centre for Economic Performance, London School of Economics, and Centre for Economic Policy Research

# John Morrow

Kings College London, Centre for Economic Performance, London School of Economics, and Centre for Economic Policy Research

Resource-based theories propose that firms grow by diversifying into products that use common capabilities. We provide evidence for commoninput capabilities, using a policy that removed entry barriers in input markets to show that the similarity of a firm's and an industry's input mix determines firm production choices. We model industry choice and economies of scope from input capabilities. When the model is estimated for Indian manufacturing, input complementarities make firms 5% more likely to produce in an industry and are quantitatively as important as time-invariant drivers of coproduction rates. Upstream entry barriers were equivalent to a 9.5% tariff on inputs.

#### I. Introduction

Multiproduct firms dominate production and export activity. They are much larger than their single-product counterparts, and their product

We are grateful to Andy Bernard, Keith Head, Beata Javorcik, Kalina Manova, Thierry Mayer, Veronica Rappoport, and John Sutton, as well as audiences at many seminars and conferences,

Electronically published October 25, 2022

Journal of Political Economy, volume 130, number 12, December 2022.

© 2022 The University of Chicago. All rights reserved. Published by The University of Chicago Press. https://doi.org/10.1086/720630 turnover contributes substantially to aggregate output growth.<sup>1</sup> Recent work in international economics and industrial organization examines how many products firms make and the impact of economic changes on these choices. It emphasizes the importance of core products for firm growth (Bernard, Redding, and Schott 2010, 2011; Eckel and Neary 2010; Iacovone and Javorcik 2010; Mayer, Melitz, and Ottaviano 2014), but less is known about why these products are "core." This paper examines firm decisions to make products across different industries and provides reduced-form evidence and structural estimates for comparative advantage arising from input capabilities and industrial coproduction.

Early theoretical and empirical work recognizes the sizable contribution of product diversification toward firm growth and aggregate productivity and examines explanations for product diversification within firms (such as Stigler 1951 and Scherer 1982, summarized in Chandler 1992 and Montgomery 1994). Explanations on the demand side include gaining market power through horizontal and vertical integration or internalizing demand complementarities and network externalities across products (e.g., Bernheim and Whinston 1990; Willig, Salop, and Scherer 1991; Jovanovic and Gilbert 1993; Hoberg and Phillips 2016). On the supply side, agency-based theories of the firm suggest that diversification is motivated by internal labor and capital markets of the firm. For example, managers choose to diversify to reduce their human capital risk, to gain rents from utilizing free cash flows, or to obfuscate when their own division is doing badly (Amihud and Lev 1981; Jensen 1986; Morck, Shleifer, and Vishny 1990). This explained early trends of reduced firm valuations from diversification (e.g., Lichtenberg 1992; see Maksimovic and Phillips 2013 for a survey). Resource-based theories of the firm, dating back to Penrose (1955), take a competing view that diversification enables firms to grow beyond the limits imposed by the

for helpful comments. An earlier version of the paper was titled "Swimming Upstream: Input-Output Linkages and the Direction of Product Adoption." Boehm thanks the Sciences Po/Banque de France partnership for financial assistance and the International Economics Section at Princeton for hospitality. Morrow thanks the London School of Economics–Tata Institute of Social Sciences–Tata grant for financial assistance. Dhingra thanks the European Research Council for financial assistance under Starting Grant 760037. Shania Bhalotia and Yan Liang provided excellent research assistance. Replication files are available in a zip file. This paper was edited by Chad Syverson.

<sup>&</sup>lt;sup>1</sup> For example, in the United States, multiproduct firms account for over 90% of manufacturing output, and multiproduct exporters account for over 95% of exports. They are larger than single-product firms in the same industry in terms of shipments (0.66 log points), employment (0.58), labor productivity (0.08), and total factor productivity (TFP; 0.02). About 89% of multiproduct firms vary their product mix within 5 years, and these changes in the product mix make up a third of the increase in US manufacturing output (Bernard et al. 2007; Bernard, Redding, and Schott 2010). In India, multiproduct firms (that produce in more than one of 262 different industries) account for 32% of firms and 62% of sales (as we discuss below). Among publicly listed firms, Goldberg et al. (2010b) find that multiproduct firms, which produce in more than one of 108 4-digit NIC (National Industrial Classification) industries, make up 47% of firms and 80% of sales. They are 107% bigger in output than single-product firms within the same industry.

size of a single-product market. Entering new products requires resources, such as know-how or inputs, that are costly to acquire and to transfer outside the firm. Firms therefore gain economies of scope by diversifying into products that require similar know-how or inputs to what their existing products use, as experienced during wartime, when auto manufacturers quickly switched to making tanks, chemical companies to making explosives, and radio manufacturers to making radar (Teece 1982; Wernerfelt 1984; see Baumol 1977; Panzar and Willig 1981).

Theories of product diversification have influenced a vast literature in economics, finance, and management that examines which products firms choose to make.<sup>2</sup> For example, recent microdata reveal that firms are much more likely to produce in certain pairs of industries (Bernard, Redding, and Schott 2010). Many firms that make fabricated metal products also make industrial machinery. A challenge, however, has been to move beyond systematic correlations in coproduction and product characteristics to disentangling evidence for specific theories of the firm. In the example, systematic coproduction of metal and machinery could arise because they share common inputs and technologies such as metal and metalworking (as in the resource-based view). Or it could arise because consumers who order fabricated metal products also need industrial machinery (as in demand-side theories) or still further because machinery and metal are produced in firms with excess capital (as in agency-based theories). This paper examines product diversification within firms in the light of these theories of the firm and with a view to understanding the comparative advantage of firms in the product space.

Using plant-level data from Indian manufacturing, this paper starts with the striking observation that product diversification within firms is systematically related to shared input use across these products. Figure 1 shows that firms produce more in pairs of industries that require similar intermediate inputs. Figure 1*A* illustrates the extent of coproduction of industry pairs within plants (across 1–253 different industries), and figure 1*B* is for the input similarity between industry pairs.<sup>3</sup> This is also borne out in findings from the United States, where stark examples

<sup>&</sup>lt;sup>2</sup> In early work, Scherer (1982) estimates technology flows across industries to examine business lines within firms and the related slowdown in aggregate productivity growth in the United States. Recent work has built on these findings to show a systematic relationship between demand relatedness, technological relatedness or input relatedness of products made by firms, and various firm performance measures (e.g., Robins and Wiersema 1995; Fan and Lang 2000; Bowen and Wiersema 2005; Rondi and Vannoni 2005; Bryce and Winter 2009; Liu 2010).

<sup>&</sup>lt;sup>3</sup> The coproduction cells contain the size-weighted average sales shares of plants that derive the largest share of revenue from products in the row industry. Darker values indicate higher shares. The input similarity cells contain the inner product of the industries' vector of intermediate-input expenditure shares, calculated from single-product plants in each industry.



3028

FIG. 1.—Coproduction and input similarity. *A*, For plants with primary sales in the row industry, the fraction of sales coming from products in the column industry. *B*, Inner product between the row and column industry's intermediate-input expenditure share vectors. Intermediate-input shares (*B*) are constructed from single-industry plants only. Darker colors indicate larger relative values within each panel.

of industry pairs that are coproduced and have similar input requirements include textiles and apparel, lumber and paper, and primary metal and fabricated metal (Bernard, Redding, and Schott 2010).<sup>4</sup>

To disentangle input-based product diversification from other explanations, the paper leverages plausibly exogenous variation in input supply to relate product choices within firms with input similarity across products. Starting in the late 1990s, the Indian government dismantled size-based entry barriers in several products that were previously reserved for production by small-scale firms. The removal of entry barriers was driven primarily by an agenda to reform postindependence economic policy.<sup>5</sup> As the entry barriers were lifted, firms acquired better access to inputs. Those firms that intensively used these inputs became more likely to grow by diversifying into products that are intensive in the use of liberalized inputs. To give a concrete example, when entry barriers to cotton are lifted, a cotton apparel maker becomes more likely (than a silk apparel maker) to move into cotton textile production (than into silk textile production). In fact, even within the cotton apparel industry, a

<sup>4</sup> Similar patterns emerge in firm-level data from the United Kingdom and Belgium (Hutchinson, Konings, and Walsh 2010; Bernard et al. 2019), and early work by Wernerfelt (1984) suggests systematically examining these patterns through "resource-product" matrices.

<sup>&</sup>lt;sup>5</sup> The original aim of the reservation policy was employment generation through smallscale units that were expected to be more labor intensive than larger firms (though Martin, Nataraj, and Harrison 2017 show that the dismantling of this policy in fact generated relatively more employment).

firm that is relatively more intensive in cotton use becomes relatively more likely to move into cotton textile production. In other words, firms diversified into industries in which they had input-based comparative advantage.

This paper examines the extent to which better input supply enables firms to acquire comparative advantage in industries related by common inputs. According to comparative-advantage theory, industries differ in their use of technologies or factor requirements, and countries differ in their technological prowess or factor endowments. Countries therefore produce relatively more in industries in which they are more capable through better technologies or greater reliance on abundant factors. Translating this from countries and technologies/factors to firms and inputs, we examine whether firms produce more in their input-intensive industries relative to the typical firm in those industries and relative to other industries that firms could enter. Common input requirements provide well-measured proxies for shared technical know-how across products, which overcomes long-standing constraints associated with measuring technologies and intangible know-how (Atalay et al. 2019). The focus on intermediate inputs enables direct examination of the empirical relevance of economies of scope and the extent to which these economies are determined by policy choices. In concrete terms, firms intensively using metal inputs have metalworking know-how and skills, and the policy change in metal inputs provides variation in supply complementarities that can be directly linked to production outcomes.

While the policy episode is best suited to examining economies of scope through inputs, the setting is amenable to disentangling competing explanations for product diversification. The production data can be used to distinguish various resources within firms, such as intermediate inputs and primary factors, that have been shown to be important constraints to firm growth in developing economies (Tybout 2000; Bloom et al. 2010). The rich microdata can also be leveraged to construct measures suggested by competing theories, such as vertical upstream/downstream linkages, capital intensity, and the diversification discount. In particular, firm-time, firm-industry, and industry-time fixed effects control for unobserved reasons such as firms' financial and managerial conditions, industry technological shifts, and other time-invariant firm-industry reasons for coproduction. In the absence of direct measures of substitutability/complementarities across products and their evolution, industry-mix fixed effects can also be included to account for unobservable demand-side reasons for coproduction. Accounting for the different theories of the firm, we find that the dominant explanation in our context and policy episode is that firms specialize in products where they have comparative advantage based on shared input use.

Having established the importance of input-based comparative advantage for firms, the paper provides a theoretical framework for input capabilities and their contribution to firm sales. Starting from the primitive of industry-specific production functions, differences across industries arise from differences in their input requirements. Differences across firms arise from their endowed industry productivities and from their decisions to invest in input capabilities, which can be shared across industries. Economies of scope induce coproduction in industries that are intensive in the use of dynamically acquired input capabilities. Removal of entry barriers in input markets provides better access to those inputs and confers an advantage to firms that have higher use for those inputs. These firms step up production, but much more so in industries that use these inputs more. In sum, policy-induced improvements in input supply enable firms to diversify into industries in which they have input-based comparative advantage relative to other firms, even within their industry.

The theory allows for love or hate for input variety, and a key insight of our framework is that economies of scope within multiproduct firms imply that production choices and input capabilities are jointly determined. Since firms are heterogeneous in their costly-to-transfer resources, this joint determination of downstream input capabilities and production choices is around the revealed "core competencies" of the firm, resulting in input-based comparative advantage. The framework generates structural estimating equations that explain the portfolio of industries a firm adopts on the basis of contemporaneous input similarity with each industry. The latter, in turn, is determined by policy changes that improve access to inputs and by demand and supply shocks that interact with a firm's industry mix. The theory guides estimation of common industry demand innovations and policy changes in input supply to predict contemporaneous input similarity, which in turn determines industry choice.

The estimates show that input capabilities are quantitatively important in determining the industry choice and scope of firms. On average, inputbased comparative advantage makes single-industry firms 5.2 percentage points (pp) more likely to produce in an industry. Multi-industry firms exhibit a distribution of such premia. Shared input capabilities provide advantages across multiple industries, but this decays as a firm diversifies into industries with more varied inputs. For instance, triple-industry firms are 6.1 pp more likely to enter their first industry, 2.3 pp more likely to enter their second industry, and only 1.9 pp more likely to enter their third industry. However, as multi-industry firms are larger across the board, sizeweighted premia range as high as 46.8 pp, showing that they are important for firm growth. Overall, input-based comparative advantage is quantitatively as important a determinant of firm entry into industries as timeinvariant industry-pair determinants of coproduction rates.

*Related literature.*—The results relate to the multiproduct-firm literature, which usually focuses on how many, not which, products firms make. We contribute to this literature by identifying the role of input

linkages as a determinant of the core competencies of multiproduct firms.<sup>6</sup>

A large literature studies the role of access to inputs on firm productivity.<sup>7</sup> While we ask a different question, the focus on input supply is consistent with these studies. Specifically, Goldberg et al. (2010b) highlight the importance of input supply in Indian manufacturing. They find that large firms in India increased the range of products they offered in response to India's input tariff liberalization of the nineties. Their focus is on the number of products firms make. We instead examine which products firms make and, in doing so, uncover input capabilities-based comparative advantage of firms. Goldberg et al. (2010a) differentiate the role of price and new variety channels of imported inputs in expanding firm product scope, finding a crucial role for new imported varieties and allowing for potential technological complementarity within firms or product lines, something we also find and structurally model. Vandenbussche and Viegelahn (2018) also show that Indian firms move away from inputs facing domestic antidumping measures by decreasing sales of products using these inputs. Similarly, we find that intermediate inputs drive output decisions. Distinct from their work, our reduced-form analysis examines measures suggested by competing theories of product diversification, and we provide a structural estimating equation for the relationship between input linkages and output decisions.

While our focus is on supply-side policies in a developing-country context, the approach of characterizing firms and industries is similar to that of Bloom, Schankerman, and Van Reenen (2013) and Conley and Dupor (2003). Bloom et al. construct technological and product market proximity measures to identify the causal effect of R&D spillovers across US firms by using changes in federal and state tax incentives for R&D. Conley and Dupor construct input similarity measures between sectors. They show that cross-sector productivity covariance tends to be greatest between sectors that are similar in inputs and that this channel contributes substantially to the variance in aggregate productivity. We build on these ideas and show how plants internalize input linkages to achieve product diversification.

The question of product choice in a developing-country setting is related to work by Hausmann, Hwang, and Rodrik (2007) and Hidalgo et al. (2007), who examine the product space of countries and the network structure of their products. They propose that products differ in

<sup>&</sup>lt;sup>6</sup> See also Foster, Haltiwanger, and Syverson (2008), Eckel and Neary (2010), Liu (2010), Dhingra (2013), Mayer, Melitz, and Ottaviano (2014), and Eckel et al. (2015) in the multiproduct literature and Hottman, Redding, and Weinstein (2016) and Bernard et al. (2022) in the firm-heterogeneity literature.

<sup>&</sup>lt;sup>7</sup> See, e.g., Acemoglu, Antràs, and Helpman (2007), Amiti and Konings (2007), Kasahara and Rodrigue (2008), Kugler and Verhoogen (2009, 2012), Antràs and Chor (2013), and Halpern, Koren, and Szeidl (2015). In recent work, Lu, Mariscal, and Mejía (2016) model the inherently dynamic process of accumulating input capabilities and its role in increasing firm productivity.

the capabilities needed to make them and that countries differ in the capabilities they have. Countries make products for which they have the requisite capabilities, and they tend to move to goods close to those they are currently specialized in (Hidalgo et al. 2007). Introducing quality capabilities to this framework, Sutton and Trefler (2016) show a nonmonotonic relationship between advances in countries' wealth and changes in their product mix and quality. We apply these ideas at the microeconomic level of a production unit and find empirical support for input-based diversification of the product space. This confirms the view of Hausmann and Hidalgo (2011) that developing a regional jet aircraft is likely to be less costly for those who have previously developed a transcontinental aircraft and a combustion engine, compared to those who previously produced only raw cocoa and coffee.

In innovative work at the firm level, Flagge and Chaurey (2014) use a moment-inequality methodology to estimate bounds on the costs of adding products, including the role of product-proximity measures. Like theirs, our work connects to studies documenting relatedness across products made by firms, though we differ in using policy variation to identify input-based comparative advantage. Using a different approach, Aw and Lee (2009) focus on four Taiwanese electronics industries and estimate cost functions to arrive at the incremental marginal cost of the core product when the firm adds a new product. The industrial policy we exploit eased entry barriers in previously reserved industries and has been of interest in understanding competition, employment generation, productivity growth, and misallocations in manufacturing (Bollard, Klenow, and Sharma 2013; García-Santana and Pijoan-Mas 2014; Galle 2015; Martin, Nataraj, and Harrison 2017). We show a new channel, input-side complementarities, through which the policy affected the economy.

Our work is related more broadly to the literatures on industry linkages and entry barriers.<sup>8</sup> We quantify entry barriers in terms of tariff rates that have equivalent effects on firm decisions to move into industries. On average, entry barriers from the policy to reserve products for small-scale plants are equivalent to input tariffs of 9.5%. Domestic policies, such as size-based entry barriers, are well understood to be a nontariff barrier to doing business. Given their prevalence as a protectionist tool, a large literature in international economics has tried to quantify such policies in terms of tariffs that have an equivalent effect on outcomes of interest. But such quantification is typically fraught with difficulties for reasons such as limited variation

<sup>&</sup>lt;sup>8</sup> There are a growing number of studies relating linkages to productivity (see the handbook chapter by Combes and Gobillon 2015). In particular, López and Südekum (2009) find that upstream, but not downstream, linkages are associated with higher productivity, perhaps in part because of the stronger effect of upstream linkages on product adoption that we find.

in policies and correlation of policy changes with other shocks.<sup>9</sup> The Indian context overcomes these problems to reveal the constraints placed by domestic policy on firms and its comparison with trade policy. Recent work has started to examine international trade as a driver of product choice of firms (Ding 2019; Rachapalli 2021).

The paper is also relevant for macroeconomic studies that stress the importance of input linkages in amplifying micro shocks and policy effects.<sup>10</sup> The development literature emphasizes their role in aggregate productivity and volatility (Koren and Tenreyro 2013) and in motivating policies, such as domestic content requirements, that have interested governments across the developing world (Harrison and Rodríguez-Clare 2010). While we do not look at product linkages across firms, our results for within-firm product linkages demonstrate the existence of cross-product spillovers through inputs. These have been harder to identify across firms because of confounding factors, such as unobserved demand shocks. Looking within firms controls for many of these confounding factors and provides a causal interpretation of shared input capabilities in product choice by drawing on variation driven by policy changes.

The remainder of the paper is organized as follows. Section II contains a description of the context, data, and stylized facts. Section III shows the empirical relationship between input similarity and the industry mix of firms. Section IV presents a model of capability choice and limit-pricing suppliers, deriving structural estimation equations and an instrumentation strategy. Section V contains the results from estimation and quantification of input capabilities. Section VI concludes.

# II. Data and Stylized Facts

## A. Data Description

We use annual data on manufacturing firms from the Annual Survey of Industry (ASI), which is conducted by the Ministry of Statistics and Programme Implementation (MOSPI) of the Government of India. The ASI is the Indian government's main source of industrial statistics on the formal manufacturing sector and consists of two parts: a census of all manufacturing plants that are larger than 100 employees and a random sample of one-fifth of all plants that employ between 20 and 100 workers (between 10 and 100 workers if the plant uses power). The ASI's sampling methodology and product classifications have changed several times over the course

<sup>&</sup>lt;sup>9</sup> In their handbook chapter, Bown and Crowley (2016, 78) summarize that "the existing literature and data sources are not sufficiently developed" to answer key questions such as the extent to which domestic policies affect economic activity and how they compare with trade policy instruments.

<sup>&</sup>lt;sup>10</sup> For example, Acemoglu et al. (2012), di Giovanni, Levchenko, and Méjean (2014), and early work by Jovanovic (1987) and Durlauf (1993).

of its history. In order to ensure consistency, we focus on the time frame of the fiscal years (April–March) 2000–2001 to 2009–10.

The ASI has two unique aspects that make it particularly suitable for our analysis. First, it contains detailed information on both intermediate inputs and outputs, hence allowing us to link the firm's input characteristics to their product-mix decisions. All sales figures include exports, and all purchases include imports. The same product codes are used to describe both inputs and outputs of plants. The data report inputs and outputs at the 5-digit level (of which there are 5,204 codes). To look at the question of production in multiple industries, we aggregate these codes to the 3-digit level, which corresponds to 253 codes, which we call "industries" and take to be our unit of analysis for diversification choices. We focus on 3-digit industries because the purpose is to capture differences in input needs across products. It also avoids the possibility of misclassification, which is more acute at finer levels. Importantly, it keeps our analysis computationally feasible.<sup>11</sup>

The 3-digit industries are in 60 2-digit sectors. To give a sense of the level of detail in this classification, consider the sector "cotton, cotton yarn, and fabrics" (ASIC [ASI commodity code] 63), which has various 3-digit industries, such as "cotton fabrics including cotton hosiery fabrics" (ASIC 633), "made up articles of cotton including apparel" (ASIC 634), and "processing or services of cotton, cotton yarn and fabrics" (ASIC 638). To take another example, the 3-digit industry "stainless steel in primary and finished form" (ASIC 714) is an industry in the sector "iron & steel (incl. stainless steel), and articles thereof" (ASIC 711). A comparison of the dimensionality of products with other sources and descriptive statistics are provided in Boehm, Dhingra, and Morrow (2016, 2018).

Second, the ASI is collected with the definition that the unit of production (factory or factories) must have the same management, combined accounts, and resources that are not separately identifiable. This is particularly well suited for examining the capability (or resource) theory of the firm. But it implies that we pick up plant-wide explanations and not all firm-wide drivers of firm decisions. While we do not have firm identifiers and hence cannot aggregate plants under common ownership, we know that less than 7.5% of all plants are part of a multiplant firm with sister plants that file separate survey returns. With that caveat in mind, we call the units of observation in our data "firms."

## B. The Industry Mix of Indian Manufacturing Firms

We turn to documenting a set of facts related to the industry mix of firms in our sample. This set of facts motivates our subsequent empirical analysis.

<sup>&</sup>lt;sup>11</sup> According to the ASI, the product classification is stratified into 2-digit sectors, 3-digit industries, and 5-digit products.

# 1. Multi-industry Firms Dominate Production

Like their counterparts in the United States and other countries, firms that span multiple industries account for a disproportionately large share of economic activity. Table 1 shows the prevalence of multi-industry firms in our sample. Multi-industry firms account for 32.2% of observations but 62.2% of all sales. Firms that span three or more industries (11.2% of all observations) still account for more than 39% of total sales. This fact is well known and mirrors the results reported by Bernard, Redding, and Schott (2010) for the United States and by Goldberg et al. (2010b) for the set of listed Indian firms.

# 2. Coproduction Is Not Random

We now turn to the question of which industries the firms produce in. Figure 1, in section I, shows two matrices. Figure 1A shows the degree of coproduction between industries. Each row contains the size-weighted average sales shares of plants (in the column industry) that derive the largest share of revenue from products in the row industry. Darker colors indicate higher shares. Hence, by construction, the diagonal contains the highest value in each row. There is substantial coproduction across industries, as indicated by the off-diagonal dark areas. In particular, there is much coproduction occurring within the metal product and machinery manufacturing sectors (the large shaded square on the bottom right) and in the chemicals and pharmaceuticals industries (the industries with indices between 55 and 93), as well as within the textiles and apparel sectors (150–70). Firms from a diverse range of industries (columns 100–112).

No. of Industries	2-Dig	IT SECTOR	s	<b>3-Digit Industries</b>			
	Observations	% Firms	% Sales	Observations	% Firms	% Sales	
1	250,028	81	50	208,881	68	38	
2	43,048	14	28	63,997	21	23	
3	10,113	3	12	22,723	7	14	
4	2,972	1	7	6,843	2	8	
5	864	0	2	2,835	1	6	
6	216	0	1	1,198	0	6	
7	43	0	0	539	0	2	
8	7	0	0	183	0	1	
9	3	0	0	69	0	1	
10 +				26	0	1	

 TABLE 1

 Frequency and Sales Shares of Multi-industry Firms

SOURCE.—Authors' calculations from ASI data.

NOTE.—Observations are firm-years.

These patterns are similar to the coproduction documented by Bernard, Redding, and Schott (2010) for the United States.

Figure 1*B* shows a matrix that captures the similarity of the row and column industries' mix of intermediate inputs. Each element (k, k') is the inner product of the industries' vector of intermediate input expenditure shares:

$$\overline{\text{InputSimilarity}}_{kk'} = \sum_{i} \overline{\theta}_{ik} \overline{\theta}_{ik'},$$

where  $\bar{\theta}_{ik}$  is the sum of expenditure of single-industry firms that produce only *k* on intermediate inputs from *i*, divided by total expenditure of these firms on intermediate inputs. This measure captures the overlap in intermediate-input mixes between industries *k* and *k'*. While not identical, the two matrices look very similar. The metal product and machinery industries all rely on primary metals as inputs; the textiles and apparel industries share a dependence on textile fibers and yarns. Many base chemicals are applicable in different industrial processes. This correlation motivates an examination of firms' input mixes in determining their comparative advantage in the next section.

# III. The Input Mix and Comparative Advantage of Firms

Motivated by the strong positive relationship between coproduction and common use of intermediate inputs at the aggregate level, we focus in particular on the role of firms' intermediate-input mix in explaining revealed comparative advantage. We find that firms' intermediate-input mix explains their subsequent movements in the product space and that these input mixes interact with policy changes to shape revealed comparative advantage. Our regressions motivate a structural model of firm heterogeneity in input-biased productivity, which we present and estimate in section IV. The estimating equation in that model bears a close resemblance to the reduced-form regressions from this section but provides a structural interpretation to the estimated coefficients.

# A. Input Similarity

A natural way to bring the industry-level input similarity from above to the firm level is to consider the inner product of the firm's vector of intermediate-input expenditure shares,  $\theta_{j}$ , with the vector of intermediate-input expenditure shares of an industry *k*:

InputSimilarity<sup>*t*</sup><sub>*jk*</sub> = 
$$\sum_{i=1}^{N} \theta_{iji} \overline{\theta}_{ik}$$
,

where *i* indexes the expenditure shares of spending on 3-digit inputs and t denotes time. We construct the aggregate intermediate-input shares  $\bar{\theta}_{ik}$ by aggregating the microdata of single-industry plants that produce only in industry k. The input similarity measure ranges from zero, when firm jand sector k have no 3-digit inputs in common, to one, when the input expenditure shares of firm j and sector k are identical. The crucial difference between this firm-level input similarity and the aggregate input similarity constructed in section II.B.2 is that this one incorporates idiosyncratic firm-specific variation in input mixes. The firm's input mixes may deviate from the one observed in input-output tables because of the firm producing outputs belonging to multiple industries or because of other sources of variation. This firm-specific variation is quantitatively important: a set of input-output dummies explains only 61% of the overall variation in firm's cost shares  $\theta_{ii}$ . The firm-industry input similarity measure is related to the measure of technological proximity of Bloom, Schankerman, and Van Reenen (2013). Our model in section IV provides a structural interpretation of the measure as the part of firm-level comparative advantage that comes from shared capabilities in intermediate input use.

# B. Estimating the Role of Input Similarity in Industry Adoption

We use the input similarity measure to predict firm movements in the product space. To avoid the possibility that changes in the input mix predate an anticipated change in the product mix, we use the firms' sales and intermediate-input shares at the time of the first observation (and denote the corresponding similarity measure by a "0" superscript).<sup>12</sup> Our baseline specification is a linear model for the probability of firm *j* adding industry *k* between time *t* and *t* + 1:

$$Add_{jkt} = \beta \cdot InputSimilarity_{jk}^{0} + \alpha_{jt} + \alpha_{kt} + \alpha_{kk't} + \varepsilon_{jkt}.$$
 (1)

Here,  $\operatorname{Add}_{jkt}$  is 1 if and only if firm *j* does not produce in industry *k* at time *t* but does so at time t + 1;  $\alpha_{jt}$  is a firm-time fixed effect that captures the average rate of adding industries for each firm-year, leaving the regression to identify only the direction of change in the industry mix and not changes in the number of industries that the firm operates in;  $\alpha_{kt}$  is an industry-time fixed effect that captures any economic changes that determine entry into a particular industry at a particular point in time (such as demand shocks for *k* or input cost shocks that affect all potential *k*-producing firms uniformly). In some specifications, we refine this to industry pair-time

<sup>&</sup>lt;sup>12</sup> That said, the data on reported intermediate-input use in the ASI are the expenditures on intermediate inputs that are being consumed in the current year. Hence, purchases of inventories should not show up in these variables.

fixed effects,  $\alpha_{kk',v}$  with an additional dimension of the firm's industry k' from which it derives the highest fraction of revenue. These effects control for all shocks that might make all firms in industry k' more or less likely to start producing in industry k. Finally,  $\varepsilon_{jkt}$  is an idiosyncratic error term. Appendix A shows summary statistics and correlation tables for all the variables in the regression.

Table 2 shows the results of estimating equation (1), with the inclusion of increasingly stringent fixed effects from left to right. The first specification contains only firm-year fixed effects, thereby estimating the direction of movement in the industry space. The estimated coefficient of the input similarity measure is positive and statistically significant: firms that have an initial input mix that is relatively intensive in inputs that an industry k relies on are more likely to start producing in k (than in the average industry). The second specification additionally includes industry-time fixed effects for every period, which control for any systematic demand or supply shocks that could affect the probability of firms starting to produce in a particular industry. Finally, the third specification of table 2 is very stringent, in that it absorbs the average rate of product adoption for each product k and the main industry of each firm k' (as measured by sales) for each period through  $k \times k' \times t$  fixed effects. This means that any economic shock (supply, demand, technology, infrastructure, etc.) that might affect industry coproduction is accounted for and that what remains are estimates of the direction of intraindustry product changes driven by idiosyncratic input-output linkages of each firm within its main industry. As the table shows, input similarity remains important even in this specification.

Our preferred specification is presented in column 2 of table 2, which controls for annual rates of product adoption at the firm level in addition to annual supply and demand shocks that occur at the product level. A 0.10 unit increase in the similarity between firm and industry

INDUSTRY ENTRY: INPUT SIMILARITY							
	DEPENDENT VARIABLE: Add <sub>jkt</sub>						
	(1)	(2)	(3)				
InputSimilarity $_{jk}^0$	.0391** (.0004)	.0383** (.0004)	.0281** (.0006)				
Firm $\times$ year fixed effects $\alpha_{it}$	Yes	Yes	Yes				
Industry $\times$ year fixed effects $\alpha_{kt}$		Yes					
$k \times k' \times t$ fixed effects $\alpha_{kkt}$			Yes				
$R^2$	.0097	.0117	.0575				
Observations	52,691,029	52,691,029	52,666,907				

TABLE 2 Industry Entry: Input Similarit

NOTE.—Standard errors are in parentheses, clustered at the firm-industry level. \*\* p < .01.

cost shares is associated with a 0.38 pp (or 75%) higher entry probability. In distributional terms, a 1 standard deviation increase in input similarity is associated with a 174% higher industry entry rate.

As in previous work examining product diversification, the results above constitute compelling correlations between firm characteristics and subsequent entry into industries by firms. To establish a causal channel and distinguish from competing theories, we now turn to exploiting a policy change that interacted with the firm's input mix to determine the direction of change in the output industry mix.

## C. Dereservation of Products from Small-Scale Production

Since the 1950s, India has given particular attention to the development of its small-scale industry (SSI) sector, which contributes almost 40% to gross industrial value added and is the second-largest employer after agriculture.<sup>13</sup> Starting in 1967, the government implemented a policy of reservation of certain products for exclusive manufacture by SSI firms. The stated aim of this policy was to ensure employment expansion, to achieve a more equitable distribution of income and "greater mobilization of private sector resources of capital and skills" (Government of India 2009). By the end of 1978, more than 800 products had been reserved; in 1996 it was more than a thousand.

By the early 1990s, the government realized that the reservation policy was inconsistent with the vast liberalization that had begun in the late 1980s and culminated in the new economic policy of 1991. According to the expert committee set up by the government to look into SSI policy, reservation did little to promote small enterprises and had negative consequences by keeping out large enterprises in these products. With free imports of most goods after liberalization, the reservation policy was no longer relevant. It also did not cover the large majority of products manufactured by the small-scale sector. Those industries that were covered, such as light engineering and food processing, were unable to grow and invest in better technologies because of the limitations imposed by SSI reservation. Consequently, the government was repeatedly advised to dereserve products from the SSI list (Hussain 1997). Over the course of the years 1997–2008, the government dereserved almost all products (see table 3). The remaining 20 products were dereserved in 2015.

The definition of SSIs, and therefore the scope of reservation, changed over the period during which the reservation was in place. In 1955, an SSI was defined as an establishment with fixed investments of less than

<sup>&</sup>lt;sup>13</sup> Development Commissioner, Ministry of Micro, Small, and Medium Enterprises, India (2018). Available at http://www.dcmsme.gov.in/publications/reserveditems/resvex.htm.

TABLE 3Dereservation of Products, by Year												
Year	1997	1999	2001	2002	2003	2004	2005	2006	2007	2008	2010	2015
No. of products	15	9	15	51	75	85	108	180	212	107	1	20

SOURCE.-Government of India, Ministry of Micro, Small and Medium Enterprises.

500,000 Indian rupees (Rs) that employed less than 50 workers when working with power or less than 100 workers when not working with power. The employment criterion was dropped in 1960, and the SSI definition was based on the original value of investment in plant and machinery. The investment value was revised over time, and by 1999, the investment ceiling was Rs10 million in plant and machinery (at historical cost).

The impact of the product dereservation on output markets has been thoroughly studied in the literature. The consensus is that the dereservation policy was not systematically related to industry characteristics. In the official report to the government, Hussain (1997, §2.2) states that there was "no explanation in official documents anywhere how the list of reserved items have been selected, . . . the choice of products was somewhat arbitrary." The dereservation policy led to entry of large firms into the dereserved markets, which boosted overall industry output and employment: Martin, Nataraj, and Harrison (2017) find that the aggregate employment response is, on average, above 40%, that output increased by about 30% and wages by 6% and that the number of producers grew by about 13%. Most of the policy response occurred among new firms entering the dereserved product space, rather than old firms adding new products (Amirapu, Gechter, and Smagghue 2018).

In contrast to the existing literature, we use the dereservation as an unexpected change in the conditions that firms face on intermediateinput markets; we are thus looking at firms that are downstream from the dereserved markets. Unit values paid by downstream firms using inputs from dereserved markets drop by about 8%–12% upon dereservation (see the online appendix for full results of regressing log unit values of domestic 5-digit inputs on a dereservation indicator with various fixed effects). We use the policy to obtain variation in input supply that is plausibly exogenous to the production decisions of using firms that were not in the small-scale sector.

# D. Input Similarity Weighted by Dereservation

The official list of dereserved items is taken from the ministry's website and manually matched to 5-digit ASIC products. To define dereservation, let  $\delta_{ijt}$  be 1 if and only if firm *j* at some point uses a 5-digit product in the 3-digit industry *i* that has been dereserved during or before year *t'*. We

	DEPENDENT VARIABLE: Add <sub>jkt</sub>					
	(1)	(2)	(3)	(4)		
InputSimilarity <sup>0</sup> <sub>jk</sub>	.0379** (.0004)	.0371** (.0004)	.0273** (.0006)	.0268**		
InputSimilarity-Dereservation $^{0}_{jkt}$	.0429**	.0424**	.0203**	.0192**		
InputSimilarity-Tariff $^{0}_{jkt}$	(10020)	(	(10021)	$0701^{**}$ (.0095)		
Firm $\times$ year FE $\alpha_{it}$	Yes	Yes	Yes	Yes		
Industry × year FE $\alpha_{kt}$		Yes				
$k \times k' \times t \operatorname{FE} \alpha_{kkt}$			Yes	Yes		
$R^2$	.0098	.0118	.0575	.0576		
Observations	52,691,029	52,691,029	52,666,907	52,666,907		

TABLE 4								
INDUSTRY ENTRY: 7	THE IMPACT OF	DERESERVATION						

Note.—Standard errors are in parentheses, clustered at the firm-industry level. FE = fixed effects.

\*\* *p* < .01.

then interact the similarity measures by these dereservation indicators as follows:

(InputSimilarity-Dereservation)<sup>*t*</sup><sub>*jkt*</sub> = 
$$\sum_{i=1}^{N} \delta_{ijt} \theta_{ijt} \bar{\theta}_{ik}$$
. (2)

This measure "selects" the portion of input industries in the inner product that have been dereserved.

To study how the dereservation interacts with firms' input mix in shaping their comparative advantage, the specification of equation (1) is estimated with the input similarity measure weighted by dereservation. Table 4 shows that the estimated coefficient of the dereservation-weighted input similarity coefficient is positive and statistically significant in all specifications: when input *i* gets dereserved or faces reduced tariffs, firms that have been using *i* intensively are more likely to add products that rely heavily on *i*. Column 4 includes a tariff-change-weighted input similarity measure, analogous to the dereservation-weighted input similarity.<sup>14</sup> Later in the paper, the structural estimation provides a tariff equivalent for dereservation.

To put these numbers in perspective, a firm experiencing a dereservation of one of its inputs gets a shock equivalent to a 0.30 increase in input similarity for its most affected output industry and 0.19 and 0.12 increases for its second- and third-most affected output industries, respectively. A 1 standard deviation increase in the initial input similarity–dereservation measure is associated with an 18% higher industry entry rate, versus a 13% higher industry entry rate for the tariff-based measure, suggesting

<sup>&</sup>lt;sup>14</sup> This is constructed by replacing the dereservation indicator  $\delta_{ijt}$  with the change in India's import tariffs  $\Delta \tau_{ijt}$ . For the precise definition and data description, see app. C.2.

a much larger direct role for removal of entry barriers, compared to input tariff cuts.  $^{\rm 15}$ 

# E. Case Study

Dereservation reduced firms' input prices, and we use the policy to obtain variation in input supply that is plausibly exogenous to the production decisions of using firms that were not in the small-scale sector. The reasoning for using the dereservation policy to study input-based comparative advantage can be motivated by a notable example in comparative advantage driven by better input supply from dereservation.

India is the leading producer, consumer, and exporter of spices in the world and produces 28% of the world's spices. The spice industry in India traditionally specialized in bulk spice commodity production but has now become a world supplier of high-value spice products (including oleoresins, seasonings, sterilized spices, and nutraceuticals). According to the Asian Development Bank, one of the main constraints faced by high-value spice producers is the difficulty in getting high quality and reliable supply of spices, for which they rely on small, unorganized firms.

Spices were reserved for small-scale production till 2008. On October 10, 2008, the government of India dereserved one of the main product categories: ground and processed spices, which serves as an input into several related industries. The National Productivity Council of India documented that the dereservation led to a rise in employment per unit and an expansion in capital investment per unit in the ground and processed spices industry.

Immediately after the dereservation in November 2008, the industry magazine *Spice India* suggested that it is "for the spice industry now to make use of the dereservation" to expand its processing capabilities and to enhance development in high-value-added segments. One of the top five sellers of spice oleoresins in the world is a good example of how the product mix of firms changed with the dereservation of spices.

Headquartered in Cochin, Kerala, the Akay Group is a large Indian firm with sales of over US\$45 million in 2017. It exports mostly to the United States, Europe, and China and is a leading producer of high-value spice products. It initially specialized in food coloring, certain spices, and flavored oil. Following the dereservation, Akay expanded its product offerings to new products that rely heavily on dereserved inputs, such as spiceuticals (spice-based health supplements) and various oleoresins (which are semisolid spice oils such as capsicum oleoresin and cardamom

<sup>&</sup>lt;sup>15</sup> As the paper addresses mechanisms driving product entry of firms, we abstract from indirect input-output linkages that could be important if the question of interest is the aggregate effect of dereservation or tariff policies.

oleoresin). Therefore, building on its earlier product portfolio, Akay has scaled up operations in products that use related dereserved inputs. Similar examples of moving toward spice-intensive products can be found in the ASI data for firms that were in related industries before the dereservation. Therefore, the case study provides a real-world example of the findings from the reduced-form evidence.

## F. Alternative Theories of Product Diversification

While the previous subsections control for a rich set of fixed effects to account for unobserved differences across firms, industries, and industry pairs over time, coproduction could arise because of other rationales suggested by theories of the firm, such as vertical integration and demand complementarities or substitutability. We discuss these in the remainder of this section and find that they are not confounding the findings for shared inputs driving product diversification.

# 1. Vertical Diversification Measures

Firms could diversify up and down their value chain to gain vertical efficiency (e.g., Stigler 1951; Chandler 1992). We use the constructed inputoutput shares  $\bar{\theta}$  to measure whether a sector *k* is upstream or downstream from the firm's current product mix. Let  $\sigma_{zjt}$  denote the sales of firm *j* in industry *z* at time *t*, divided by the total of *j*'s sales at time *t*. Accordingly, we define

Upstream<sup>t</sup><sub>jk</sub> = 
$$\sum_{z=1}^{N} \sigma_{zjt} \overline{\theta}_{kz}$$
, Downstream<sup>t</sup><sub>jk</sub> =  $\sum_{z=1}^{N} \sigma_{zjt} \overline{\theta}_{zk}$ , (3)

where *z* runs over the set of 3-digit industries. To make sense of these definitions, consider the following analogy. Imagine a firm *j* with observed sales shares  $\sigma_{ji}$ . Then, given the firm's output mix  $\sigma_{ji}$  and the industry's average input expenditures for these outputs, one would expect the expenditure share upstream of *j* on *k* to be Upstream<sup>*t*</sup><sub>*jk*</sub>. This measure, for example, is positive for the car components industry when the firm being considered is in the car industry, and the value of the upstream measure rises with the share of car sales of the firm and with the input share of car components in making cars. Likewise, Downstream<sup>*t*</sup><sub>*jk*</sub> is proportional to the expected expenditure share of downstream industry *k* on a firm with the output mix  $\sigma_{jt}$ . It is positive, for example, for the car industry when the firm being considered makes car components, and the value of the downstream measure rises with the firm's sales share in car components and the input share of car components in the car industry.

#### 2. Output Similarity Measures

Firms might also enjoy other complementarities in outputs, by which firms that produce in one industry, or a certain set of industries, are able to obtain relatively higher prices or sales for products from another industry. We construct a measure of output similarity analogously to our input similarity index as an inner product between firm j's sales shares and the aggregate industry k's sales shares:

OutputSimilarity<sup>*t*</sup><sub>*jk*</sub> = 
$$\sum_{i=1}^{N} \sigma_{ijt} \bar{\sigma}_{ik}$$
,

where *i* runs over the set of 3-digit industries. The vector  $\bar{\sigma}_k$  denotes the (size-weighted) average  $\sigma_{ij'}$  among firms j' that derive their highest fraction of revenue from sales in *k*. Again, this measure captures the degree of overlap between firm *j*'s portfolio of sales (across industries) and the average portfolio of firms that sell most in *k*. We also construct an output similarity weighted by the dereservation dummies analogous to the input similarity measure in equation (2).

Output similarity summarizes similar distributions of sales, which would be implied by the horizontal diversification motives of demand-side theories (e.g., Brander and Eaton 1984; Shaked and Sutton 1990; Willig, Salop, and Scherer 1991; Jovanovic and Gilbert 1993; Dhingra 2013; Bernard et al. 2019). If firms diversify to internalize demand complementarities across products, it would show up as higher values of output similarity, and there would then be a positive relationship between the output similarity measure and firm's product diversification. If firms diversify to gain market power by taking over substitutable products, then output similarity would be low, and there would be a negative relationship with firm diversification.

The advantage of the output similarity measure is that it encapsulates various factors determining coproduction in a way that is broadly applicable, relying on information from input-output tables. But like many inferred measures, output similarity embeds both demand- and supply-side motivations.<sup>16</sup> For our purposes, this means that we cannot disentangle the different demand- and supply-side factors embedded in output similarity, but we can assess whether these factors confound the effects of input similarity in determining firm diversification.

<sup>&</sup>lt;sup>16</sup> Product substitutability has also been measured in related work by Broda and Weinstein (2006) and Hoberg and Phillips (2016), who construct demand substitutability measures through, respectively, trade data across various countries (where demand-side parameters are inferred structurally) and computational linguistic methods to determine product similarity from textual product descriptions (which correlate well with demand-side expenditures such as advertising). These methods either are applicable only to exported products or rely on English-language descriptions, which makes them less suited to other settings.

Output and input similarity would be positively correlated with each other when the latter is an important contributor to coproduction, as suggested by figure 1. There is, however, substantial independent variation across the two measures, especially when they are interacted with the dereservation policy. The correlation of the dereservation interactions of input and output similarity is tiny because the dereserved products could not generally be produced by firms (which were not "small," according to the policy).

Table 5 shows the result of estimating equation (1) when controlling for the output similarity variable, the dereservation-weighted version of it, and the two vertical relatedness measures. Firms are also slightly more likely to move upstream from their product mix and slightly less likely to move downstream, showing again a role for inputs in product diversification. The estimated coefficient of output similarity is positive and significant, in particular in the specifications with  $k \times k' \times t$  fixed effects. Putting the numbers in perspective, a 1 standard deviation higher output similarity or upstream linkage is associated with a 62% or 152% higher entry probability, respectively (based on col. 2). A 1 standard deviation rise in the downstream linkage is associated instead with a 38% lower entry probability. The existence of these channels is not entirely surprising

		DEPENDENT VARIABLE: Add <sub>jkt</sub>					
	(1)	(2)	(3)	(4)	(5)		
InputSimilarity <sup>0</sup> <sub>ik</sub>	.0379**	.0251**	.0245**	.0199**	.0195**		
. ,,	(.0004)	(.0004)	(.0005)	(.0006)	(.0006)		
InputSimilarity-Dereservation <sup>0</sup>	.0429**	.0383**	.0378**	.0155**	.0145**		
1 / J <sup>ni</sup>	(.0025)	(.0024)	(.0024)	(.0023)	(.0023)		
OutputSimilarity <sup>0</sup>	· /	.0136**	.0136**	.100**	.100**		
1 · · · · · · · · · · · · · · · / jĸ		(.0006)	(.0006)	(.0018)	(.0018)		
OutputSimilarity-Dereservation <sup>0</sup>		.0344**	.0334**	.0171**	.0171**		
1 / jai		(.0016)	(.0016)	(.0022)	(.0022)		
Upstream <sup><math>0</math></sup> <sub><math>\pi</math></sub>		.0335**	.0315**	.0291**	.0291**		
I JR		(.0009)	(.0009)	(.0030)	(.0030)		
Downstream <sup>0</sup> <sub>#</sub>		00826 **	0076**	0035*	0036*		
jn		(.0007)	(.0006)	(.0014)	(.0014)		
InputSimilarity-Tariff <sup>0</sup>		()	()	(	- 0640**		
P Jki					(.0095)		
Firm $\times$ year FE $\alpha_{i}$	Yes	Yes	Yes	Yes	Yes		
Industry $\times$ year FE $\alpha_{\mu}$			Yes				
$k \times k' \times t FE \alpha_{\mu\nu}$			100	Yes	Ves		
$R^2$	00981	0122	0140	0646	0646		
Observations	52,691,029	52,691,029	52,691,029	52,666,907	52,666,907		

TABLE 5 INDUSTRY ENTRY: ROBUSTNESS

NOTE.—Standard errors are in parentheses, clustered at the firm-industry level. FE = fixed effects.

\* p < .05. \*\* p < .01.

and could be suggestive of other complementarities, such as demand complementarities, motivating product diversification. Most importantly, the estimated coefficients of input similarity and dereservation-weighted input similarity remain positive and statistically significant across different specifications. They are also economically more important than the other measures through higher estimated coefficients and larger values of the measure on average.

# 3. Alternative Output Complementarity Measures

A nonparametric way of capturing demand complementarities is to include a vector of output industry mix × year fixed effects for all observed output combinations (i.e., fixed effects at the industry mix–time level rather than just at the industry pair–time level). Table 6 shows the entry regressions with fixed effects for  $k \times K(j) \times t$  groups, where K(j) is a set of dummies for the mix of goods produced by j at the time of first observation. These specifications show that among all producers of a particular product mix at a given time and for a particular industry k, entry rates are higher for firms that use input bundles more similar to those needed in industry k. The output mix fixed effects absorb time variation in the entry probabilities for combinations of industries, including those arising from

	DEPENDENT VARIABLE: Add <sub>jkt</sub>					
	(1)	(2)	(3)	(4)		
InputSimilarity <sup>0</sup> <sub>ik</sub>	.0145**	.0143**	.0147**	.0145**		
	(.0006)	(.0006)	(.0006)	(.0006)		
InputSimilarity-Dereservation <sup>0</sup> <sub>jkt</sub>	.0125**	.0121**	.0122**	.0118**		
	(.0025)	(.0025)	(.0025)	(.0025)		
InputSimilarity-Tariff <sup>0</sup> <sub>jkt</sub>		0268 **		0269 **		
Ÿ		(.010)		(.010)		
OutputSimilarity $_{ik}^0$			0228 **	0228 **		
			(.0038)	(.0038)		
OutputSimilarity-Dereservation <sup>0</sup> <sub>ikt</sub>			.0066**	.0066**		
<b>-</b> <i>j</i>			(.0025)	(.0025)		
$Upstream_{ik}^{0}$			.0039	.0039		
¥ 9.			(.0080)	(.0080)		
Downstream $_{ik}^{0}$			0000	0000		
			(.0025)	(.0025)		
Firm $\times$ Year FE $\alpha_{it}$	Yes	Yes	Yes	Yes		
$k \times K(j) \times t \operatorname{FE} \alpha_{kK(j)t}$	Yes	Yes	Yes	Yes		
$R^2$	.123	.123	.123	.123		
Observations	47,136,891	47,136,891	47,136,891	47,136,891		

 TABLE 6

 Industry Entry with Output-Mix Indicators

Note.—Standard errors are in parentheses, clustered at the firm-industry level. FE = fixed effects.

3046

\*\* *p* < .01.

demand complementarities. The coefficient of the input similarity measure remains positive and statistically significant. It is slightly smaller than the baseline specification with  $k \times k' \times t$  fixed effects (in col. 3 of table 4), as might be expected because the average input similarity effect is subsumed in the new fixed effects.

Table 6 confirms that input similarity is not confounded with demandside motivations for changes in the output mix of firms as controlled for by the  $k \times K(j) \times t$  groups of all initial output mixes.<sup>17</sup> It also reveals that firms move into industries similar in output to their main industry but dissimilar in output from the rest of their industry mix. If output similarity is interpreted as capturing demand complementarities, then this is consistent with diversification into industries that receive positive demand spillovers from the main industry and into industries that differentiate the firm from its competitors that have the same industry mix.

# 4. Diversification Discount

Agency-based theories of the firm suggest that diversification reduces the value, growth, or productivity of firms and that managers undertake diversification to deploy distressed assets away from core activities. It therefore might be that firm diversification into industries with shared inputs reflects weaker firms moving into new activities that would not be pursued by stronger firms to achieve growth. As a first examination of this weak-firm hypothesis, table 7 estimates the heterogeneous effects across firms on the basis of their initial size, which is often taken as a proxy for stronger performance. We run the baseline entry regressions but interact input similarity (and the dereservation-weighted version) with log sales of the firm at the time of first observation (to proxy for "strong" vs. "weak" firms). The regressions clearly show that larger firms have a stronger correlation between input similarity and industry entry. The negative coefficient on InputSimilarity $_{ik}^{0}$  is completely dominated by the positive coefficient on the interaction, resulting in an almost-zero correlation for the smallest firms (as log sales for the bottom percentile is about 12). The estimated coefficients are very similar to the ones from this table when log sales at time t is used instead of that at the time of first observation. Therefore, there is little evidence to support that weaker firms select into inputsimilar industries.

It may, however, be that firms that diversify into input-similar industries start to experience slower growth, which would again point toward weaker (low-growth) firms selecting into input-similar industries. Table 8 studies

<sup>&</sup>lt;sup>17</sup> See, for instance, Mayer, Melitz, and Ottaviano (2021), who use a similar approach to control for firm-export destination effects.

	DEPENDENT VARIABLE: Add <sub>jkt</sub>						
	(1)	(2)	(3)	(4)	(5)		
InputSimilarity <sup>0</sup> <sub>ik</sub>	0346**	0348 **	0328**	0396**	0382**		
* • • •	(.0029)	(.0028)	(.0029)	(.0029)	(.0030)		
InputSimilarity <sup>0</sup> <sub>#</sub> $\times$ logSales <sup>0</sup> <sub>#</sub>	.0042**	.0041**	.0040**	.0039**	.0038**		
1 7 j <sup>n</sup> 0 jn	(.0002)	(.0002)	(.0002)	(.0002)	(.0002)		
InputSimilarity-Dereservation <sup>0</sup>	0102	00888	00314	0180	0139		
1 , , , , , , , , , , , , , , , , , , ,	(.020)	(.020)	(.020)	(.019)	(.019)		
InputSimilarity-Dereservation $^{0}_{ikt}$ ×							
$\log \text{Sales}^0_{ik}$	.0029*	.0028*	.0024*	$.0020^{+}$	$.0018^{+}$		
o ja	(.0011)	(.0011)	(.0011)	(.0011)	(.0011)		
InputSimilarity-Tariff <sup>0</sup>	. ,		156		0898		
1 / jni			(.10)		(.094)		
InputSimilarity-Tariff <sup>0</sup> <sub>at</sub> $\times$ logSales <sup>0</sup> <sub>at</sub>			.0047		.0024		
1 / j <sup>ni</sup> O j <sup>n</sup>			(.0053)		(.0049)		
Firm $\times$ year FE $\alpha_{ii}$	Yes	Yes	Yes	Yes	Yes		
Industry $\times$ year FE $\alpha_{kl}$		Yes	Yes				
$k \times k' \times t \operatorname{FE} \alpha_{kk'}$				Yes	Yes		
$R^2$	.0100	.0120	.0121	.0577	.0577		
Observations	52,691,029	52,691,029	52,691,029	52,666,907	52,666,907		

 TABLE 7

 Revealed Comparative Advantage: Sales Interactions

NOTE.—Standard errors are in parentheses, clustered at the firm-industry level. FE = fixed effects.

 $\label{eq:posterior} \begin{array}{l} {}^{+} & p < .10. \\ {}^{*} & p < .05. \\ {}^{**} & p < .01. \end{array}$ 

the relationship between postentry sales growth and input similarity, and tables B8–B10 provide further checks with different sets of fixed effects to compare across firms by their entering and continuing industries. In particular, we examine the specification

POSTENTRY GROWTH: WITHIN ENTERING INDUSTRIES								
	Depe log	NDENT VA	ARIABLE: $s_{jk}^{l,,l+3}$ )	Depe log	NDENT VA	ARIABLE: $s_{jk}^{t,,t+5}$ )		
	(1)	(2)	(3)	(4)	(5)	(6)		
InputSimilarity $^0_{jk}$	1.5438** (.0859)	1.4885** (.0864)	1.3314** (.0878)	1.7403** (.0869)	1.6836** (.0875)	1.5259** (.0887)		
InputSimilarity-Dereservation <sup>0</sup> <sub>jkt</sub>	(,	1.4308** (.3349)	1.0711** (.3313)	(,	1.4672** (.3349)	1.1060** (.3275)		
InputSimilarity-Tariff <sup>0</sup> <sub>jkt</sub>		. ,	$-21.5375^{**}$ (1.3242)		. ,	$-21.6295^{**}$ (1.3368)		
Industry $\times$ year FE $\alpha_{kl}$	Yes	Yes	Yes	Yes	Yes	Yes		
$R^2$	.181	.181	.186	.188	.189	.194		
Observations	55,318	55,318	55,318	55,318	55,318	55,318		

 TABLE 8
 Postentry Growth: Within Entering Industries

NOTE.—Dependent variable is the log average sales by firm *j* in industry *k* between t + 1 and t + h for h = 3 or 5 years. Years where the firm is not surveyed are excluded in the calculation of the average. Standard errors are in parentheses, clustered at the firm-industry level. FE = fixed effects.

\*\* *p* < .01.

$$\log\left(\frac{\sum_{\tau=1}^{h} \text{Sales}_{jkt}}{\sum_{\tau=1}^{h} 1(\text{firm } j \text{ observed at } t)}\right) = \beta_0 \text{InputSimilarity}_{jk}^0 + \alpha_{kt} + \varepsilon_{jkt}$$

on the sample of observations (j, k, t), where firm j is entering industry k between t and t + 1 for h = 3 or 5 years. The dependent variable is log average sales of the firm in the industry it entered over a 3- or 5-year horizon, where the average is taken across all years where we observe the firm (note that this may include zeros if the firm has exited the industry during that time window).<sup>18</sup> Table 8 includes industry-year fixed effects and therefore compares firms entering the same industry k at time t that may have produced different outputs before. We find that firms whose input mix is more similar to k have higher postentry sales and not a diversification discount, as predicted by many agency-based theories.

# 5. Other Explanations and Robustness

Within the resource-based view of the firm, our focus has been on shared intermediate inputs because of the policy variation that we can directly leverage. Primary factors, such as labor and capital, could also be shared across industries within the firm. This is explored in full detail in section V, after a simpler exposition with just intermediate inputs.

In appendix B, we report a number of additional results and robustness checks: input similarity shapes revealed comparative advantage not only through industry entry but also through the probability of dropping an industry from the mix and through the intensive margin of production. The probability of dropping an industry falls with input similarity, while sales in an industry rise with input similarity. This suggests that product turnover is not driving the relationship between diversification and input similarity, which is also reaffirmed in the postentry sales growth specifications. We also show that results hold when focusing on (1) the set of large firms  $(\geq 100 \text{ employees})$  that are sampled every year in the ASI, (2) the set of firms that are single-plant firms, (3) the sample when excluding industry pairs (k, k') where there is never any coproduction, (4) the sample excluding few producing firms per industry-year, (5) the sample of firms that exclude those defined as wholesalers, and (6) the sample of firms that are single product to begin with. Finally, the results are also robust to changing the estimator from OLS (ordinary least squares) to logit to better account for the discrete nature of the dependent variable. We conclude that

<sup>&</sup>lt;sup>18</sup> We choose this dependent variable because in any given year, smaller firms are surveyed with only about 20% probability. Given that we are conditioning on entry, sales in year t + 1 are necessarily positive.

input-based comparative advantage is robust to a number of explanations for product diversification proposed by theories of the firms.

The next section investigates these reduced-form findings by building a structural model that explains them and quantifies the role of firmlevel comparative advantage based on input-output mechanisms.

# IV. Theory of the Firm: Product Diversification and Input Similarity

This section presents a theory of multiproduct firms including economies of scope. We focus on the simplest setting, which yields a relationship between policy changes in the input market, supply of inputs, and production choices of multiproduct firms.

The model starts with the primitive of industry-specific production functions, which firms use with their endowed industry-specific productivities. Economies of scope arise because firms can invest in acquiring inputspecific capabilities that can be shared across the industries that they produce in. This generates input-based comparative advantage, which makes firms more likely to produce in industries that share inputs. Increases in the depth of input supply, such as the removal of upstream entry barriers or reductions in input tariffs, operate to heighten these economies of scope. But as a firm goes on expanding its product range, its acquired capabilities get stretched further and the return to comparative advantage declines, as in models of core competencies. This endogenizes the flexiblemanufacturing hypothesis of Eaton and Schmitt (1994), Eckel and Neary (2010), and Mayer, Melitz, and Ottaviano (2014), where unit costs of production rise as firms move away from their core competencies.<sup>19</sup>

The production model allows us to isolate upstream-downstream linkages and their role in multiproduct final-goods production. As we see below, even here the interdependency of suppliers' entry choices and producers' capability choices that allow them to use better-quality suppliers opens up a rich framework. Key to establishing the existence of a supplier equilibrium is avoiding Jevons's paradox, namely, that increased efficiency of an input (as a result of more efficient coal engines) can result in a net increase in demand for the input (coal), causing an outwardly spiraling feedback loop. In our case, increased entry lowers the input costs of downstream firms, spurring further demand and therefore entry. The supplier model also shows that as parameters approach Jevons's paradox, multiple supplier equilibria are possible, precisely as one would get with textbook

<sup>&</sup>lt;sup>19</sup> This also rationalizes the finding of Fontagné, Secchi, and Tomasi (2018) that exporters have typical (essentially median) product vectors that are common across many markets but that there is also considerable fickleness of distinct product baskets across markets (perhaps because of country-specific demand).

falling-average-cost-curve models. This echoes the literature on external scale economies going back to Ethier (1982), although in our case these arise endogenously from a downstream demand response to cutthroat competition by suppliers rather than from a technological assumption.

The model first solves for the optimal pricing and production choices of downstream firms for fixed capabilities, then turns to supplier behavior and establishes their production, pricing, and entry choices. It then determines the capability choice of firms for exogenous industry-time demand shifters.<sup>20</sup> In that sense, the model is not completely closed, remaining agnostic about the relationship of the demand shifters, since the empirics will allow for rich empirical substitution/complementarity patterns over time.

The key insight from the model is that unit costs across industries for multiproduct firms are interdependent through the portfolio of demand a firm faces, because capabilities are chosen to maximize total profits, not minimize costs in any single industry. This extends the pioneering work by Baumol (1977) and Panzar and Willig (1981), as the existence of economies of scope brings in joint-optimization considerations that alter the usual duality results. The framework generates an intuitive estimating equation that mirrors and interprets the input similarity results of the previous section. The portfolio of products a firm produces and the impact that policy changes have on observed portfolios are determined by a firm's optimal input distance from an industry. Distance from an industry carries a productivity penalty, as per the idea of core competency, which examination shows contains the input similarity measure



which can be interpreted in the light of theory. The theory motivates an instrumental variable (IV) strategy that uses common industry-time demand shocks to approximate how endogenous firm revenue shares would change, which maps onto input distance changes (separately from firm-industry-time changes holding last period's capability choices constant). This will allow us, in the next section, to use the structural estimates to quantify entry barriers in terms of equivalent tariffs and to determine the

<sup>&</sup>lt;sup>20</sup> The closest piece to our model of supplier behavior is de Blas and Russ (2015), who use Fréchet cost draws and limit pricing for a discrete number of firms that sell to end consumers, and our setting delivers analogs to their propositions 1 and 2. In contrast, we examine the interplay of supplier entry and downstream feedback effects through demand that creates the possibility of Jevons's paradox and multiple or nonexistence of equilibria while still delivering a tractable model that can be readily estimated.

extent to which input-driven economies of scope explain the portfolios of multiproduct firms.<sup>21</sup>

### A. Demand and Unit Costs

There are a finite number of downstream firms, which are large relative to suppliers. Firm *j* can produce in multiple industries, indexed by *k*. In period *t*, firm *j* pays a fixed cost of  $f_{kt}$  to operate in industry *k* and faces inverse demand in industry *k* of <sup>22</sup>

$$p_{jkt}(q_{jkt}) = D_{kt}q_{jkt}^{\rho-1},$$

where  $p_{jkt}$  are prices,  $q_{jkt}$  are quantities, and  $D_{kt}$  is an industry-time demand shifter. To produce a quantity  $q_{jkt}$  in industry k at time t, firm j combines inputs from industry i,  $M_{ijkt}$ , using a constant-returns-to-scale Cobb-Douglas technology with industry input expenditure shares  $\bar{\theta}_{ik}$  and industry productivity labeled  $\varphi_{jk}$ .<sup>23</sup> As firms pay a fixed cost to produce in any industry each period, we can think of them as "production loci" with firm-industry productivity vector types  $\varphi_j$  that produce different combinations of final goods each period, depending on demand and supply conditions. At input prices  $S_{ijt}\psi_{it}$ , the unit cost of firm j to produce in industry k at time t is therefore

$$c_{jkt} \equiv \prod_{i} \left( \frac{S_{ijt} \psi_{it}}{\overline{\theta}_{ik} \varphi_{jk}} \right)^{\theta_{ik}}$$

Thus,  $c_{jkt}$  is a vector of unit costs that are influenced by input prices and industry productivities.

# B. Capabilities

Inputs  $M_{ijk}$  at the industry level are a composite of quantities  $m_{iijk}$  of varieties, indexed by  $\iota$ ;  $M_{ijk}$  is the CES aggregator of varieties of input i:

$$M_{ijkt}^{(\sigma-1)/\sigma} = \int_0^\infty m_{iijkt}^{(\sigma-1)/\sigma} \, d\iota,$$
(4)

where variety  $\iota$  of input *i* has a price  $s_{\iota i\iota}$ . Firms have capabilities of using inputs with prices  $[\underline{c}_{ij\iota}, \infty)$ , where  $\underline{c}_{ij\iota}$  is chosen by the firm. Here, a lower

<sup>&</sup>lt;sup>21</sup> In what follows, we take all prices in terms of a numeraire input or commodity (e.g., labor).

<sup>&</sup>lt;sup>22</sup> As is well known, this structure can be microfounded with CES (constant elasticity of substitution) preferences over varieties at the industry level. How one chooses to aggregate across industries has implications for the patterns across  $\{D_{u}\}$  each period. We remain agnostic to allow for flexibility in the estimation.

<sup>&</sup>lt;sup>23</sup> In keeping with this section's focus on input capabilities,  $\varphi_{jk}$  could be modeled as a Cobb-Douglas aggregation of firm-input productivities:  $\varphi_{jk} = \prod_i A_{ij}^{\theta_k}$ .

 $\underline{c}_{jit}$  corresponds to both a greater variety of inputs and lower average prices. This can be interpreted as firms screening their input suppliers by choosing a lower cost cutoff for suppliers that they meet. Firms then minimize costs to produce  $M_{ijkt}$  conditional on  $\underline{c}_{ijt}$ .

All firms have an innate capability for inputs from industry *i*,  $\underline{c}_{i0}$ , and can adjust this capability according to demand and supply conditions, subject to a Hicks-neutral cost across production in all industries.<sup>24</sup> Letting  $\underline{c}_{ji}$  denote the vector of acquired capabilities, the unit costs of a multiproduct firm are given by  $\omega(\underline{c}_{ji})c_{jki}$  in each industry, where

$$\omega(\underline{c}_{jt}) \equiv \exp\left\{\sum_{i} \left(\ln \underline{c}_{i0} - \ln \underline{c}_{ijt}\right)^2 / 2\right\}.$$

A firm can use its acquired capabilities across any number of products and reoptimizes by choosing  $\underline{c}_{ijt}$  each period. In order to simplify the subsequent notation, we normalize  $\underline{c}_{i0} = 1.^{25}$ 

# C. Upstream Suppliers

Suppliers can enter input market *i* by paying an entry cost  $f_s$  and receive a cost draw  $b_i$  with  $\Pr(b_i \ge b) = (b/s_m)^{-\lambda}$ , with  $0 < \lambda < 1$ . The resulting mass of entrants is  $N_{it}$ . Suppliers are monopolistically competitive across varieties  $\iota$ , but within varieties inputs are perfect substitutes, and so suppliers engage in limit pricing akin to Bernard et al. (2003). Supplier  $\iota$  chooses price  $s_{iit}$ , supplying a quantity  $m_{iijt}$  to downstream firm j, and earns profits  $\zeta_{iit}$  by selling to any interested firms with the capability to purchase their variety. Since the minimum cost draw among  $N_{it}$  entrants is  $\Pr(b_{iit} \ge b) = (b/s_m)^{-\Omega_u}$ , with  $\Omega_{it} \equiv \lambda N_{it}$ , increases in entry uniformly decrease supplier costs, which are passed on downstream, and prices drop even further from limit pricing. It is noteworthy that supplier depth (per unit mass of entry)  $\lambda$  and the mass of entrants  $N_{it}$  both have similar price-reducing effects through stochastic dominance in lower supplier prices.

## D. Equilibrium

We are mainly concerned with how policies might influence supplier depth and thereby capability choice and input distance with multiproduct firms. We therefore use a partial equilibrium setting in which suppliers enter, receive heterogeneous cost draws, but compete as in a Bertrand model within their variety to supply downstream firms. Downstream firms

<sup>&</sup>lt;sup>24</sup> The innate capability is assumed to be common for econometric reasons. It can be heterogeneous but will then have to be estimated with fixed effects beyond the combination of industry-time.

<sup>&</sup>lt;sup>25</sup> This will not influence our estimating equations, as it is an industry-time effect.

are large, compared to suppliers, and purchase a continuum of input varieties, as determined by their choice of capabilities, to enter across industries and maximize profits in the presence of scope economies across industries. The jointness of the mutiliproduct firm production and sourcing decisions and the external scale economies of suppliers interact to provide a rich setting.<sup>26</sup> Our equilibrium concept is as follows.

DEFINITION. An equilibrium consists of choices in which firms and suppliers compete under monopolistic competition in product markets such that

- 1. Each firm *j* maximizes joint profits, given production technologies,  $\{f_{kd}, \{D_{kd}, \{\psi_{il}\}, \{\bar{\theta}_{ik}\}, \{\varphi_{jk}\}, s_m, \lambda$ , input price distributions  $\Pr(s_{iit} \geq s) = (s/s_m)^{-\lambda N_a}$ , and  $\{N_{ii}\}$  by choosing  $\{q_{jkd}\}, \{M_{ijd}\}, \{\underline{c}_{ijt}\}$ , and  $\{m_{iijd}\}$ .
- Each supplier ι in industry i maximizes profits, given the parameters and distributions above, {N<sub>i</sub>, and downstream firm choices {q<sub>jkd</sub>, {M<sub>ij</sub>}, and {<u>c<sub>ij</sub></u>, by choosing price s<sub>cit</sub>.
- 3. Suppliers enter until expected profits  $E[\zeta_{ii}]$  equal the entry cost  $f_s$ .

We now solve for the optimal behavior of (upstream) suppliers and (downstream) firms.

# E. Unit Costs

In what follows, we assume that  $N_{it} \ge \max\{1 + (\sigma - 2)/\lambda, 1\}$ , which is a continuous analog of having at least two competitors and is sufficient for limit-pricing effects from entry. While supplier entry is endogenous, we characterize when downstream demand from final-goods firms is sufficient to ensure this condition. In this setting, a downstream firm's optimal choice of inputs can be summarized by the following proposition (all proofs are in app. D).

PROPOSITION 1. Assume that  $\Omega_{it} > 1 - \sigma$ , which is necessary for nondegenerate variety choices. Define the cost index of input *i* for firm *j* as  $S_{ijt}$  for costs  $S_{ijt}M_{ijkt}$ .

 The price index for input *i* for firm *j* is a function of capabilities <u>*c*</u><sub>jit</sub> and supplier entry N<sub>it</sub>:

$$S_{ijt}^{1-\sigma} = \frac{\lambda N_{it} \vartheta_{it}}{\lambda N_{it} + \sigma - 1} s_m^{\lambda N_i} \underline{c}_{jit}^{-\lambda N_i + 1-\sigma},$$

<sup>&</sup>lt;sup>26</sup> The model could be further closed with consumers who supply inputs to suppliers in general equilibrium, but those features are not relevant to the subsequent analysis.

where

$$\vartheta_{ii} \equiv 1 + \frac{\sigma - 1}{\lambda(N_{ii} - 1) + \sigma - 1} \left[ 1 - \left(\frac{\sigma - 1}{\sigma}\right)^{\lambda(N_a - 1) + \sigma - 1} \right]$$

- 2. Since  $d \ln S_{ijt}/d \ln \underline{c}_{ijt} = 1 + \lambda N_{it}/(\sigma 1)$ , it follows that when inputs are
  - a. substitutes ( $\sigma > 1$ ), increasing varieties lowers costs (love for variety); or
  - b. complements ( $\sigma < 1$ ), decreasing varieties lowers costs (hate for variety).
- (3) Unit costs  $c_{jkt}$  are given by

$$c_{jkt} = \underbrace{\frac{1}{\varphi_{jk}}}_{\text{firm-industry } (jk)} \underbrace{\prod_{i} \left[ \psi_{il} \left( \frac{\Omega_{il} \vartheta_{il}}{\Omega_{il} + (\sigma - 1)} \right)^{1/(1-\sigma)} \frac{S_{m}^{\Omega_{c}/(1-\sigma)}}{\overline{\theta}_{ik}} \right]^{\theta_{a}}}_{\text{supplier } (kt)} \underbrace{\prod_{i} \left( \underline{C}_{ijt}^{1-\Omega_{c}/(1-\sigma)} \right)^{\overline{\theta}_{a}}}_{\text{capability } (jkt)}$$

We next discuss the interplay between capabilities and supplier entry.

#### F. Supplier Entry

Since  $\rho R_{jk}$  are expenditures by a firm j in industry k and  $\theta_{ik}$  is the share of those expenditures on input i,  $\sum_{j} \sum_{k} \rho \overline{\theta}_{ik} R_{jkt}$  are total expenditures on input i. Expected profits corresponding to these revenues under limit pricing are calculated by multiplying revenues by the aggregate Lerner index  $\mathcal{L}_{jit} = \mathrm{E}[\min\{1/\sigma, 1 - b_1/b_2\}|b_1 \ge \underline{c}_{jit}]$ , where  $b_1/b_2$  is the ratio of the lowest to the second-lowest cost draw of suppliers of the same variety. While the Lerner index varies by firm, the aggregate Lerner index depends only on the intensity of supplier competition at the industry level, even though each firm buys from a fraction  $(s_m/\underline{c}_{jit})^{\lambda N_a}$  of all suppliers. Compared to the usual role of the Lerner index  $1/\sigma$  under CES preferences, this implies that limit pricing is eating into supplier profits and passing the benefits of external scale economies downstream. The aggregate Lerner index is characterized by the following proposition.

**PROPOSITION 2.** The aggregate Lerner index that converts supplier industry revenues weighted by active suppliers into profits,

$$\mathcal{L}_{ii} = \frac{\mathrm{E}[\zeta_{iii}]}{\sum_{j} \sum_{k} \rho \bar{\theta}_{ik} R_{jkl} \left( s_m / \underline{c}_{jii} \right)^{\lambda N_u}}, \qquad (5)$$

decreases in entry, approaches zero, and is given by

$$\mathcal{L}_{ii} = rac{1}{\sigma} \Big(rac{\sigma}{\sigma-1}\Big)^{-\lambda(N_a-1)} + rac{1}{\lambda(N_{ii}-1)+1} \left[1-\Big(rac{\sigma}{\sigma-1}\Big)^{-\lambda(N_a-1)-1}
ight] \ \le \ rac{1}{\sigma-1}.$$

The aggregate Lerner index functions like an average markup that decreases with entry as a result of increased competition from limit pricing. The free-entry condition is expected supplier profits per mass of entrants  $E[\zeta_{it}]/N_{it}$  equal to the entry cost  $f_s$ , which, with equation (5), gives the entry condition:

$$\mathcal{L}_{it} = \frac{f_{\rm s}}{\sum_{j} \sum_{k} \rho \bar{\theta}_{ik} R_{jkt} \left( s_m / \underline{c}_{jit} \right)^{\lambda N_{\rm s}} / N_{it}}.$$
(6)

Equation (6) will obtain if the left-hand and right-hand curves considered as functions of entry cross. The aggregate Lerner index decreases in entry, so this has the potential to hold if the right-hand side increases in entry, that is, if average weighted revenues  $\sum_{j} \sum_{k} \rho \bar{\theta}_{ik} R_{jkl} (s_m / \underline{c}_{jil})^{\lambda N_s} / N_{il}$  decrease.

However, there is no reason this has to be the case. As the depth of suppliers increases, downstream input costs decrease, which spurs further entry. Famously, this occurs under Jevons's paradox, namely, that increased efficiency of an input (due to more efficient coal engines) can result in a net increase in demand for the input (coal), causing an outwardly spiraling feedback loop. In our case, increased entry lowers the input costs of downstream firms, spurring further demand and therefore entry. Even while fixing the complex interrelationship of capability choice with input markets, it is crucial to examine whether this upstream-downstream feedback loop leads to Jevons's paradox. This feedback loop can be understood through profit shifting. Increased entry by suppliers decreases downstream costs through downward price competition that shifts profits into cost savings downstream. This triggers the expansion of both profits and quantities downstream, spurring higher profits for suppliers and therefore entry. Looking at this profit channel, it is intuitive that if the profit taking downstream, as measured by the Lerner index  $1 - \rho$ , is greater than the upper bound  $1/\sigma$  for all upstream Lerner indexes, then the paradox is avoided.27

**PROPOSITION** 3. Holding downstream entry and capabilities constant, supplier entry drives average weighted revenues eventually down to zero, including demand feedbacks from lower input costs, provided that  $(1 - \rho)\sigma \ge 1$ , which guarantees that downstream markups are high enough to absorb the upstream cost reductions.

<sup>&</sup>lt;sup>27</sup> Famously, this occurs for inferior inputs, an idea dating at least to Jevons in 1865, who pointed to a paradox by which increased efficiency of an input (due to more efficient coal engines) can result in a net increase in demand for the input (coal). Hicks later refers to such inputs as regressive inputs, while subsequent work calls them Giffen inputs or inferior inputs (see, e.g., Bear 1965; Portes 1968). In our case, if the "no-Jevons's-paradox condition"  $(1 - \rho)\sigma \ge 1$  does not hold, then for high entry costs or low downstream demand, the market cannot support limit pricing, akin to markets that are too small to support the entry of more than one firm.

Returning to equation (6), one can picture the aggregate Lerner index, drawn across levels of supplier entry, dropping from  $1/(\sigma - 1)$  toward an index of zero, where limit pricing consumes all industry profits per unit mass of suppliers and average weighted revenues fall toward zero, sweeping the right-hand side of equation (6) unboundedly upward. These curves will cross so long as average revenues are high enough, which is true for sufficient downstream effective demand. This crossing need not be unique, as even for fixed downstream entry choices, expected supplier profits  $E[\zeta_{ii}]$  (the product of the aggregate Lerner index and average revenues) are not necessarily decreasing in entry. However, they will be, given the stronger conditions below, which further dampen the feedback loop that is Jevons's paradox. Once downstream entry is considered, then as input costs drop from supplier entry, downstream firms may enter new industries, spiking up revenues abruptly, so for equation (6) to hold, supplier entry must increase, potentially leading to multiple equilibria.<sup>28</sup> We summarize these arguments and provide a sufficient condition for equilibrium holding downstream entry fixed in the following proposition.

PROPOSITION 4. Holding downstream entry and capabilities constant, an equilibrium exists when average weighted revenues are at least  $(\sigma - 1)f_s$  at supplier mass  $N_{it} = 1 + (\sigma - 2)/\lambda$ . Provided that  $\sigma > 1 + \lambda$  and  $(1 - \rho)\sigma \ge 2$ , supplier entry drives average expected profits monotonically down to zero, including demand feedbacks from lower input costs.

If capabilities are instead not held fixed, this implies that entry changes for one input *i* will have cascading effects on input demand for all inputs used in common production, causing changes in supplier entry in all such inputs. Further analytical results would depend on demand structures (see Dhingra and Morrow 2019), and general results are not likely because of the nonmonotonicities just detailed. Just as in models of the location of production, this is to be expected from a rich model of interlinkages that allows for "accidents of history" to occur.

When the condition of proposition 4 holds, equation (6) also provides two comparative statics we appeal to in the empirical specification. First, if supplier entry costs  $f_s$  are reduced as from the removal of entry barriers, the right-hand side of equation (6) shifts out, showing that the equilibrium mass of suppliers increases. Second, suppose that some of the suppliers are foreign and for expositional purposes, supply only to the domestic market. For  $N_{it}$  fixed, if supplier costs are reduced as through a tariff decrease in industry *i*, it is easy to show that the price index  $S_{it}$  decreases, increasing average weighted revenues and again shifting out the right-hand

<sup>&</sup>lt;sup>28</sup> However, the result implies average weighted revenues must decrease to zero barring further entry, leading to an inductive proof of at least one equilibrium with the conditions above.

side of equation (6) and increasing the equilibrium level of supplier entry.<sup>29</sup> This yields the following proposition.

**PROPOSITION 5.** Under the assumptions of proposition 4,

- 1. decreases in supplier entry costs increase supplier entry; and
- 2. decreases in tariffs increase supplier entry.

# G. Capability Choice

As derived above, unit costs are a function of chosen input capabilities, which we now model. Economies of scope arise in this model because firms can use their acquired capabilities across industries. The returns to acquired capabilities, however, decrease as firms become active in more industries. Then, firms have to spread their input capabilities across a larger range of inputs and according to the different factor intensities of their outputs. The acquired capabilities are therefore not as tailored to the needs of each industry as the industry mix gets wider.

*Profits and revenues.*—The profit function of firm j at time t across all industries k is then

$$egin{aligned} \pi_{jt} &= \sum\limits_k \pi_{jkt} \,=\, \sum\limits_k p_{jkt} \, q_{jkt} \,-\, \sum\limits_k \sum\limits_i \omegaig( \underline{c}_{jt} ig) \, S_{it} M_{ijkt} \ &=\, \sum\limits_k ig( D_{kt} q_{jkt}^
ho - \omegaig( \underline{c}_{jt} ig) \, c_{jkt} \, q_{jkt} ig). \end{aligned}$$

A firm's profit-maximizing capability and production choices, considering product markets jointly, are summarized in the following proposition.

PROPOSITION 6. Assume that  $\lambda N_{it} > 1 - \sigma$ . For firm-input expenditure shares  $\theta_{iib}$  the optimal capability choice is

$$\ln \underline{c}_{ijt} = -\Theta_{it}\theta_{ijt},$$

where  $\Theta_{it} \equiv 1 + \lambda N_{it}/(\sigma - 1)$  is the elasticity of input price with respect to capability. Firm-industry revenues are given by

$$\ln R_{jkl} = \underbrace{\ln\left(\rho^{\rho/(1-\rho)}D_{kl}^{1/(1-\rho)}\right)}_{\text{demand }(kl)} - \underbrace{\frac{\rho}{1-\rho}\sum_{i}\overline{\theta}_{ik}\ln\psi_{il}\theta_{il}^{1/(1-\sigma)}\left(1-\Theta_{il}^{-1}\right)^{1/(1-\sigma)}\frac{s_{m}^{\Theta_{u}-1}}{\overline{\theta}_{ik}}}_{\text{supplier }(kl)} + \underbrace{\frac{\rho}{1-\rho}\ln\varphi_{jk}}_{\text{RCA }(jk)} + \underbrace{\frac{\rho}{2(1-\rho)}\sum_{i}\Theta_{il}^{2}\overline{\theta}_{ik}^{2}}_{\text{supplier tech }(kl)} - \underbrace{\frac{\rho}{2(1-\rho)}\sum_{i}\Theta_{il}^{2}\left(\theta_{ijl}-\overline{\theta}_{ik}\right)^{2}}_{\text{input distance }(jkl)},$$
(7)

<sup>29</sup> This can be modeled as a decrease in iceberg transport costs or a first-order stochastic shift downward in the cost distribution, since constant markups and limit pricing will pass this through into lower prices.

with the dimension of variation listed below each term (RCA = revealed comparative advantage).

Since  $\ln \underline{c}_{ijt} = -\Theta_{ii}\theta_{ijt}$ , it follows that firms sourcing from industry *i* increase their range of inputs under love for variety and decrease them under hate for variety. Since competency is costly, firms do not invest in capabilities for inputs they do not source, that is, when  $\theta_{ijt} = 0$ . The addition to equation (7) of comparative advantage is beyond standard models and yields input-based comparative advantage, highlighted by the resource-based theory of the firm, through capability adjustment and different effective pools of suppliers for each firm.<sup>30</sup> The demand and supplier terms can be estimated with industry-time fixed effects, which capture production shifts from the changing demand and supply environment. The RCA terms capture idiosyncratic advantages a firm has across industries that are static and can be estimated with industry-specific combinations of idiosyncratic input productivities.

The remaining comparative-advantage term captures the dynamic redeployment of input capability and is sensitive to the depth of input markets (through  $N_{it}$  in  $\Theta_{it}$ ). To interpret this term, we split it into two parts to highlight the difference between single-product and multiproduct firms. The special case of the "average" single-product firm is useful with  $\theta_{ijt} = \overline{\theta}_{ik}$ , in which case the input-distance term vanishes (the firm is exactly in its "core" and tailors its inputs fully) and only a supplier-technology effect of the benefits from supplier depth by input intensity remains. To the extent that a multiproduct firm deviates from its core competency, this will be reflected in input shares  $\theta_{ijt}$  deviating from each  $\overline{\theta}_{ik}$  and penalizing industries far from the firm's core. In the case of identical supplier depth across markets—that is,  $\Theta_{it} = \Theta$ —the penalty takes the intuitive form of a coefficient times Euclidean distance squared,  $\Sigma_i (\theta_{ijt} - \overline{\theta}_{ik})^2$ .

A final result stemming from profit-maximizing behavior is how firms approximately update their core distance from changes in industry-level demand  $\{D_{kl}\}$ , holding capabilities constant. We use this theory-driven relationship in the instrumentation strategy below to correct for potential biases from unobserved firm-industry-level shocks.

PROPOSITION 7. Input distance can be approximated by considering each firm as holding capabilities constant and optimally updating to respond to industry demand shocks through the relationship

$$\sum_{i} \left( \theta_{ijt} - \overline{\theta}_{ik} \right)^2 \approx \sum_{i} \left( \theta_{ijt-1} - \overline{\theta}_{ik} \right)^2 - \gamma_{kt} \sum_{i} \chi_{jkt-1} \left( \theta_{ijt-1} - \overline{\theta}_{ik} \right)^2,$$

<sup>&</sup>lt;sup>30</sup> Following proposition 6, the effective mass of suppliers a firm chooses is  $N_{it} \cdot (s_m/\underline{c}_{jit})^{\lambda N_a} = N_{it} s_m^{\lambda N_a} e^{\lambda N_a [1+\lambda N_a/(\sigma-1)]\theta_{\theta}}$ . Since the vector of expenditure shares  $\theta_{jt}$  is a function of contemporaneous demand and supply conditions and comparative advantage,  $\theta_{jt}(D_k, N_i, \varphi_j)$ , there is firm heterogeneity in effective suppliers and ranks of supplier purchases from capability choice stemming from the joint production and sourcing decision.

where  $\chi_{jkt}$  are revenue shares of industry *k* for firm *j* in year *t* and  $\gamma_{kt}$  is a common industry demand innovation equal to  $2(D_{kt}/D_{kt-1} - 1)/(1 - \rho)$ .

# H. Estimating Policy Effects

Now consider an observable policy *P* that changes the depth of input markets of the form  $\Omega_{ii} = \lambda N_i = \Omega_{i0} + \alpha_P P_{ii}$ . Linearizing equation (7) around the initial policy state  $\Omega_{i0}$  and letting  $\kappa_x$  represent a fixed effect for characteristic *x* yields the following estimating equation:

$$\ln R_{jkt} = \kappa_{kt} + \kappa_{jk} + \underbrace{\frac{\rho}{1-\rho} \sum_{i} \left[\Theta_{i0}^{2} + \frac{2\Theta_{i0}}{\sigma-1} \alpha_{P} (P_{it} - P_{i0})\right] \left(\theta_{ijt} \bar{\theta}_{ik} - \frac{\theta_{ijt}^{2}}{2}\right)}_{\text{comparative advantage (jkt)}}.$$
 (8)

The theory implies that  $\Theta_{ii}$  has the same sign as  $\sigma - 1$ , so estimating  $\alpha_P \cdot \Theta_{i0}/(\sigma - 1)$  allows for testing hypotheses about the sign of  $\alpha_P$ .

Two policy changes over this period that can be expected to increase the depth of the supplier market are dereservation and tariff changes, which change the number of potential suppliers available. We model these two policy changes as a discrete effect of entry barriers (reservation)  $\alpha_B$  within the 3-digit level (with  $B_{ijt}$  equal to 1 if a 5-digit product the firm ever uses is reserved in industry *i* and 0 otherwise) and a linear effect  $\alpha_\tau$  of tariffs on entry for three digit tariffs  $\tau_{ijt}$  (these are aggregated at the firm level from observed average firm-level imports at the 5-digit level).

For ease of estimation, we impose that all supplier markets have the same depth  $\Omega_{i0} = \overline{\Omega}$ , so that

$$\Omega_{it} = \Omega + \alpha_B B_{ijt} + \alpha_\tau \tau_{ijt}.$$

In light of the theory above, we can interpret these policy shifts as changing the depth of input markets with theory signing both  $\alpha_B$  and  $\alpha_\tau$  to be negative, so that with no entry barriers and zero tariffs,  $\Omega_{i0} = \overline{\Omega}$  is the "maximal" market depth. Therefore, equation (8) approximates around a policy space of no entry barriers and no tariffs. This then implies the estimating equation

$$\ln R_{jkt} = \kappa_{kt} + \kappa_{jk} + \kappa_0 \sum_i \left( \theta_{ijt} \overline{\theta}_{ik} - \frac{\theta_{ijt}^2}{2} \right) + \kappa_1 \sum_i \left( \alpha_B B_{ijt} + \alpha_\tau \tau_{ijt} \right) \left( \theta_{ijt} \overline{\theta}_{ik} - \frac{\theta_{ijt}^2}{2} \right).$$
(9)

with  $\kappa_0 = \Theta_{i0}^2 \rho / (1 - \rho)$  and  $\kappa_1 = 2\Theta_{i0} \rho / (1 - \rho)(\sigma - 1)$ . This estimating equation says that (log) firm revenues depend on industry-time demand and supply effects  $\kappa_{kt}$ , firm-industry effects  $\kappa_{jk}$ , a distance effect of industry *k* from the firm's core competency  $\kappa_0$ , and policy effects that exacerbate
distance for each input *i* through  $\kappa_1 \alpha_B B_{ijt}$  and  $\kappa_1 \alpha_\tau \tau_{ijt}$ . Since deeper supply increases the returns to capabilities and entry barriers and tariffs decrease supplier entry (proposition 5),  $\kappa_1 \alpha_B$  and  $\kappa_1 \alpha_\tau$  should be negative.

The tariff equivalent of dereservation can then be computed from  $\alpha_B \kappa_1 / \alpha_r \kappa_1 = \alpha_B / \alpha_r$ . Because of the selection issues involved in log specifications of sales, we estimate the extensive margin of production implied by equation (9) which accounts for zeros in firm-industry sales. Firms will produce in industry *k* exactly when  $R_{jkl} > (1 - \rho)f_{kl}$ , so we estimate equation (9) as a linear probability model for the outcome that observed revenues of the firm-industry are positive each period.<sup>31</sup> As we are estimating probabilities, we can think of how comparative advantage shifts the production probability frontier of firms.

# I. Structural Instrumentation

In equation (9), firm expenditure shares  $\theta_{iji}$  are a function of timevarying input prices  $\psi_{ib}$  demand shocks  $D_{kb}$ , firm-industry productivities  $\varphi_{jb}$ , and fixed technology  $\overline{\theta}_{ik}$ . Input price and demand shocks are estimated through industry-time fixed effects. Productivities are estimated through firm-industry fixed effects, expressed as RCA. Technology is estimated with a large number of observations, so the risk of measurement error contaminating  $\overline{\theta}_{ik}$  is small, and similarly for demand and input shocks.<sup>32</sup>

There might be omitted variables from our structural equation that cause  $\theta_{ijt}$  to change, which could bias our estimates of the role of capabilities. For example, demand or cost shocks at levels more disaggregated than the firm-industry would change input expenditures and revenues of a firm for reasons other than changes in input capabilities. It can be shown in these two cases, for instance, that bias will exist but run in opposite directions:

 contemporaneous demand shocks *D<sub>jkt</sub>* at the firm level would be positively correlated with input similarity through the composition of firm activity; and

<sup>&</sup>lt;sup>31</sup> This can be naturally extended to an extensive-margin formulation with a logit-type model; see app. D2. We implement this for the structural form as a robustness check but have difficulties with IV-logit as a result of the high-dimensional parameter space and well-known sensitivity of that estimator.

<sup>&</sup>lt;sup>32</sup> One potential concern is that dereservation systematically changes technology  $\bar{\theta}_{ik}$  (such as through price sensitivity of input shares for single-product firms), in which case we could have instrumented for the change in input similarity with the interaction between reservation and initial input similarity, under the assumption that better input supply affects revenues only through the channel of input expenditure shares. Regression coefficients of the percentage of reserved inputs within a 3-digit category on  $\bar{\theta}_{ik}$ , however, have a mean of -0.010 with a standard deviation of 0.017, which is to say about zero in significance and magnitude. Since, in addition, the value percentage of reserved inputs is generally much less than 100%, the implied changes are negligible. See the online appendix for the histogram of estimated coefficients.

• contemporaneous supply shocks  $\psi_{ijkt}$  at the firm level would be negatively correlated with input similarity through the composition of firm activity away from industries intensive in using input *i* (high  $\bar{\theta}_{ik}$ ).

A key econometric insight of proposition 7 is that omitted demand and supply shocks interact with a firm's industry mix, which alters their input use and hence input similarity across industries, potentially introducing bias in estimating economies of scope or policy impacts. The theory motivates a novel IV strategy that uses common industry-time demand shocks to approximate how endogenous firm revenue shares would change input use. The instrumentation strategy is based on the assumption of common industry-level demand innovations  $D_{kt}/D_{kt-1}$  across firms, which can be estimated precisely from the large number of observations and projected onto firm behavior through theory. Recovering these common demand shocks allows us to predict changes in  $\theta_{ii}$  on the basis of shifts in the within-firm distribution of activity.<sup>33</sup> In fact, examining the estimating equation (8), what is needed is not instruments for each  $\theta_{ijt}$  but rather an instrument for terms of the form  $\sum_{i} (\theta_{iji} \bar{\theta}_{ik} - \theta_{ijt}^2/2)$  and  $\sum_{i} (P_{it} - P_{i0}) (\theta_{ijt} \bar{\theta}_{ik} - \theta_{ijt}^2/2)$ . The key insight here is that we need to instrument for comparative advantage and comparative advantage interacted with exogenous policy changes.

Proposition 7 motivates the following instrumentation strategy. The current level of input similarity can be predicted from the levels of the past period, plus a linear approximation of the change in acquired comparative advantage one would expect from common industry demand shocks. Intuitively, this is akin to predicting current input expenditure levels from the previous year (and the RCA they contain) and then projecting them forward one period with a Bartik-type instrument based on input expenditures from the input-output table.<sup>34</sup> In the case of a single

<sup>33</sup> In doing so, we hold the role of capabilities constant in the instrumentation stage to avoid nonlinearity, as the full expression for input similarity is recursive. Even assuming common input markets for all inputs ( $\lambda N_{it} = \Omega$ ), the expression becomes

$$\sum_{i} \overline{\theta}_{ih} \theta_{ijt} = \frac{\sum_{i} \theta_{ih} \sum_{k} \overline{\theta}_{ik} D_{kt}^{1/(1-\rho)} \left( s_{kt} \underline{c_0}^{-[1+\Omega/(\sigma-1)]^2} \underline{\Sigma}^{\overline{\theta}_{k} \theta_{ij}} / \varphi_{jk} \right)^{-\rho/(1-\rho)}}{\sum_{k} D_{kt}^{1/(1-\rho)} \left( s_{kt} \underline{c_0}^{-[1+\Omega/(\sigma-1)]^2} \underline{\Sigma}^{\overline{\theta}_{k} \theta_{ij}} / \varphi_{jk} \right)^{-\rho/(1-\rho)}}$$

with  $s_{kl} \equiv \prod_{i} (\psi_{il} \{\Omega / [\Omega + (\sigma - 1)]\}^{1/(1-\sigma)} s_m^{\Omega/(1-\sigma)} / \overline{\theta}_{ik})^{\overline{\theta}_k}$ .

<sup>34</sup> While this instrumentation strategy works for contemporaneous shocks, if the omitted shocks are serially correlated, they would appear in the previous-period terms that are the basis for the IV strategy, and the exclusion restriction would fail. There are two potential approaches to this problem: either (1) explicitly modeling the serial process, such as shocks being AR(1), or (2) using even further lags in the first stage so that the serial correlation is lower. The former approach is popular but in our case would introduce a nonlinear-in-parameters estimator that presents computational issues. The latter suffers from a large loss in observations, which would select toward larger firms because of the nature of the data sample.

instrument for terms of the form  $(\theta_{ij}\bar{\theta}_{ik} - \theta_{ijt}^2/2)$ , the first stage of an IV strategy following from proposition 7 is then

$$\sum_{i} \left( \theta_{ijl} \overline{\theta}_{ik} - \frac{\theta_{ijl}^{2}}{2} \right) = \kappa_{kl} + \kappa_{jk} + \lambda \sum_{i} \left( \theta_{ijl-1} \overline{\theta}_{ik} - \frac{\theta_{ijl-1}^{2}}{2} \right) - \gamma_{kl} \sum_{i} \chi_{jkl-1} \left( \theta_{ijl-1} \overline{\theta}_{ik} - \frac{\theta_{ijl-1}^{2}}{2} \right).$$
(10)

Equation (10) is composed of three parts: the fixed effects found in the main structural equation for revenues, a lagged term for the endogenous sum  $\sum_i (\theta_{ijt}\bar{\theta}_{ik} - \theta_{ijt}^2/2)$ , and a linear adjustment based on predicted input share changes from lagged revenue shares  $\chi_{jkt-1}$  and contemporaneous industry-level demand shocks  $\gamma_{kt}$ . This last term is essentially a (lagged) sales-weighted "technological distance" measure of the firm away from an industry *k* times the magnitude of the demand innovation that predicts the change in  $\sum_i (\theta_{ijt}\bar{\theta}_{ik} - \theta_{ijt}^2/2)$  between periods.

However, as we need to instrument for both changes in input shares and these input shares interacted with two policy changes, we need three instruments of the type in equation (10), one for the shares and two for their two policy interactions. For this 2SLS (two-stage least squares) estimator, we also need a system that includes all instruments in each firststage prediction equation. Accordingly, define both  $\tilde{\theta}_{ijkl} \equiv (\theta_{ijl}\bar{\theta}_{ik} - \theta_{ijl}^2/2)$ and  $\tilde{\chi}_{ijkl} \equiv \chi_{jkl}(\theta_{ijl}\bar{\theta}_{ik} - \theta_{ijl}^2/2)$  and the following sums for parameters  $\lambda$ and the  $K \times T$  vector  $\gamma$ :

$$egin{aligned} &I^M_{jkt+1} \equiv \lambda \sum_i ilde{ heta}_{ijkt} - \gamma \sum_i ilde{\chi}_{ijkt}, \ &I^B_{jkt+1} \equiv \lambda \sum_i B_{ijt} ilde{ heta}_{ijkt} - \gamma \sum_i B_{ijt} ilde{\chi}_{ijkt}, \ &I^{\pi}_{jkt+1} \equiv \lambda \sum_i au_{ijt} ilde{ heta}_{ijkt} - \gamma \sum_i au_{ijt} ilde{\chi}_{ijkt}. \end{aligned}$$

The resulting first-stage equations for our estimator are as follows:<sup>35</sup>

$$\sum_{i} \tilde{\theta}_{ijkt} = \kappa_{kt} + \kappa_{jk} + I_{jkt}^{M} \left( \lambda^{11}, \gamma^{11} \right) + I_{jkt}^{B} \left( \lambda^{12}, \gamma^{12} \right) + I_{jkt}^{\tau} \left( \lambda^{13}, \gamma^{13} \right) + \eta_{jkt}, \quad (11)$$

$$\sum_{i} B_{ijk} \tilde{\theta}_{ijkt} = \kappa_{kt} + \kappa_{jk} + I_{jkl}^{M} \left( \lambda^{21}, \gamma^{21} \right) + I_{jkt}^{B} \left( \lambda^{22}, \gamma^{22} \right) + I_{jkt}^{\tau} \left( \lambda^{23}, \gamma^{23} \right) + \eta_{jkt}^{B}, \quad (12)$$

$$\sum_{i} \tau_{ijt} \tilde{\theta}_{ijkt} = \kappa_{kt} + \kappa_{jk} + I_{jkt}^{M} \left( \lambda^{31}, \gamma^{31} \right) + I_{jkt}^{B} \left( \lambda^{32}, \gamma^{32} \right) + I_{jkt}^{\tau} \left( \lambda^{33}, \gamma^{33} \right) + \eta_{jkt}^{\tau}.$$
(13)

<sup>35</sup> In practice, sales within a firm-industry group are unlikely to be a balanced panel, as the extensive margin of a firm's industries is liable to change (we in fact model and estimate this with a logit model). Consequently, our one-period-lag strategy may lose some observations, but it reduces the number of parameters that must be estimated simultaneously.

We implement the IV estimator of the structural coefficients in equation (9) as a manual 2SLS estimator, which allows us to calculate the fitted values of the first stage without having to recover the high number of demand innovation coefficients  $\gamma_{kt}$  of the instruments in equations (11)– (13), and accordingly we do not report them. We correct for the wellknown misspecification of the residual variance estimator in manual 2SLS (see chap. 4.2.1 of Angrist and Pischke 2008) and cluster standard errors at the firm-industry level, as proposed by Cameron and Miller (2015). The resulting estimator is equivalent to those obtained through one-stage IV estimation with clustered standard errors.

# V. Results and the Economic Relevance of Input Capabilities

This section first presents our structural estimates of the industry portfolio of firms. The estimates predict which industries firms operate in following policy changes, showing how acquired comparative advantage generates core competencies. We then turn to examples and counterfactuals that demonstrate the role of input capabilities in predicting firm industry scope. A quantification of input-based comparative advantage follows, relative to aggregate industry movements and coproduction rates, before the section ends with an extension of the structural estimates to primary factors.

## A. Structural Estimates

Table 9 shows the OLS and IV estimates for the extensive-margin version of equation (9). The estimated coefficient on the deviation of the input similarity measure is  $\kappa_0 = 0.0086$  in the OLS, which rises to 0.1630 in the IV.<sup>36</sup> The policy coefficient of interest for the entry barriers is  $\kappa_1 \alpha_B = -0.0004$  in the OLS, which increases in magnitude to -0.0016 in the IV. Comparing this with the coefficient on tariffs interacted with the input similarity deviation,  $\kappa_1 \alpha_\tau = -0.0168$ , the effect of entry barriers is a tenth as large. Both entry barriers and higher tariffs reduce the role of input distance, since fewer suppliers disincentivize investing in input capabilities. The tariff equivalent of dereservation is then  $\alpha_B/\alpha_\tau = 0.0016/0.0168 = 0.095$ . Entry barriers from reservation of inputs for small-scale firms therefore lower industry adoption, and their estimated effect is equivalent to a 9.5% tariff on inputs.

<sup>&</sup>lt;sup>36</sup> For the IV sample, the OLS coefficients for the right-hand-side variables are similar: 0.0092 (0.0002), -0.0004 (0.0002), and -0.0001 (0.0004) for  $\kappa_0$ ,  $\kappa_1\alpha_B$ , and  $\kappa_1\alpha_7$ , respectively. Relevant summary statistics are in table A3.

	POSITIVE SAI	LES FOR PLAN	f <i>j</i> in Industi	$\operatorname{RY} k \ (R_{jkt} > 0)$
	(1)	(2)	(3)	(4)
Input distance $\Sigma_i(\theta_{iil}\overline{\theta}_{ik} - \theta_{iil}^2/2)$	.0085**	.0086**	.1362**	.1630**
L	(.0002)	(.0002)	(.0229)	(.0226)
Input distance $\sum_{i} B_{iit} \cdot (\theta_{iit} \overline{\theta}_{ik} - \theta_{iit}^2/2),$				
entry barriers	0004 **	0004 **	0016 **	0016**
,	(.0001)	(.0001)	(.0004)	(.0004)
Input distance $\sum_{i} \tau_{iit} \cdot (\theta_{iit} \overline{\theta}_{ik} - \theta_{iit}^2/2),$				
tariffs		0005		0168 **
		(.0003)		(.0027)
K <sub>ik</sub>	Yes	Yes	Yes	Yes
K <sub>kt</sub>	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	IV	IV
Observations	77,745,382	77,745,382	46,185,150	46,185,150
$R^2$	.762	.762	.760	.760

TABLE 9	
STRUCTURAL ESTIMATES FOR MULTI-INDUSTRY SALES I	Premium

NOTE.—Standard errors are in parentheses, clustered at the firm-industry level. \*\* p < .01.

### B. Input-Based Comparative-Advantage Estimates

The structural estimates can be used to quantify the importance of input capabilities in shaping firm-level comparative advantage. Input-based comparative advantage (CA) can be summarized by the premium arising from input linkages in the production probability frontier. By expanding  $\ln R_{jkt}$ , taking out industry-time, industry-firm, and firm-time fixed effects, we have

$$CA_{jkl} \equiv \hat{\kappa}_0 \sum_i \bar{\theta}_{ik} \theta_{ijl} + \hat{\kappa}_1 \sum_i \left( \hat{\alpha}_B B_{ijl} + \hat{\alpha}_\tau \tau_{ijl} \right) \bar{\theta}_{ik} \theta_{ijl}, \qquad (14)$$

where parameters with a hat denote our IV estimates of the parameters. Note that because of fixed effects, these estimates are within firm-industry, so they are inferred from shifts in comparative advantage, and they are also within industry-time, so they measure shifts relative to other firms in an industry. Therefore, this measure captures movements in comparative advantage.

Table 10 shows summary statistics of CA for firms across industries they produce in by sales rank. CA increases a firm's production probability by 4.3 pp on average, relative to that predicted by industry-wide demand and supply factors (captured in the industry-time fixed effects). For the top-tenth percentile, this increase is more than 13 pp. On average, CA is higher for single-industry firms because they can choose their input capabilities in a way that is tailored to their industry. In line with the model, CA decreases as firms are active in more industries, since firms have to spread their input capabilities across a larger range of inputs and factor intensities.

Industry Rank	Observations	Mean	10th Percentile	90th Percentile			
1	307,294	.054	.004	.153			
2	98,413	.026	.001	.071			
3	34,416	.017	.000	.040			
4	11,693	.013	.000	.032			
5	4,850	.011	.000	.028			
6	2,015	.010	.000	.028			
7	817	.009	.000	.024			
8	278	.009	.000	.024			
9	95	.008	.001	.018			
10 +	38	.005	.000	.010			
Total	459,909	.043	.002	.132			

TABLE 10 CA by Industry Sales Rank

We now study CA for industries that firms do not produce in, which is the additional probability that a firm would produce in a new industry by virtue of their input capabilities, holding fixed their capability choice. Since the space of inputs is large and many industries will not have inputs in common with the firm, CA is often close to zero for any given firm and industry. But for more input-similar firm-industry combinations, as suggested by figure 1, CA is economically significant. Table 11 contrasts the average CA for single-product firms in three industries they might enter. Single-industry firms in the edible fruits and nuts/edible vegetables industry (code 121), on average, enjoy a CA in the fruit and vegetable juices industry (135) of 8.5 pp, whereas the single-industry firms in the (perhaps technologically more similar) industry of soft drinks and mineral water (152) would, on average, get only a 0.6 pp premium. In this example, the edible fruits and nuts/edible vegetables industry is upstream to the fruit and vegetables juices industry and may therefore share intermediate inputs. Many industry pairs where CA is economically relevant, however,

CA FOR THE SECOND INDUSTRY	
Industry (3-Digit ASI Code)	CA (pp)
Fruit and vegetable juices (135):	
Edible fruits and nuts, edible vegetables (121)	8.5
Soft drinks and mineral water (152)	.6
Animal oils and fats (115):	
Other produce of animal origin (119)	5.3
Vegetable oils and fats (125)	1.1
Leather bags, purses, etc. (441):	
Leather footwear (443)	6.8
Plastic footwear (423)	.4

TABLE 11 CA for the Second Industry

Note.—Average firm-level comparative advantage among singleindustry plants of two contrasting industries for the italicized industry. "Other produce of animal origin" covers mostly bone, horn, and meals thereof.

are not vertically related. Consider the leather bags and purses industry (441), which is not vertically related to either leather footwear (443) or plastic footwear (423). Given the leather footwear industry's shared input use of leather with the leather bags and purses industry, its premium is 6.8 pp, whereas the plastic footwear industry's premium is only 0.4 pp. Table E1 states the average CA with the highest premium for 25 industries. It shows that the examples below are not outliers: in many industries, input capabilities shape firm-level comparative advantage to an extent that is economically relevant to firms.

Table 12 further highlights the core-competencies feature of inputbased comparative advantage. The columns contain the number of industries firms operate in, and the rows contain the firm sales ranking of each industry. For firms that produce in a single industry (top left), tailoring input capabilities to the needs of the industry increases production probabilities by 5.2 pp. Firms that produce in two industries experience a 6 pp premium on their core industry and about half of that, 2.9 pp, on their secondary industry. As firms diversify into more industries, the returns to capabilities for an individual industry decline. This occurs along the rows and the columns, showing that the estimated industry adoption falls for firms that offer a wider industry mix and also for core industries, because the acquired capabilities are less tailored to the needs of a single industry.

Table 12 shows that more diversified multiproduct firms experience lower returns from input-based comparative advantage in percentage terms. This, of course, conceals the large economic magnitudes of premia associated with input-based comparative advantage in more diversified firms, which are much bigger than other firms. To highlight this selection effect, entries in table 13 contain the size-weighted CA of firms. We normalize sales weights by the average sales of a single-product firm in that industry, so that the interpretation is premia weighted by the equivalent

	Number of Industries with Positive Sales									
INDUSTRY RANK	1	2	3	4	5	6	7	8	9	10 +
1	.052	.060	.061	.033	.026	.021	.020	.019	.014	.020
2		.029	.023	.018	.017	.016	.014	.015	.010	.022
3			.019	.015	.013	.013	.011	.011	.014	.015
4				.013	.013	.012	.011	.009	.010	.016
5					.011	.011	.011	.009	.011	.009
6						.010	.010	.010	.010	.006
7							.010	.009	.009	.007
8								.009	.008	.008
9									.008	.009
10 +										.005

 TABLE 12

 Core-Competency Sales Premium (in pp) from CA

	Nume	Number of Industries with Positive Sales (CA Weighted by Size)								y Size)
INDUSTRY RANK	1	2	3	4	5	6	7	8	9	10 +
1	.055	.072	.130	.157	.143	.179	.178	.284	.468	1.727
2		.005	.012	.039	.158	.301	.266	.332	.018	3.499
3			.002	.005	.007	.048	.019	.041	.245	1.375
4				.001	.007	.057	.017	.024	.019	.185
5					.004	.009	.014	.008	.019	.047
6						.004	.007	.008	.006	.011
7							.002	.006	.005	.019
8								.002	.001	.006
9									.005	.004
10 +										.002

TABLE 13 Core-Competency Sales Premium (in pp) from CA—Size Weighted

number of typical single-product firms. The single-industry premium from acquiring capabilities is hardly changed at 5.5 pp, compared to that of the typical single-industry firm. Firms in multiple industries now show large premia even when we move along the rows of core industries for firms that operate in more and more industries. For example, a firm operating in nine industries has a 46.8 pp higher (size-weighted) premium in its core industry, compared to a 7.2 pp core premium for a two-industry firm. Moving down the columns, firms see larger premia on their core products, compared to their peripheral products. The lowest-ranked industries of a firm show small premia of under 1 pp. Examination of the analogous tables for the model extended to factors of production in appendix F shows broadly the same patterns at roughly half the size.

Tables 12 and 13 therefore confirm the core-competencies feature of input-based comparative advantage. Together they show that multiproduct firms experience growth as a result of economies of scope in inputs but that these decline as firms diversify into more and more industries.

## C. Economic Significance

To examine the economic importance of input-based comparative advantage, we compare the distribution of industry-level variation in coproduction, which is captured in the fixed effects, with the model-implied comparative-advantage terms. A large literature in international economics seeks to quantify the importance of alternative drivers of productivity and welfare, such as by decomposing the margins for welfare gains from trade. Taking a similar approach, this subsection decomposes the margins of coproduction within firms. It relates them to potential underlying drivers, such as industry-level demand and supply shocks, which are

being explicitly modeled in the theory and are being picked up in the form of fixed effects in the empirical analysis of sections III and V.

Figure 2*A* compares the distribution of CA, in lighter yellow, to the distribution of (unconditional) coproduction rates within industry-pair cells, in darker blue. Comparative advantage is estimated as in equation (14), while the unconditional coproduction rates capture, for example, demand complementarities such as left shoes being coproduced with right shoes. The distributions overlap substantially, showing that input-based



FIG. 2.—Comparing CA with other drivers of entry. FE = fixed effects.

comparative advantage shifts entry probabilities to an extent that is similar to time-invariant coproduction rates within industry pairs. In other words, CA is as important in driving entry as summary statistics of observed coproduction rates across industries. Further, input-based comparative advantage explains much of the variation in entry probabilities across industry pairs, implying a limited role that confounders could play in altering the contribution of CA toward entry into an industry.

Figure 2*B* compares the distribution of CA, in yellow, to the distribution of entry probabilities within industry-year cells (i.e., what an industryyear fixed effect would pick up), in blue. Although smaller, the overlap continues to be substantive. The quantitative importance of the model-implied CA in driving production choices is therefore between the importance of industry pair–level drivers and that of industry-year shocks.

## D. Extension to Primary Factors

The model and structural estimation can be readily extended to primary factors, such as capital and labor. This extension is motivated by product-level findings of Schott (2004), who shows that countries' within-product specialization reflects factor-based comparative advantage. It is also related to Crozet and Trionfetti (2013), which examines factor intensity and firm exports, and to Fontagné, Secchi, and Tomasi (2018), which examines the typical product vectors of firm exports.

Adding primary factors f in an analogous way to intermediate inputs (details in app. D4), the revenue equation then contains a factor similarity term as follows:

$$\ln R_{jkt} = \kappa_{kt} + \kappa_{jk} + \sum_{i} \left[ \kappa_{0}^{I} + \kappa_{1}^{I} \left( \alpha_{B} B_{ijt} + \alpha_{\tau} \tau_{ijt} \right) \right] \left( \theta_{ijt} \overline{\theta}_{ik} - \frac{\theta_{ijt}^{2}}{2} \right) \\ + \sum_{i} \kappa_{1}^{F} \left( \theta_{jjt} \overline{\theta}_{jk} - \frac{\theta_{jjt}^{2}}{2} \right),$$
(15)

with  $\kappa_0^I = \Theta_{i0}^2 \rho/(1-\rho)$ ,  $\kappa_1^I = 2\Theta_{i0}\rho/(1-\rho)(\sigma-1)$ , and  $\kappa_1^F = 2\Theta_{f0}\rho/(1-\rho)(\sigma-1)$ . The IV estimator is analogous to the one above, with an additional instrument for firm-level factor shares to correct for biases from firm-factor shocks, detailed in appendix D4.

Taking the extended model to data, table 14 shows structural estimates accounting for primary factors capital *K* and labor *L*. It finds similar results, with a slightly smaller coefficient on the input similarity term and minor reductions in the magnitudes of the interaction terms. The results suggest that firms are also more likely to move into industries that have a similar primary factor mix, and they provide some evidence for theories of the firm suggesting coproduction in high-capital-intensity

	Pos	SITIVE SALES	Dummy ( $R_{jkt}$ >	> 0)
	(1)	(2)	(3)	(4)
Input distance $\sum_{i \notin \{K,L\}} (\theta_{iil} \bar{\theta}_{ik} - \theta_{iil}^2/2)$	.0068**	.0068**	.1277**	.1112**
	(.0002)	(.0002)	(.0127)	(.0110)
Factor distance $\sum_{l \in \{K,L\}} (\theta_{lil} \overline{\theta}_{lk} - \theta_{lil}^2/2)$	.0084**	.0084**	.1094**	.0600**
	(.0003)	(.0003)	(.0122)	(.0080)
Input distance $\Sigma_i B_{iit} \cdot (\theta_{iit} \overline{\theta}_{ik} - \theta_{iit}^2/2)$ ,				
entry barriers	0000	0000	0040 **	0036**
,	(.0002)	(.0002)	(.0006)	(.0006)
Input distance $\Sigma_i \tau_{iit} \cdot (\theta_{iit} \overline{\theta}_{ik} - \theta_{iit}^2/2)$ ,				
tariffs		.0008		0210 **
		(.0005)		(.0027)
K <sub>ik</sub>	Yes	Yes	Yes	Yes
Kkt	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	IV	IV
Observations	77,745,382	77,745,382	46,185,150	46,185,150
$R^2$	.7622	.7622	.7598	.7598

TABLE 14		
STRUCTURAL ESTIMATES FOR MULTI-INDUSTRY SALES PREMIUM, W	WITH PRIMARY	Factors

NOTE.—Standard errors are in parentheses, clustered at the firm-industry level. \*\* p < .01.

industries. Tables B4 and B5 provide reduced-form results with primary factors.

## VI. Conclusion

Even though multiproduct firms account for a disproportionately large share of economic activity, systematic theory and evidence examining product diversification are thin. In this paper, we examine the role of common use of input capabilities as a determinant of the evolution of firms' product space. This formalizes the resource-based view of the firm in ways that enable an assessment of its economic significance. We bring this theory to Indian manufacturing data to study the relevance of input capabilities both in reduced form and through structural estimation. We use the removal of size-based entry barriers in input markets to establish a causal channel from input capabilities to the firm's industry mix. Estimating the structural parameters that govern the elasticity of revenue with respect to the capabilities component of cost, we find that input capabilities determine the content of a firm's "core competencies" and that they are quantitatively as important as time-invariant coproduction rates across industries.

A key theoretical insight of our framework is that economies of scope within multiproduct firms imply that production choices and input capabilities are jointly determined. Production choices are interdependent on the relative demand and supply conditions a firm faces, and the portfolio of industries a firm enters depends on the extent of its input similarity to that industry. The theory motivates an IV strategy that shows that input capabilities are quantitatively important in determining the production patterns of firms.

Broadly speaking, the fact that the mechanisms of this paper are quantitatively important underscores that multiproduct firms do not behave like collections of single-product firms. Therefore, in aggregate, industries may respond to policy in ways that will not be captured by singleproduct firm models. Coupled with the obvious role of input-output linkages central to economies of scope shown here, this calls for additional research on these linkages both between firms and at the macroeconomic level to look for policy effects within firms that so far may have been missed.

## Appendix A

### **Descriptive Statistics**

TABLE A1 Summary Statistics (N = 52,637,029)

	Mean	SD	Minimum	Maximum
Industry Add Dummy	.0011	.0324	.00	1
InputSimilarity <sup>0</sup> <sub>ik</sub>	.0110	.0529	.00	1
InputSimilarity-Dereservation <sup>0</sup> <sub>ikt</sub>	.0005	.0088	.00	1
InputSimilarity-Tariff <sup>0</sup> <sub>ikt</sub>	0001	.0023	33	0
OutputSimilarity <sup>0</sup> <sub>ik</sub>	.0038	.0515	.00	1
OutputSimilarity-Dereservation <sup>0</sup> <sub>ikt</sub>	.0005	.0192	.00	1
Upstream <sup>0</sup> <sub>ik</sub>	.0035	.0335	.00	1
Downstream $_{jk}^0$	.0058	.0402	.00	1

	Col	RRELATION MATRIX OF	SIMILARITY INDICES			
	InputSimilarity $^0_{jk}$	OutputSimilarity <sup>0</sup>	InputSimilarity-Dereservation $\int_{jkt}^{0}$	OutputSimilarity- Dereservation <sup>0</sup> <sub>jit</sub>	$\mathrm{Upstream}^{0}_{jk}$	$\operatorname{Downstream}^0_{jk}$
$\operatorname{InputSimilarity}_{\#}^{0}$	1.00					
OutputSimilarity <sup>0</sup>	.37	1.00				
InputSimilarity-Dereservation <sup>0</sup>	.17	.06	1.00			
OutputSimilarity-Dereservation <sup>0</sup> / <sub>#t</sub>	60.	.37	.10	1.00		
Upstream $^{0}_{*}$	.46	.54	.05	.10	1.00	
$\mathbf{Downstream}_{jk}^{0}$	.54	.45	.06	.08	.50	1.00

	INDIC
. A2	F SIMILARITY
TABLE	MATRIX O
	RELATION

STRUCTURAL SUMMARY STATISTICS								
	Count	Mean	SD	Minimum	Maximum			
Indicator $R_{ikt} > 0$ (OLS)	77,745,382	.0059	.08	.00	1			
$\Sigma_i(\overline{\theta}_{ik}\theta_{ijt}-\theta_{ijt}^2/2)$	77,745,382	3585	.15	50	1			
$\Sigma_i B_{il} \cdot (\overline{\theta}_{ik} \theta_{ijl} - \theta_{ijl}^2/2)$	77,745,382	0142	.07	50	0			
$\Sigma_i \tau_{it} \cdot (\overline{\theta}_{ik} \theta_{ijt} - \theta_{ijt}^2/2)$	77,745,382	0067	.02	-1.05	0			
Indicator $R_{ikt} > 0$ (IV)	46,185,150	.0060	.08	.00	1			
$\Sigma_i(\overline{\theta}_{ik}\theta_{ijt}-\theta_{ijt}^2/2)$	46,185,150	3477	.15	50	1			
$\Sigma_i B_{il} \cdot (\overline{\theta}_{ik} \theta_{ijl} - \theta_{ijl}^2/2)$	46,185,150	0128	.06	50	0			
$\Sigma_i \tau_{it} \cdot (\overline{\theta}_{ik} \theta_{ijt} - \theta_{ijt}^2/2)$	46,185,150	0076	.03	83	0			
$\Sigma_i(\overline{\theta}_{ik}\theta_{ijt-1} - \theta_{ijt-1}^2/2)$	46,185,150	3374	.17	-1.46	1			
$\Sigma_i B_{il} \cdot (\overline{\theta}_{ik} \theta_{ijl-1} - \theta_{ijl-1}^2/2)$	46,185,150	0130	.07	-1.00	0			
$\Sigma_i \tau_{il} \cdot (\overline{\theta}_{ik} \theta_{ijl-1} - \theta_{iil-1}^2/2)$	46,185,150	0061	.02	91	0			
$\sum_{i} \chi_{ikt-1} (\overline{\theta}_{ik} - \theta_{ijt-1})^2$	46,185,150	.0000	.00	.00	1			
$\sum_{i} \widetilde{B}_{il} \cdot \chi_{ikl-1} (\overline{\theta}_{ik} - \theta_{ijl-1})^2$	46,185,150	.0011	.02	.00	3			
$\sum_i {\pmb{ au}}_{it} \cdot {\pmb{\chi}}_{jkt-1} (\overline{\pmb{ heta}}_{ik} - {\pmb{ heta}}_{ijt-1})^2$	46,185,150	.0001	.01	.00	1			

TABLE A3

## Appendix B

# **Robustness of Estimates and Further Results**

### B1. Robustness of Industry Add, Drop, and Sales Regressions

Table B1 shows the results of the most stringent specification of the industry addition regressions on particular subsamples. Column 1 shows the benchmark results on the full sample. Column 2 shows results for single-plant firms. Given that the vast majority of plants are single-plant firms, the results are virtually unchanged. Column 3 shows results for the plants that get surveyed every year (what the ASI calls the "census": all plants that have more than 100 employees). Finally, in column 4, we exclude all industries k that never have any coproduction with the main industry (defined as the one where *j* has the highest amount of sales). This removes about 90% of observations from the sample (which always have zeros on the left-hand side). Table B2 shows how the probability to drop an industry from the industry mix is shaped by input similarity. Table B3 shows how log sales are correlated with input similarity with a wide range of input-output linkage controls.

		Dependent V	ARIABLE: Add	kt
	(1)	(2)	(3)	(4)
InputSimilarity <sup>0</sup> <sub>jk</sub>	.0199**	.0196**	.0251**	.0360**
InputSimilarity-Dereservation <sup>0</sup> <sub>jkt</sub>	(.0006) $.0155^{**}$	(.0006) .0146**	(.0009) .0110**	(.0011) $.0214^{**}$
Output Similarity <sup>0</sup>	(.0023) 100**	(.0027) 0970**	(.0028) 119**	(.0042) 0858**
	(.0018)	(.0020)	(.0025)	(.0019)
OutputSimilarity-Dereservation <sup>0</sup> <sub>jkt</sub>	.0171** (.0022)	.0178** (.0025)	.0179** (.0030)	.0147** (.0021)

TABLE B1 **REVEALED COMPARATIVE ADVANTAGE: ROBUSTNESS** 

		Dependent V	ARIABLE: Add	l <sub>jkt</sub>
	(1)	(2)	(3)	(4)
Upstream <sup>0</sup> <sub><math>ik</math></sub>	.0291**	.0219**	.0358**	.0186**
1 J.	(.0030)	(.0032)	(.0041)	(.0035)
Downstream <sup>0</sup> <sub>ik</sub>	0035*	0044 **	00384*	0148 **
J.	(.0014)	(.0016)	(.0020)	(.0029)
Sample	Full	Single-	Census	Coproduction
1		plant firms	plants	industries
Firm $\times$ year fixed effects $\alpha_{ii}$	Yes	Yes	Yes	Yes
$k \times k' \times t$ fixed effects $\alpha_{kk'}$	Yes	Yes	Yes	Yes
$R^2$	.0646	.0660	.0848	.0965
Observations	52,666,907	43,120,945	27,076,486	5,165,511

TABLE B1 (Continued)

NOTE.—Standard errors are in parentheses, clustered at the firm-industry level. \* p < .05. \*\* p < .01.

	Depe	ndent Variable: I	Drop <sub>jkt</sub>
	(1)	(2)	(3)
InputSimilarity <sup>0</sup> <sub>ik</sub>	.0128+	1952**	1391**
1 7 5	(.0078)	(.0126)	(.0139)
InputSimilarityKL <sup>0</sup> <sub>#</sub>	7804**	9606**	6631**
1 Jn	(.0996)	(.1518)	(.1834)
InputSimilarity-Dereservation <sup>0</sup>	4002**	$1326^{+}$	$1442^{+}$
1 , Jui	(.0658)	(.0703)	(.0872)
OutputSimilarity <sup>0</sup>	2492**	2629**	2127**
1 / jn	(.0040)	(.0049)	(.0074)
OutputSimilarity-Dereservation <sup>0</sup>	1312**	1012**	0721**
i , , , , , , , , , , , , , , , , , , ,	(.0095)	(.0110)	(.0137)
Firm $\times$ year fixed effects $\alpha_{it}$	Yes	Yes	Yes
Industry $\times$ year fixed effects $\alpha_{kt}$		Yes	
$k \times k' \times t$ fixed effects $\alpha_{kkt}$			Yes
$R^2$	.536	.572	.656
Observations	159,001	158,920	134,861

TABLE B2 INDUSTRY DROP REGRESSIONS

NOTE.—Standard errors are in parentheses, clustered at the firm-industry level.

 $p^+ p < .10.$ \*\* p < .01.

	Depeni	DENT VARIABLE: log	$Sales_{jkt}$
	(1)	(2)	(3)
InputSimilarity $^{0}_{ik}$	.5261**	1.0434**	.5077**
1 7.54	(.0359)	(.0522)	(.0505)
InputSimilarityKL <sup>0</sup> <sub>ik</sub>	1.8510**	4.8912**	4.9163**
1 7 5"	(.5145)	(.7342)	(.7247)
InputSimilarity-Dereservation <sup>0</sup>	1.2502**	.6725**	.8664**
1 / jni	(.2563)	(.2516)	(.2438)
OutputSimilarity <sup>0</sup>	4.1096**	3.4748**	1.5068**
1 7 jn	(.0186)	(.0222)	(.0244)
OutputSimilarity-Dereservation <sup>0</sup>	3806**	5480 **	2616**
1 / Jni	(.0391)	(.0445)	(.0406)
Firm $\times$ year fixed effects $\alpha_{it}$	Yes	Yes	Yes
Industry $\times$ year fixed effects $\alpha_{kt}$		Yes	
$k \times k' \times t$ fixed effects $\alpha_{kk'}$			Yes
$R^2$	.802	.832	.911
Observations	251,026	250,963	220,613

TABLE B3 Intensive Margin of Sales

NOTE.—Standard errors are in parentheses, clustered at the firm-industry level. \*\* p < .01.

# B2. Robustness of Industry Add Regressions to the Inclusion of Factors

Tables B4 and B5 include analogous measures of factor similarity (capital and labor) in the reduced form, showing similar results.

		De	PENDENT V	ARIABLE: AG	ld <sub>jkt</sub>	
	(1)	(2)	(3)	(4)	(5)	(6)
InputSimilarity <sup>0</sup> <sub>jk</sub>	.0543**	.0543** (.0006)	.0530** (.0006)	.0529** (.0006)	.0377** (.0009)	.0376** (.0009)
InputSimilarityKL $^0_{jk}$	(10000)	.0032** (.0002)	(10000)	.0077** (.0003)	(10000)	.0045** (.0003)
Firm × year FE $\alpha_{jt}$ Industry × year FE $\alpha_{kt}$	Yes	Yes	Yes Yes	Yes Yes	Yes	Yes
$k \times k' \times t$ FE $\alpha_{kkt}$ $R^2$ Observations	.0088 46,387,505	.0088 46,387,505	.0109 46,387,505	.0110 46,387,505	Yes .0572 46,360,567	Yes .0572 46,360,567

 TABLE B4

 Industry Entry Correlations, with Primary Factors

Note.—Standard errors are in parentheses, clustered at the firm-industry level. FE = fixed effects.

\*\* *p* < .01.

		Dependent V.	ARIABLE: Add <sub>jk</sub>	
	(1)	(2)	(3)	(4)
InputSimilarity $_{jk}^0$	.0525**	.0511**	.0365**	.0357**
InputSimilarityKL $^0_{jk}$	.00324**	.00768**	.0045**	.0045**
InputSimilarity-Dereservation $^{0}_{jkt}$	.0713**	.0705**	.0339**	.0325**
InputSimilarity-Tariff $_{jkl}^0$	(.0013)	(.0012)	(.0010)	$0997^{**}$
Firm × year fixed effects $\alpha_{jt}$ Industry × year fixed effects $\alpha_{kt}$	Yes	Yes Yes	Yes	Yes
$k \times k' \times t$ fixed effects $\alpha_{kkt}$ $R^2$ Observations	.0089 46 387 505	.0111	Yes .0573 46 360 567	Yes .0573 46 360 567

 TABLE B5

 Industry Entry, with Primary Factors

Note.—Standard errors are in parentheses, clustered at the firm-industry level. \*\*  $p \le .01$ .

# B3. Robustness of Industry Add Regressions to Logit

Table B6 shows the results of the logit estimation of the industry addition regressions, corresponding to the baseline specifications of table 4.

			DEPENDENT V.	ARIABLE: Add <sub>ju</sub>		
	(1)	(2)	(3)	(4)	(5)	(9)
nputSimilarity $_{jk}^{0}$	6.665**	5.209**	6.585** 7 0997	8.204**	5.136**	5.095**
nputSimilarity-Dereservation $\int_{a_{i}}^{0}$	(100.)	(070.)	(.0 <i>32)</i> 2.134**	(.041) 2.448**	(.011) 1.455**	(.072) 1.388**
, T			(.172)	(.198)	(.247)	(.247)
nputSimilarity-Tariff $^0_{jkt}$						$-3.731^{*}$
Plant × year FE $\alpha_{ji}$	Yes	Yes	Yes	Yes	Yes	(.899) Yes
ndustry × year FE $\alpha_{ht}$ $t \times k \times t$ FE $\alpha_{ht}$		Yes		Yes	Yes	Yes
Istimator	Logit (ML)	Logit (ML)	Logit (ML)	Logit (ML)	Logit (ML)	Logit (ML)
Observations	52,691,029	52,691,029	52,691,029	52,691,029	52,691,029	52,691,029
NOTE.—Standard errors are in r	parentheses, clustered	l at the firm-industry	v level. FF. = fixed ef	fects: MI. = maximu	m likelihood.	

TABLE B6 Revealed Comparative Advantage: Robustness
--

\* p < .05.

#### B4. Robustness of Unit Value Regressions

Table B7 shows results of a regression of log unit values of domestically sourced intermediate inputs (by 5-digit input category *i*) on a dummy that takes the value 1 when input *i* used to be reserved and has been dereserved in the current or a past year. The regressions include either input *i* fixed effects or firm-input fixed effects and therefore show the impact that the dereservation had on average prices paid on *i*. The ASI unit value data are noisy, and we correct for known problems. One particular problem is that from 2005 onward, the magnitudes of reported quantities (and therefore unit values) jump inexplicably by a factor of 100 or 1,000 within firm-input observations. In columns 3 and 4 of table B7, we report results for a sample of "safe" observations where we are pretty sure that this problem is not present to begin with (more precisely, all observations that are within a factor of 90 of the median of the pre-2005 distribution of unit values for that product code).

	I	Dependent VA	ARIABLE: log p	jit
	(1)	(2)	(3)	(4)
$t \ge$ year <i>i</i> was dereserved	0696**	0617**	0450**	0654**
,	(.012)	(.013)	(.012)	(.014)
Sample	All	All	Safe	Safe
Year FE	Yes	Yes	Yes	Yes
Input product fixed effects	Yes		Yes	
Firm $\times$ input product fixed effects		Yes		Yes
$R^2$	.864	.966	.882	.969
Observations	900,003	481,036	761,703	418,361

 TABLE B7

 Domestic Input Unit Values After Dereservation: Robustness

NOTE.—Standard errors are in parentheses, clustered at the firm-industry level. \*\* p < .01.

### **B5.** Diversification Discount: Robustness

The dependent variable in the following tables is log average sales of the firm in the industry it entered over a 3-year or a 5-year horizon, where the average is taken across all years where we observe the firm. Table B8 includes fixed effects for the firms' industry  $\times$  year, so we are comparing entries of firms from the same industry into different industries. We find that entry into industries with more similar input mixes is associated with higher sales performance. Table B9 also includes industry-year fixed effects and finds similar results. Finally, table B10 performs a stringent exercise by comparing firms that are both in the same industry and are entering the same industry. Estimates for unweighted input similarity are closer to zero than in the tables before, but policy-weighted input similarity is still of a similar magnitude and statistically significant. At least for dereservation-induced entry, input similarity is positively associated with postentry performance.

	DEPENDEN	r Variable: log(Avg	$\operatorname{Sales}_{jk}^{t,\dots,t+3}$	DEPENDENT V	VARIABLE: log(Avg S	$\operatorname{ales}_{jk}^{t,\dots,t+5}$
	(1)	(2)	(3)	(4)	(5)	(9)
InputSimilarity $^0_{jk}$	.9713** ( 0899)	.8957**	.7614**	1.1604**	1.0839** / 0836)	.9493**
InputSimilarity-Dereservation $_{jkt}^{0}$	(7700.)	2.1992**	1.8308**		2.2253**	1.8558**
InputSimilarity-Tariff $^0_{ijt}$		(.3333)	$(.3298) -20.0752^{**}$		(.3341)	(.3281) -20.1344**
Firm industry $\times$ vear FF $\alpha_{ii}$	Yes	Ves	(1.3439) Ves	$V_{es}$	Yes	(1.3607)
$R^2$	.159	.159	.164	.166	.167	.172
Observations	55,296	55,296	55,296	55,296	55,296	55,296

surveyed are excluded in the calculation of the average. Standard errors are in parentheses, clustered at the firm-industry level. FE = fixed effects. \*\* p < .01.

	DEPENDEN	T VARIABLE: log(Avg	$\mathrm{Sales}_{jk}^{t,\dots,t+3})$	DEPENDENT V	/ARIABLE: log(Avg S	$ales_{ji}^{t,\dots,t+5}$
	(1)	(2)	(3)	(4)	(5)	(9)
InputSimilarity $_{jk}^{0}$	1.2924**	1.2304**	1.0765**	1.4854**	1.4227**	1.2684**
InputSimilarity-Dereservation $\frac{0}{2n}$	(6100)	(.0000) 1.5378**	(.0090) 1.2308**	(0600.)	$1.5566^{**}$	(.0900) 1.2490**
, J <sup>ns</sup>		(.3355)	(.3301)		(.3391)	(.3317)
InputSimilarity-Tariff $_{ik}^{0}$			$-19.6582^{**}$			$-19.7009^{**}$
			(1.2664)			(1.2797)
Industry $k \times$ year FE $\alpha_{ht}$	Yes	Yes	Yes	Yes	Yes	Yes
Firm industry $k \times \text{year FE} \alpha_{kt}$	Yes	Yes	Yes	Yes	Yes	Yes
$R^2$ , .	.259	.260	.264	.266	.267	.271
Observations	55,163	55,163	55,163	55,163	55,163	55,163
NoTE.—Dependent variable is the	he log average sales	by firm j in industry	k between $t + 1$ and	$t + h$ for $h = 3$ or $\frac{1}{2}$	by years. Years where	the firm is not

TABLE B9	POSTENTRY GROWTH: WITHIN CONTINUING AND ENTERING INDUSTRIES
----------	---

surveyed are excluded in the calculation of the average. Standard errors are in parentheses, clustered at the firm-industry level. FE = fixed effects. \*\* p < .01.

	DEPENDEN	T VARIABLE: log(Avg	$\operatorname{Sales}_{jk}^{t, \dots, t+3})$	DEPENDENT	VARIABLE: log(Avg 5	$\operatorname{sales}_{jk}^{t, \dots, t+5}$
	(1)	(2)	(3)	(4)	(5)	(9)
InputSimilarity $_{jk}^{0}$	.2185+	.1549	.0125	.3344**	.2734*	.1317
InputSimilarity-Dereservation $^0_{\mu\iota}$	(0671.)	(.1271) 1.6002**	$1.4386^{**}$	(0071.)	(.1501) 1.5331**	(.1202) 1.3723**
		(.3937)	(.3868)		(.3947)	(.3877)
InputSimilarity-Tariff $_{ju}^{v}$			-17.4555** $(1.4814)$			$-17.3741^{**}$ (1.4820)
$k  imes k  imes t \operatorname{FE} lpha_{\scriptscriptstyle h \! k t}$	Yes	Yes	Yes	Yes	Yes	Yes
$R^2$	.445	.445	.448	.450	.450	.453
Observations	39,695	39,695	39,695	39,695	39,695	39,695
Norr.—Dependent variable is th surveyed are excluded in the calcul p < .10. * $p < .05$ . ** $p < .01$ .	e log average sale lation of the avera	s by firm <i>j</i> in industry ge. Standard errors a	/k between $t + 1$ and re in parentheses, clu	t + h for $h = 3$ or istered at the firm-in	5 years. Years where ndustry level. FE =	the firm is not fixed effects.

TABLE B10	STENTRY GROWTH: WITHIN-INDUSTRY FAIRS
-----------	---------------------------------------

# Appendix C

# Data

- C1. Data Sources
- 1. Manufacturing Plant Data

Our manufacturing plant data come from the "detailed unit-level data with factory identifier" of the Indian ASI for the years 2000–2001 to 2009–10. The data can be obtained by writing to ASI Processing and Report (Deputy Director General, CSO (IS Wing) 1, Council House Street, Kolkata; email: asidata.cc-mospi@ gov.in).

# 2. Tariff Data

The Indian import tariff data come from UNCTAD-TRAINS (United Nations Conference on Trade and Development–Trade Analysis and Information System, accessed through WITS [World Integrated Trade Solution]; http://wits.world bank.org/).

## 3. Dereservation Data

Notices of dereservation of products are from the website of the Development Commissioner, Ministry of Micro, Small, and Medium Enterprises. (http://www.dcmsme.gov.in/publications/reserveditems/resvex.htm). We manually concord the product codes to 5-digit ASIC codes on the basis of the text description of the dereserved items.

# C2. Variable Definitions

- Add dummies Add<sub>jkl</sub>: equal 1 if and only if j does not produce any product in 3-digit industry k at time t and does produce a product in k at time t + 1. We exclude outputs with zero or missing sales from the set of produced products.
- Drop dummies Drop<sub>jkl</sub>: equal 1 if and only if j does produce a product in 3-digit industry k at time t and does not produce any product in k at time t + 1. We exclude outputs with zero or missing sales from the set of produced products.
- Sales<sub>*jki*</sub>: *j*'s total sales of products in 3-digit industry *k* at time *t*, including exports.
- Plant expenditure shares  $\theta_{iji}$ : expenditure on intermediate inputs in 3-digit category *i* by *j* at time *t*, divided by total expenditure on individually listed intermediate inputs of *j* at time *t*. These listed intermediate inputs include all agricultural, mining, and manufacturing products that are being consumed in the production process (including imports) during the current period and exclude energy and services inputs.

- Aggregate expenditure shares  $\bar{\theta}_{ik}$ : sum of expenditures of single-industry plants that produce only products in 3-digit industry *k* on intermediate inputs from 3-digit category *i*, divided by total expenditure of these plants on individually listed intermediate inputs (including imports).
- Plant sales shares  $\sigma_{kjb} \chi_{jkc}$ ; plant *j*'s total gross sales revenue of products in 3-digit category *k* divided by *j*'s gross sales of individually listed physical outputs (which excludes revenue from services, renting out capital, interest, etc.), both at time *t*, including exports.
- Aggregate sales shares  $\bar{\sigma}_{ik}$ ,  $\bar{\chi}_{ik}$ : total gross sales in 3-digit category *i* of plants that derive the highest fraction of their revenue from sales of products in 3-digit category *k*, divided by total gross sales of individually listed physical outputs of these plants, including exports.
- Dereservation dummies  $\delta_{ijt}$  and  $B_{ijt}$  equal 1 if and only if there is a 5-digit input in the 3-digit basket *i* that has been dereserved during or before *t* and shows up at some point in *j*'s basket of intermediate inputs. In section IV, the reservation dummy  $B_{ijt}$  equals 1 when there is 5-digit product in the 3-digit basket *i* that the firm is using at some point and that is reserved at time *t*.
- Tariff change  $\Delta \tau_{iji}$ : difference between year *t* Indian import tariff and year 2000 tariff on 5-digit products in 3-digit category *i*, weighted by *j*'s average expenditure share on 5-digit imports in *i*. We concord tariffs from the 6-digit Harmonized System (HS) codes reported by TRAINS to ASIC codes via the ASIC 2009/10–NPCMS (National Product Classification for Manufacturing Sector) concordance published by MOSPI, and the CPC (Central Product Classification)–HS concordance published by UNStats (United Nations Statistics Division; the first five digits of NPCMS are CPC v2.0 codes). Tariffs are effective applied tariffs where available and most-favored-nation tariffs otherwise. We focus on nonagricultural tariffs to avoid endogeneity concerns with agricultural tariffs, which often vary as a result of policy responses to domestic economic conditions that can affect firm sales directly. In section IV,  $\tau_{iji}$  is defined analogously as the level of that tariff.
- Input similarity: Input Similarity'\_{jk} =  $\sum_{i=1}^{N} \theta_{iji} \bar{\theta}_{ik}$ , where N is the number of 3-digit industries.
- Output similarity: OutputSimilarity<sup>t</sup><sub>jk</sub>  $\equiv \sum_{i=1}^{N} \sigma_{ijt} \bar{\sigma}_{ik}$ .
- Input similarity weighted by policy changes:

InputSimilarity-Dereservation<sup>*l*</sup><sub>*jk*</sub> 
$$\equiv \sum_{i=1}^{N} \delta_{ijt} \theta_{ijt} \overline{\theta}_{ik}$$
,  
InputSimilarity-Tariff<sup>*l*</sup><sub>*jk*</sub>  $\equiv \sum_{i=1}^{N} \Delta \tau_{ijt} \theta_{ijt} \overline{\theta}_{ik}$ .

• Output similarity weighted by a policy change:

OutputSimilarity-Dereservation<sup>t</sup><sub>jk</sub> 
$$\equiv \sum_{i=1}^{N} \delta_{ijt} \sigma_{ijt} \bar{\sigma}_{ik},$$
  
OutputSimilarity-Tariff<sup>t</sup><sub>jk</sub>  $\equiv \sum_{i=1}^{N} \Delta \tau_{ijt} \sigma_{ijt} \bar{\sigma}_{ik}.$ 

• Upstream and downstream:

Upstream<sup>t</sup><sub>jk</sub> = 
$$\sum_{i=1}^{N} \sigma_{ijl} \bar{\theta}_{ki}$$
, Downstream<sup>t</sup><sub>jk</sub> =  $\sum_{i=1}^{N} \sigma_{jil} \bar{\theta}_{ki}$ .

#### C3. Sample Definition

Our sample consists of all plant-year observations between 2000–2001 and 2009– 10 that report being in operation and that report both physical intermediate inputs and outputs.

## Appendix D

# Theory

- D1. Firm Input Choice
- 1. Proof of Proposition 1

Final-goods firms purchase from the lowest-price supplier. If  $b_1$  and  $b_2$  are the lowest and second-lowest supplier-cost draws, then the price  $s_{iit}$  charged to firms will be either the monopolistically competitive markup over  $b_1$  or the limit price  $b_2$ , implying  $s_{iit} = \min\{\sigma b_1/(\sigma - 1), b_2\}$ . Letting *G* denote the cdf (cumulative distribution function) of supplier cost draws and suppressing *i*, *j*, and *t* subscripts, a general positive moment of  $S_{ijt}^{\alpha}$  is

$$\begin{split} \mathbf{E}[S^{\alpha}] &= \mathbf{E}[S^{\alpha}|\sigma b_{1}/(\sigma-1) \leq b_{2}] + \mathbf{E}[S^{\alpha}|\sigma b_{1}/(\sigma-1) > b_{2}] \\ &= \int_{\underline{c}}^{\infty} \left(\frac{\sigma}{\sigma-1} b_{1}\right)^{\alpha} \int_{\underline{s}=1}^{\infty} \frac{N! \ b_{2}^{-\lambda(N-2)} s_{m}^{\lambda(N-2)}}{(N-2)!} \ dG(b_{2}) \ dG(b_{1}) \\ &+ \int_{\underline{c}}^{\infty} \left(\frac{\sigma}{\sigma-1} b_{1}\right)^{\alpha} \int_{\underline{s}=1}^{\infty} \frac{N! \ b_{2}^{-\lambda(N-2)} s_{m}^{\lambda(N-2)}}{(N-2)!} \ dG(b_{2}) \ dG(b_{1}) \\ &= \int_{\underline{c}}^{\infty} \left(\frac{\sigma}{\sigma-1} b_{1}\right)^{\alpha} \int_{\underline{s}=1}^{\infty} \frac{(\lambda b_{2}^{-\lambda-1})}{s_{m}^{-\lambda}} \frac{N! \ b_{2}^{-\lambda(N-2)} s_{m}^{\lambda(N-2)}}{(N-2)!} \ db_{2} \ dG(b_{1}) \\ &+ \int_{\underline{c}}^{\infty} \int_{h}^{\frac{s}{\sigma-1}} b_{2}^{\alpha} \left(\frac{\lambda b_{2}^{-\lambda-1}}{s_{m}^{-\lambda}}\right) \frac{N! \ b_{2}^{-\lambda(N-2)} s_{m}^{\lambda(N-2)}}{(N-2)!} \ db_{2} \ dG(b_{1}) \\ &+ \int_{\underline{c}}^{\infty} \int_{h}^{\frac{s}{\sigma-1}} b_{2}^{\alpha} \left(\frac{\lambda b_{2}^{-\lambda-1}}{s_{m}^{-\lambda}}\right) \frac{N! \ b_{2}^{-\lambda(N-2)} s_{m}^{\lambda(N-2)}}{(N-2)!} \ db_{2} \ dG(b_{1}) \\ &= \int_{\underline{c}}^{\infty} \left(\frac{\sigma}{\sigma-1} b_{1}\right)^{\alpha} \int_{\underline{s}=1}^{\infty} \frac{N! \ b_{2}^{-\lambda(N-1)} s_{m}^{\lambda(N-1)}}{(N-2)!} \ db_{2} \ dG(b_{1}) \\ &+ \int_{\underline{c}}^{\infty} \int_{h}^{\frac{s}{\sigma-1}} \frac{N! \ b_{2}^{-\lambda(N-1)-1} s_{m}^{\lambda(N-1)}}{(N-2)!} \ db_{2} \ dG(b_{1}) \\ &+ \int_{\underline{c}}^{\infty} N s_{m}^{\lambda(N-1)} \left(\frac{\sigma}{\sigma-1}\right)^{\alpha-\lambda(N-1)} b_{1}^{\alpha-\lambda(N-1)-1} \ dG(b_{1}) \\ &+ \int_{\underline{c}}^{\infty} N s_{m}^{\lambda(N-1)} \frac{\lambda(N-1)}{\lambda(N-1)-\alpha} \left[1 - \left(\frac{\sigma}{\sigma-1}\right)^{\alpha-\lambda(N-1)}\right] \right\} b_{1}^{\alpha-\lambda,N-1} \ db_{1} \\ &= \frac{\lambda N}{\lambda N-\alpha} s_{m}^{\lambda N} \left\{ \left(\frac{\sigma}{\sigma-1}\right)^{\alpha-\lambda(N-1)} + \frac{\lambda(N-1)}{\lambda(N-1)-\alpha} \left[1 - \left(\frac{\sigma}{\sigma-1}\right)^{\alpha-\lambda(N-1)}\right] \right\} c_{\alpha}^{\alpha-\lambda N} \\ &= \frac{\lambda N}{\lambda N-\alpha} s_{m}^{\lambda N} \frac{c}{\alpha} - \lambda^{\lambda(N-\alpha}} \left\{1 - \frac{\alpha}{\lambda(N-1)-\alpha} \left[1 - \left(\frac{\sigma}{\sigma-1}\right)^{-\lambda(N-1)+\alpha}\right] \right\}, \end{split}$$

which gives the expression above for  $\alpha = 1 - \sigma$ . The Cobb-Douglas cost index is straightforward. QED

## 2. Proof of Proposition 2

What we want is the index that converts supplier revenues to profits under limit pricing. Intuitively, supplier revenues convert to profits at a rate of price minus cost over price, which is  $1/\sigma$  under monopolistic pricing, and when the lowest-cost supplier has marginal cost  $b_1$  and must undercut the second-lowest-cost producer at marginal cost  $b_2$ , this becomes  $(b_2 - b_1)/b_2$ . The aggregate Lerner index under monopolistic pricing for a firm with capability  $\underline{c}_{jil}$ , letting *G* denote the cdf of supplier cost draws, is therefore

$$\begin{split} \int_{\underline{c}_{\mu}}^{\infty} \int_{\frac{\sigma}{\sigma+1}b_{1}}^{\infty} \frac{1}{\sigma} \frac{N_{il}! \ b_{2}^{-\lambda(N_{c}-2)} \ s_{m}^{\lambda(N_{c}-2)}}{(N_{il}-2)!} \ dG(b_{2}) \ dG(b_{1}) &= \int_{\underline{c}_{\mu}}^{\infty} \int_{\frac{\sigma}{\sigma+1}b_{1}}^{\infty} \frac{1}{\sigma} \left( \frac{\lambda s_{h}^{\lambda}}{b_{2}^{\lambda+1}} \right) \frac{N_{il}! \ b_{2}^{-\lambda(N_{c}-2)} \ s_{m}^{\lambda(N_{c}-2)}}{(N_{il}-2)!} \ db_{2} \ dG(b_{1}) \\ &= \int_{\underline{c}_{\mu}}^{\infty} \frac{1}{\sigma} \left( \frac{\sigma}{\sigma-1} \right)^{-\lambda(N_{c}-1)} \frac{N_{il}! \ b_{1}^{-\lambda(N_{c}-1)} \ s_{m}^{\lambda(N_{c}-2)}}{(N_{il}-2)!} \ \left( \frac{\lambda s_{m}^{\lambda}}{b_{1}^{\lambda+1}} \right) \ db_{1} \\ &= \frac{1}{\sigma} \left( \frac{\sigma}{\sigma-1} \right)^{-\lambda(N_{c}-1)} \frac{c_{\mu}^{-\lambda(N_{c}-1)} \ s_{m}^{\lambda(N_{c}-1)}}{(N_{il}-1)!} \left( \frac{\lambda s_{m}^{\lambda}}{b_{1}^{\lambda+1}} \right) \ db_{1} \end{split}$$

The probability of limit pricing is therefore  $\Pr(\text{LP}) = \{1 - [\sigma/(\sigma-1)]^{-\lambda(N_a-1)}\}$  $\frac{c_{jit}^{-\lambda N_a}}{s_m^{\lambda N_a}}$ , while the probability of not producing (with a Lerner index of zero) is  $1 - \frac{c_{ji}}{s_m^{\lambda N_a}}$ . The aggregate Lerner index under limit pricing is

$$\begin{split} \Pr(\mathrm{LP}) &- \int_{\underline{c}_{w}}^{\infty} b_{1} \int_{b_{1}}^{[\sigma/(\sigma-1)]b_{1}} b_{2}^{-1} \frac{N_{il}! b_{2}^{-\lambda(N_{a}-2)} s_{m}^{\lambda(N_{a}-2)}}{(N_{il}-2)!} dG(b_{2}) dG(b_{1}) \\ &= \Pr(\mathrm{LP}) - \int_{\underline{c}_{w}}^{\infty} b_{1} \int_{b_{1}}^{\frac{\sigma}{\sigma-b_{1}}} \frac{\lambda N_{il}! b_{2}^{-\lambda(N_{a}-1)-2} s_{m}^{\lambda(N_{a}-1)}}{(N_{il}-2)!} db_{2} dG(b_{1}) \\ &= \Pr(\mathrm{LP}) - \int_{\underline{c}_{w}}^{\infty} b_{1} \frac{\lambda N_{il}! b_{1}^{-\lambda(N_{a}-1)-2} s_{m}^{\lambda(N_{a}-1)}}{(\lambda(N_{il}-1)+1)(N_{il}-2)!} \left[ 1 - \left(\frac{\sigma-1}{\sigma}\right)^{\lambda(N_{a}-1)+1} \right] \left(\frac{\lambda b_{1}^{-\lambda-1}}{s_{m}^{-\lambda}}\right) db_{1} \\ &= \Pr(\mathrm{LP}) - \int_{\underline{c}}^{\infty} \frac{\lambda^{2} N_{il}! b_{1}^{-\lambda(N_{a}-1)} s_{m}^{\lambda(N_{w}}}}{(\lambda(N_{il}-1)+1)(N_{il}-2)!} \left[ 1 - \left(\frac{\sigma-1}{\sigma}\right)^{\lambda(N_{a}-1)+1} \right] db_{1} \\ &= \frac{1}{\lambda(N_{il}-1)+1} \left[ 1 - \left(\frac{\sigma-1}{\sigma}\right)^{\lambda(N_{a}-1)+1} \right] \underline{c}_{jil}^{-\lambda N_{w}} s_{m}^{\lambda N_{w}}. \end{split}$$

Adding these indexes conditional on monopolistic limit pricing and not producing gives the expression above.

We define  $\Lambda \equiv \lambda(N-1) + 1 \ge 1 - \lambda > 0$  and suppress the *i* and *t* subscripts for brevity. The percentage change in the expected Lerner index with entry is

$$\frac{d\ln\mathcal{L}}{dN} = \frac{-\frac{1}{\sigma} \left(\frac{\sigma-1}{\sigma}\right)^{\Lambda-1} \lambda \ln \frac{\sigma}{\sigma-1} + \frac{1}{\Lambda} \left(\frac{\sigma-1}{\sigma}\right)^{\Lambda} \lambda \ln \frac{\sigma}{\sigma-1} - \frac{\lambda}{\Lambda^2} \left[1 - \left(\frac{\sigma-1}{\sigma}\right)^{\Lambda}\right]}{\frac{1}{\sigma} \left(\frac{\sigma-1}{\sigma}\right)^{\Lambda-1} + \frac{1}{\Lambda} \left[1 - \left(\frac{\sigma-1}{\sigma}\right)^{\Lambda}\right]}$$

The first three terms correspond to how the expected Lerner index falls,  $d \ln \mathcal{L}/dN < 0$ , as the probability of monopoly pricing decreases (which is the highest possible markup), and while the probability of limit pricing increases, the expected Bertrand markup decreases as well. The numerator of  $d \ln \mathcal{L}/dN$  can be written

$$\lambda \left(\frac{\sigma}{\sigma-1}\right)^{-\Lambda} \left\{ \left(\frac{1}{\Lambda} - \frac{1}{\sigma-1}\right) \ln \frac{\sigma}{\sigma-1} + \left[1 - \left(\frac{\sigma}{\sigma-1}\right)^{\Lambda}\right] \frac{1}{\Lambda^2} \right\}.$$

The first term is weakly negative for  $\Lambda \ge \sigma - 1$ , which holds for  $N \ge 1 + (\sigma - 2)/\lambda$ ; the second term is negative. The upper bound for  $\mathcal{L}$  comes from evaluation at  $N = 1 + (\sigma - 2)/\lambda$ . QED

### 3. Proof of Proposition 3

We define  $\Lambda \equiv \lambda(N-1) + 1$  and suppress the *i* and *t* subscripts. For  $\bar{R} \equiv \sum_j \sum_k \rho \bar{\theta}_k R_{jk} (s_m / \underline{c}_j)^{\lambda N} / N$ , noting  $\pi_{jk} = (1 - \rho) R_{jk}$  and holding capabilities fixed, we have that

$$\begin{split} \frac{d\ln\bar{R}}{dN} &= \frac{\sum_{j}\sum_{k}\bar{\theta}_{k} \cdot d\pi_{jk} \left(s_{m}/\underline{c}_{j}\right)^{\lambda N}/dN}{\sum_{j}\sum_{k}\bar{\theta}_{k}\pi_{jk} \left(s_{m}/\underline{c}_{j}\right)^{\lambda N}} - \frac{1}{N} \\ &= \frac{\sum_{j}\sum_{k}\bar{\theta}_{k} \left(s_{m}/\underline{c}_{j}\right)^{\lambda N} \cdot d\pi_{jk}/dN}{\sum_{j}\sum_{k}\bar{\theta}_{k}\pi_{jk} \left(s_{m}/\underline{c}_{j}\right)^{\lambda N}} \\ &+ \frac{\sum_{j}\sum_{k}\bar{\theta}_{k}\pi_{jk} \cdot d\left(s_{m}/\underline{c}_{j}\right)^{\lambda N}}{\sum_{j}\sum_{k}\bar{\theta}_{k}\pi_{jk} \left(s_{m}/\underline{c}_{j}\right)^{\lambda N}} - \frac{1}{N}. \end{split}$$

Now considering the individual profit terms, holding capabilities fixed, these may increase profits through increased downstream demand in response to lower costs. Note that

$$\frac{d\ln \pi_{jk}}{dN}\bigg|_{\underline{c}_{g} \text{ fixed}} = -\frac{\rho}{1-\rho} \frac{d\ln c_{jkl}}{dN}\bigg|_{\underline{c}_{g} \text{ fixed}}$$
$$= \frac{\rho}{1-\rho} \sum_{i} \frac{\bar{\theta}_{ik}}{\sigma-1} \frac{d\ln(\Omega_{il}\theta_{il}/[\Omega_{il}+(\sigma-1)])}{dN}$$

For  $\Omega_{it} = \lambda N_{it}$  and  $\Phi_{it} \equiv \lambda (N_{it} - 1) + \sigma - 1$ ,  $\vartheta_{it} = 1 - [(\sigma - 1)/\Phi_{it}] \{1 - [(\sigma - 1)/\sigma]^{\Phi_i}\}$ . Consider that after some rearrangement,

$$\frac{d\ln\frac{\Omega_{u}\vartheta_{u}}{\Omega_{u}+(\sigma-1)}}{dN} = \frac{1}{N} + \frac{\frac{\lambda(\sigma-1)}{\Phi}\left(\frac{1}{\Phi+\lambda}+\frac{1}{\Phi}\right)\left[1-\left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right] - \frac{\sigma-1}{\Phi}\left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\lambda\ln\frac{\sigma}{\sigma-1} - \frac{\lambda}{\Phi+\lambda}}{1-\frac{\sigma-1}{\Phi}\left[1-\left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right]},$$

since  $\lambda(N-1) + \sigma - 1 > 0$  is assumed and implies that  $[(\sigma - 1)/\sigma]^{\lambda(N-1)+\sigma-1} < 1$ , which implies

$$\frac{d\ln \pi_{jk}}{dN}\Big|_{\frac{c}{\sigma}, \text{fixed}} = \frac{\rho}{1-\rho} \frac{1}{\sigma-1} \left\{ \frac{1}{N} + \frac{\lambda}{\Phi} \frac{\left(1+\frac{\Phi}{\Phi+\lambda}\right)\frac{(\sigma-1)}{\Phi} \left[1-\left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right] - (\sigma-1)\left(\frac{\sigma-1}{\sigma}\right)^{\Phi} \ln \frac{\sigma}{\sigma-1} - \frac{\Phi}{\Phi+\lambda}}{1-\frac{\sigma-1}{\Phi} \left[1-\left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right]} \right\}.$$

Returning to the sum of profits across firms and industries, we have

$$\begin{split} &\sum_{j}\sum_{k}\bar{\theta}_{k}\frac{d\pi_{jk}}{dN}\left(\frac{s_{m}}{c_{j}}\right)^{\lambda N} \\ &=\sum_{j}\sum_{k}\bar{\theta}_{k}\pi_{jk}\frac{\rho}{1-\rho}\frac{1}{\sigma-1}\left\{\frac{1}{N}+\frac{\lambda}{\Phi}\frac{\left(1+\frac{\Phi}{\Phi+\lambda}\right)\frac{(\sigma-1)}{\Phi}\left[1-\left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right]-(\sigma-1)\left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\ln\frac{\sigma}{\sigma-1}-\frac{\Phi}{\Phi+\lambda}}{1-\frac{\sigma-1}{\Phi}\left(1-\left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right)}\right\}\left(\frac{s_{m}}{c_{j}}\right)^{\lambda N} \\ &=\frac{\rho}{1-\rho}\frac{1}{\sigma-1}\left\{\frac{1}{N}+\frac{\lambda}{\Phi}\frac{\left(1+\frac{\Phi}{\Phi+\lambda}\right)\frac{(\sigma-1)}{\Phi}\left[1-\left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right]-(\sigma-1)\left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\ln\frac{\sigma}{\sigma-1}-\frac{\Phi}{\Phi+\lambda}}{1-\frac{\sigma-1}{\Phi}\left[1-\left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right]}\right\}\sum_{j}\sum_{k}\bar{\theta}_{k}\pi_{jk}\left(\frac{s_{m}}{c_{j}}\right)^{\lambda N}. \end{split}$$

We also have, from  $s_m \leq \underline{c}_j$ , that

$$W(N) = \frac{\sum_{j} \sum_{k} \bar{\theta}_{k} \left[ \pi_{jk} d \left( s_{m} / \underline{c}_{j} \right)^{\lambda N} / dN \right]}{\sum_{j} \sum_{k} \bar{\theta}_{k} \pi_{jk} \left( s_{m} / \underline{c}_{j} \right)^{\lambda N}} \\ = \frac{\sum_{j} \sum_{k} \bar{\theta}_{k} \pi_{jk} \left( s_{m} / \underline{c}_{j} \right)^{\lambda N} \lambda \ln \left( s_{m} / \underline{c}_{j} \right)}{\sum_{j} \sum_{k} \bar{\theta}_{k} \pi_{jk} \left( s_{m} / \underline{c}_{j} \right)^{\lambda N}} \leq 0.$$

So we can conclude, using  $W(N) \cdot N \leq 0$  and the arguments above, that

$$\frac{d\ln\bar{R}}{d\ln N} \leq \frac{\rho}{1-\rho} \frac{1}{\sigma-1} \left\{ 1 + \frac{\lambda N}{\Phi} \frac{\left(1 + \frac{\Phi}{\Phi+\lambda}\right) \frac{(\sigma-1)}{\Phi} \left[1 - \left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right] - (\sigma-1) \left(\frac{\sigma-1}{\sigma}\right)^{\Phi} \ln \frac{\sigma}{\sigma-1} - \frac{\Phi}{\Phi+\lambda}}{1 - \frac{\sigma-1}{\Phi} \left[1 - \left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right]} \right\} - 1.$$

In what follows, the denominator  $1 - [(\sigma - 1)/\Phi]\{1 - [(\sigma - 1)/\sigma]^{\Phi}\}$  is positive so long as  $N \ge 1$ , which is assumed. From the above equation, then, it follows that

$$\lim_{N\to\infty} \frac{d\ln R}{d\ln N} \leq \frac{\rho}{1-\rho} \frac{1-\lambda}{\sigma-1} - 1 < \frac{\rho+1-\sigma+\sigma\rho-\rho}{(1-\rho)(\sigma-1)} = \frac{1-\sigma(1-\rho)}{(1-\rho)(\sigma-1)}.$$

Clearly, then,  $\sigma(1 - \rho) \ge 1$  is sufficient for average weighted revenues to be decreasing and approach zero as  $N \to \infty$  because  $\lambda > 0$ . QED

### 4. Proof of Proposition 4

We define  $\Lambda \equiv \lambda(N-1) + 1$  and suppress the *i* and *t* subscripts for brevity. Taking logs and differentiating  $E[\zeta_i]$  with respect to *N*, holding input capabilities fixed and noting that  $\pi_{jk} = (1 - \rho)R_{jk}$ , we see that

$$\frac{d\ln \mathrm{E}[\zeta_{i}]}{dN} = \frac{d\ln \mathcal{L}}{dN} + \frac{\sum_{j} \sum_{k} \bar{\theta}_{k} \cdot d\pi_{jk} (s_{m}/\underline{c}_{j})^{\lambda N} / dN}{\sum_{j} \sum_{k} \bar{\theta}_{k} \pi_{jk} (s_{m}/\underline{c}_{j})^{\lambda N}} - \frac{1}{N}$$

Also, direct inspection shows that

$$\begin{split} \lim_{N \to \infty} \frac{d \ln \mathcal{L}}{d \ln N} &= \lim_{N \to \infty} \frac{\lambda \left(\frac{\sigma}{\sigma - 1}\right)^{-\Lambda} \left\{ \left(\frac{N}{\Lambda} - \frac{N}{\sigma - 1}\right) \ln \frac{\sigma}{\sigma - 1} + \left[1 - \left(\frac{\sigma}{\sigma - 1}\right)^{\Lambda}\right] \frac{N}{\Lambda^2} \right\}}{\frac{1}{\sigma} \left(\frac{\sigma - 1}{\sigma}\right)^{\Lambda - 1} + \frac{1}{\Lambda} \left[1 - \left(\frac{\sigma - 1}{\sigma}\right)^{\lambda (N - 1) + 1}\right]} \\ &= \lambda \ln \frac{\sigma}{\sigma - 1} \cdot \lim_{N \to \infty} \frac{(N/\Lambda) - [N/(\sigma - 1)]}{1/(\sigma - 1)} = -\infty, \end{split}$$

so from the proof of proposition 3, we have

$$\lim_{N \to \infty} \frac{d \ln \mathbb{E}[\zeta_i]}{dN} < \lim_{N \to \infty} \frac{d \ln \mathcal{L}}{d \ln N} + \frac{\rho}{1 - \rho} \frac{1 - \lambda}{\sigma - 1} - 1 = -\infty,$$

which implies  $\lim_{N\to\infty} \mathbb{E}[\zeta_i] = 0$ . We also have, from the proof of proposition 3, that

$$\begin{split} \frac{d\ln \mathbf{E}[\zeta_{\cdot}]}{dN} &< \frac{d\ln \mathcal{L}}{d\ln N} \\ &+ \frac{\rho}{1-\rho} \frac{1}{\sigma-1} \left\{ 1 + \frac{\lambda N}{\Phi} \frac{\left(1 + \frac{\Phi}{\Phi + \lambda}\right) \frac{(\sigma-1)}{\Phi} \left[1 - \left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right] - (\sigma-1) \left(\frac{\sigma-1}{\sigma}\right)^{\Phi} \ln \frac{\sigma}{\sigma-1} - \frac{\Phi}{\Phi + \lambda}}{1 - \frac{\sigma-1}{\Phi} \left[1 - \left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right]} \right\} - 1. \end{split}$$

Since, from proposition 2,  $d \ln \mathcal{L}/d \ln N < 0$  for  $N > \max\{1 + (\sigma - 2)/\lambda, 1\}$ , a sufficient condition on demand for expected profits to be decreasing in entry holding capabilities fixed is therefore

$$\frac{(1-\rho)(\sigma-1)}{\rho} > 1 + \frac{\lambda N}{\Phi} \frac{\left(1 + \frac{\Phi}{\Phi+\lambda} + \Phi \ln \frac{\sigma}{\sigma-1}\right)^{\frac{(\sigma-1)}{\Phi}} \left[1 - \left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right] - (\sigma-1)\ln \frac{\sigma}{\sigma-1} - \frac{\Phi}{\Phi+\lambda}}{1 - \frac{\sigma-1}{\Phi} \left[1 - \left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right]} \equiv B(N). \quad (D1)$$

Dropping the  $\ln(\sigma/(\sigma - 1))$  terms in equation (D1) that are in sum negative and using  $\lambda N < \Phi$  from  $\lambda < \sigma - 1$  implies

$$B(N) < \frac{1 - \frac{\sigma - 1}{\Phi} \left[ 1 - \left(\frac{\sigma - 1}{\sigma}\right)^{\Phi} \right] + \frac{\sigma - 1}{\Phi} \left[ 1 - \left(\frac{\sigma - 1}{\sigma}\right)^{\Phi} \right] + \frac{\sigma - 1}{\Phi + \lambda} \left[ 1 - \left(\frac{\sigma - 1}{\sigma}\right)^{\Phi} \right] - \frac{\lambda N}{\Phi + \lambda}}{1 - \frac{\sigma - 1}{\Phi} \left[ 1 - \left(\frac{\sigma - 1}{\sigma}\right)^{\Phi} \right]}$$
$$= \frac{\frac{\sigma - 1}{\Phi + \lambda} + \frac{\sigma - 1}{\Phi + \lambda} - \frac{\sigma - 1}{\Phi + \lambda} \left(\frac{\sigma - 1}{\sigma}\right)^{\Phi}}{1 - \frac{\sigma - 1}{\Phi} \left[ 1 - \left(\frac{\sigma - 1}{\sigma}\right)^{\Phi} \right]} = \frac{2 - \left(\frac{\sigma - 1}{\sigma}\right)^{\Phi}}{\frac{\Phi + \lambda}{\sigma - 1}} \left\{ 1 - \frac{\sigma - 1}{\Phi} \left[ 1 - \left(\frac{\sigma - 1}{\sigma}\right)^{\Phi} \right] \right\}}$$
$$= \frac{2 - \left(\frac{\sigma - 1}{\sigma}\right)^{\Phi}}{\frac{\Phi + \lambda}{\sigma - 1} - \left(1 + \frac{\lambda}{\Phi}\right) \left[ 1 - \left(\frac{\sigma - 1}{\sigma}\right)^{\Phi} \right]} < \frac{2}{\frac{\Phi + \lambda}{\sigma - 1} - \left(1 + \frac{\lambda}{\Phi}\right)} = \frac{2}{\frac{\lambda N}{\sigma - 1} - \frac{\lambda}{\lambda(N-1) + \sigma - 1}}.$$

This last equation is decreasing in N where the denominator is positive, which happens for N > 1 as assumed, and the minimum value N can take, by assumption, is  $1 + (\sigma - 2)/\lambda$ . Substituting this into the last term, we have

$$\frac{2}{[(\lambda + \sigma - 2)/(\sigma - 1)] - [\lambda/(\sigma - 2 + \sigma - 1)]} \le \frac{2}{[(\lambda + \sigma - 2)/(\sigma - 1)] - [\lambda/(\sigma - 1)]} = 2\frac{\sigma - 1}{\sigma - 2},$$

where the inequality follows from  $\sigma > 2$ , which follows from the assumption that  $(1 - \rho)\sigma \ge 2$ . Therefore,  $2(\sigma - 1)/(\sigma - 2)$  is an upper bound for B(N) for  $N \ge 1 + (\sigma - 2)/\lambda$ . We can conclude that  $(1 - \rho)(\sigma - 1)/\rho \ge 2[(\sigma - 1)/(\sigma - 2)]$  implies that expected profits are decreasing in entry, which holds for  $(1 - \rho)\sigma \ge 2$ . QED

### 5. Proof of Proposition 5

Firms solve

$$\min_{m_{ijkl}} \int_{\underline{c}_{js}}^{\infty} s_{iil} m_{iijkl} \ dG_{il}(\iota), \text{ subject to } \left( \int_{\underline{c}_{js}}^{\infty} m_{iijkl}^{(\sigma-1)/\sigma} \ dG_{il}(\iota) \right)^{\sigma/(\sigma-1)} \ge M_{ijkl}.$$

A natural question is, Why not frame this as a free-end-point problem with a choice of input varieties  $[\underline{c}_{ijt}, \overline{c}_{iji}]$ ? The reason we have not is that for the case  $\sigma > 1$ , "love for variety" implies  $\overline{c}_{ijt} = \infty$ , and for  $\sigma < 1$ , the production function exhibits "hate for variety," and allowing the producer to choose a subset of suppliers will cause them to snap to the lowest-cost supplier.

Cost minimization conditional on  $\underline{c}_{ijt}$  implies a first-order condition of <sup>37</sup>

$$\begin{split} m_{iijkl}^{(\sigma-1)/\sigma} &= M_{ijkl}^{(\sigma-1)/\sigma} \left( \frac{\sigma}{\sigma-1} \frac{s_{iil}}{\eta} \right)^{1-\sigma}, \quad \text{where} \\ \eta_{il} &= \left[ -\int_{\infty}^{c_{jj}} \left( \frac{\sigma}{\sigma-1} s \right)^{1-\sigma} dG_{il}(s) \right]^{1/(1-\sigma)}. \end{split}$$

<sup>37</sup> This is for  $\sigma > 1$ . For  $\sigma < 1$ , replace  $\sigma/(\sigma - 1)$  with  $\sigma/(1 - \sigma)$ , as the sign of the inequality constraint changes. The second-order condition holds for  $\sigma > 0$  (weakly at  $\sigma = 1$ ).

Under these distributional assumptions, we have

$$\eta_{ii} = \frac{\sigma}{\sigma - 1} \left[ \frac{\Omega_{ii}}{\Omega_{ii} + (\sigma - 1)} s_m^{\Omega_a} \underline{c}_{ji}^{1 - \sigma - \Omega_a} \right]^{1/(1 - \sigma)}$$

under the condition  $\Omega_{ii} > 1 - \sigma$ ,  $\eta_{ii}$  is finite, and the input choice is nondegenerate.<sup>38</sup> Defining the cost index of input *i* as  $S_{iji}$ , we have minimum costs of  $S_{iji}M_{ijko}$ , where

$$S_{ijt} = \left[\frac{\Omega_{it} \vartheta_{it}}{\Omega_{it} + (\sigma - 1)}\right]^{1/(1-\sigma)} \underline{c}_{ijt}^{1-\Omega_s/(1-\sigma)} s_m^{\Omega_s/(1-\sigma)},$$

and therefore

$$\frac{d\ln S_{ijt}}{d\ln \underline{c}_{ijt}} = 1 + \frac{\Omega_{it}}{\sigma - 1}.$$

Now the restriction  $\Omega_{it} > 1 - \sigma$  is especially informative, as if  $\sigma > 1$ , then  $d \ln S_{ijt}/d \ln \underline{c}_{ijt} > 0$ , consistent with love for variety, and  $d \ln S_{ijt}/d \ln \underline{c}_{ijt} < 0$  for  $\sigma < 1$ , consistent with hate for variety. Unit input costs  $c_{jkt}$  conditional on capabilities are then as above. QED

## 6. Proof of Proposition 6

Profit maximization can be considered in two steps, maximizing industry profits conditional on unit costs and then maximizing joint profits by choosing capabilities. A firm will optimally choose a markup  $p_{jkl} = c_{jkl}/\rho$  in the first maximization step, so the profit accruing from each industry is

$$\pi_{jkt} = [(1/\rho) - 1] \omega(\underline{c}_{jt}) c_{jkt} q_{jkt} = \frac{[(1/\rho) - 1] (\rho D_{kt})^{1/(1-\rho)}}{(\omega(\underline{c}_{jt}) c_{jkt})^{\rho/(1-\rho)}}.$$
 (D2)

Noting that for this particular profit form and common markups across industries, we have

$$\begin{split} \frac{d\ln\pi_{jkt}}{d\ln\underline{c}_{ijt}} &= -\frac{\rho}{1-\rho} \left( \frac{d\ln\omega(\underline{c}_{jt})}{d\ln\underline{c}_{ijt}} + \frac{d\ln c_{jkt}}{d\ln\underline{c}_{ijt}} \right) \\ &= -\frac{\rho}{1-\rho} \left[ \ln\underline{c}_{ijt} - \ln\underline{c}_{i0} + \bar{\theta}_{ik} \left( 1 - \frac{\Omega_{it}}{1-\sigma} \right) \right], \end{split}$$

it follows that the first-order condition for profit maximization

$$\frac{d\pi_{jt}}{d\underline{c}_{ijt}} = \sum_{k} \frac{\pi_{jkt}}{\underline{c}_{ijt}} \frac{d\ln \pi_{jkt}}{d\ln \underline{c}_{ijt}} 
= -\frac{\rho}{1-\rho} \sum_{k} \frac{\pi_{jkt}}{\underline{c}_{ijt}} \left[ \ln \underline{c}_{ijt} - \ln \underline{c}_{i0} + \overline{\theta}_{ik} \left( 1 - \frac{\Omega_{it}}{1-\sigma} \right) \right] = 0.$$
(D3)

<sup>38</sup> Otherwise, for  $\sigma < 1$  it is optimal to use all of the cheapest input, and for  $\sigma > 1$ , input vectors of the type  $\kappa s^{1-\sigma}$  all satisfy the production constraint, so as  $\kappa \to 0$ , costs go to zero.

Using the fact that  $\rho \pi_{jkl}/(1-\rho) = \omega(\underline{c}_{jl}) c_{jkl} q_{jkl}$ , equation (D3) implies that for firminput expenditure shares of  $\theta_{ijk}$  the optimal capability choice satisfies

$$\ln \underline{c}_{ijt} = \ln \underline{c}_{i0} - \left(1 + \frac{\Omega_{it}}{\sigma - 1}\right) \theta_{ijt}.$$

Substitution into equation (D2) and further expansion shows that revenues  $R_{jkt}$  take the above form. QED

### D2. Extensive Product Margin

Equation (9) can be modified to consider the extensive product margin choice of firms. Assume that firms face a fixed  $\cot(1 - \rho)f_{kt}$  to produce in an industry *k* each period and so produce when profits  $\pi_{jkt} = (1 - \rho)R_{jkt} > (1 - \rho)f_{kt}$ . From equation (8), with identical coefficients and fixed effects similar to equation (9) and error terms with  $-\varepsilon_{jkt}$  logistic, a firm *j* operates in industry *k* when the following equation is positive:

$$\ln \frac{R_{jkt}}{f_{kt}} = \kappa_{kt} + \kappa_{jk} - \kappa_0 \sum_i \left(\theta_{ijt} - \bar{\theta}_{ik}\right)^2 + \kappa_1 \sum_i \left(\alpha_B B_{ijt} + \alpha_\tau \Delta \tau_{ijt}\right) \left(\theta_{ijt} - \bar{\theta}_{ik}\right)^2 + \varepsilon_{jkt}, \quad (D4)$$

Equation (D4) can be estimated to recover the tariff equivalent of dereservation on the extensive margin of industry adoption.

#### D3. Input Similarity Equation: Proof of Proposition 7

Let  $\{D_{kl}\}$  be demand shifters in period *t*. Let  $C_{jk} = c_{jk}q_{jk}$  be the variable costs for firm *j* in producing in industry *k* and  $C_j = \sum_k C_{jk}$  total variable costs, so that

$$\theta_{ijt} = \frac{\sum_{k} \bar{\theta}_{ik} C_{jk}}{C_{j}} = \frac{\sum_{k} \bar{\theta}_{ik} D_{kt}^{1/(1-\rho)} c_{jkt}^{-\rho/(1-\rho)}}{\sum_{k} D_{kt}^{1/(1-\rho)} c_{jk}^{-\rho/(1-\rho)}}.$$
 (D5)

Holding  $\underline{c}_{jit}$  fixed, for  $\chi_{jkt} \equiv C_{jk}/C_j$ , the cost share of industry *k* for firm *j* (equal to revenue shares), it is the case that

$$\frac{d\theta_{ijt}}{dD_{kt}} = \frac{1}{C_j^2} \left( \frac{\bar{\theta}_{ik}}{1-\rho} \frac{C_{jk}}{D_{kt}} C_j - \frac{1}{1-\rho} \frac{C_{jk}}{D_{kt}} \sum_k \bar{\theta}_{ik} C_{jk} \right) = \frac{\chi_{jkt}}{1-\rho} \frac{\bar{\theta}_{ik} - \theta_{ijt}}{D_{kt}}$$

It follows from the mean-value theorem that for some  $\{\delta_{jk}\}$  with each  $\delta_{jk} \in [D_{kt-1}, D_{kt}]$ , cost shares  $\chi_{ik}^*$ , and expenditure shares  $\theta_{ij}^*$  evaluated at  $\{\delta_{jk}\}$ ,

$$\sum_{i} \left( \bar{\theta}_{ik} \theta_{ijt} - \frac{\theta_{ijt}^2}{2} \right) - \left( \bar{\theta}_{ik} \theta_{ijt-1} - \frac{\theta_{ijt-1}^2}{2} \right) = \sum_{i} \left( \bar{\theta}_{ik} - \theta_{ijt}^* \right) \frac{\chi_{jk}^*}{1 - \rho} \left( \bar{\theta}_{ik} - \theta_{ijt}^* \right) \frac{D_{kt} - D_{kt-1}}{\delta_{jk}}.$$

Redefining  $\delta_{jk} = D_{kl-1}$  as common across firms yields the (feasible) approximation

$$\sum_{i} \left( \bar{\theta}_{ik} \theta_{ijt} - \frac{\theta_{ijt}^2}{2} \right) \approx \sum_{i} \left( \bar{\theta}_{ik} \theta_{ijt-1} - \frac{\theta_{ijt-1}^2}{2} \right) + \sum_{i} \left( \bar{\theta}_{ik} - \theta_{ijt-1} \right)^2 \frac{\chi_{jkt-1}}{1 - \rho} \frac{D_{kt} - D_{kt-1}}{D_{kt-1}}$$

Adding  $\bar{\theta}_{ik}^2$  and rearranging gives the result. QED

#### D4. Extensions to Primary Factors

The model can be extended on the production side to include factor services that are not directly consumed, in particular factors such as capital, different types of labor, and other firm balance sheet items. In this extension, to produce a quantity  $q_{jkl}$  in industry k at time t, firm j combines inputs from industry i,  $M_{ijkl}$ , and factor services from factor f,  $F_{jjkl}$ , using a constant-return-to-scale Cobb-Douglas technology with industry input expenditure shares  $\bar{\theta}_{ik}$ ,  $\bar{\theta}_{fk}$  and idiosyncratic industry productivity labeled  $\varphi_{jk}$ . At input prices  $S_{ijl}\psi_{il}$  and factor service prices  $W_{jjk}\psi_{jl}$ , the unit cost of firm j to produce in industry k at time t is therefore

$$c_{jkt} \equiv \prod_{i} \left( \frac{S_{ijt} \psi_{it}}{\overline{\theta}_{ik} \varphi_{jk}} \right)^{\theta_{s}} \cdot \prod_{f} \left( \frac{W_{fit} \psi_{ft}}{\overline{\theta}_{fk} \varphi_{jk}} \right)^{\theta_{p}}$$

Thus,  $c_{jkl}$  is a vector of unit costs that are influenced by input prices and industry productivities. Industry-level inputs  $M_{ijk}$  and F are again composite quantities of varieties through a CES aggregator with elasticity of substitution  $\sigma$  for inputs and  $\sigma_f$  for factor services where varieties follow a Pareto distribution, with  $\Pr(s_{iil} \geq s) = (s/s_m)^{-\Omega_s}$  for inputs and  $\Pr(w_{ijl} \geq w) = (w/w_m)^{-\Omega_s}$  for factors with  $\Omega_{jl} = \lambda N_{jl}$ , and  $N_{jl}$  is the mass of suppliers as modeled above but for factor services. Firms have capabilities of using inputs with prices  $[\underline{c}_{ijl}, \infty)$  for inputs and  $[\underline{c}_{fil}, \infty)$  for factors, where  $\underline{c}_{ijl}$  and  $\underline{c}_{fil}$  are chosen by the firm. The analogous version of proposition 1 goes through with corresponding and symmetric terms for factors. Letting  $\underline{c}_{jl}$  denote the vector of acquired capabilities, the unit costs of a multiproduct firm are given by  $\omega(\underline{c}_{jl})c_{jkt}$  in each industry and are assumed to follow

$$\omega(\underline{c}_{ji}) \equiv \exp\left\{\frac{\sum_{i} (\ln \underline{c}_{iji})^2}{2} + \frac{\sum_{i} (\ln \underline{c}_{iji})^2}{2}\right\}.$$

A firm can use its acquired capabilities across any number of products and reoptimizes by choosing capabilities each period.

In period t, firms pay a fixed cost of  $f_{kt}$  to operate in industry k and face inverse demand in industry k of

$$p_{jkt}(q_{jkt}) = D_{kt}q_{jkt}^{\rho-1},$$

as above. A firm's profit-maximizing capability and production choices considering product markets jointly are analogous to proposition 6. In particular, for  $\Theta_{ii} \equiv 1 + \Omega_{ii}/(\sigma - 1)$  and  $\Theta_{\beta} \equiv 1 + \Omega_{\beta}/(\sigma_f - 1)$ , firm revenues are given by

$$\ln R_{jkt} = \underbrace{\frac{\rho}{1-\rho} \ln \varphi_{jk}}_{\text{RCA}(jk)} + \underbrace{\ln\left(\rho^{\rho/(1-\rho)} D_{kt}^{1/(1-\rho)}\right)}_{\text{demand }(kt)}$$

$$- \underbrace{\frac{\rho}{1-\rho} \sum_{i} \overline{\theta}_{ik} \ln \psi_{il} \vartheta_{jl}^{1/(1-\sigma)} \left(1-\Theta_{il}^{-1}\right)^{1/(1-\sigma)} \frac{s_{lm}^{1-\Theta_{s}}}{\overline{\theta}_{ik}}}_{\text{input supplier }(kt)} - \underbrace{\frac{\rho}{1-\rho} \sum_{j} \overline{\theta}_{jk} \ln \psi_{jl} \vartheta_{jl}^{1/(1-\sigma)} \left(1-\Theta_{jl}^{-1}\right)^{1/(1-\sigma_{j})} \frac{w_{lm}^{1-\Theta_{s}}}{\overline{\theta}_{jk}}}_{\text{factor supplier }(kt)}$$

$$+ \underbrace{\frac{\rho}{2(1-\rho)} \left[ \sum_{i} \Theta_{il}^{2} \overline{\theta}_{ik}^{2} + \sum_{j} \Theta_{jl}^{2} \overline{\theta}_{jk}^{2} \right]}_{\text{supplier-tech }(kt)} - \underbrace{\frac{\rho}{2(1-\rho)} \left[ \sum_{i} \Theta_{il}^{2} \left(\theta_{ijt} - \overline{\theta}_{ik}\right)^{2} + \sum_{j} \Theta_{jl}^{2} \left(\theta_{jjt} - \overline{\theta}_{jk}\right)^{2} \right]}_{\text{core competency }(jk)},$$

where  $\vartheta_{j_l} \equiv 1 + \{(\sigma - 1)/[\lambda(N_{j_l} - 1) + \sigma - 1]\}\{1 - [(\sigma - 1)/\sigma]^{\lambda(N_p - 1) + \sigma - 1}\}$ . Linearizing  $\Omega_{i_l}$  and  $\Omega_{j_l}$  around the initial policy states  $\Omega_{i_0} = \overline{\Omega}$  and  $\Omega_{f_0} = \overline{\Omega}_f$  and letting  $\kappa_x$  represent a fixed effect for characteristic *x* yields equation (15). The IV estimator is analogous to the one above; for  $\tilde{\theta}_{j_lkl} \equiv (\theta_{j_ll}\bar{\theta}_{j_k} - \theta_{j_ll}^2/2)$  and  $\tilde{\chi}_{j_lkl} \equiv \chi_{j_{kl}}(\theta_{j_ll}\bar{\theta}_{j_k} - \theta_{j_{ll}}^2/2)$ , we define

$$\begin{split} I^{M}_{jkt}(\lambda,\gamma) &\equiv \lambda \sum_{i} \tilde{\theta}_{ijkt-1} - \gamma \sum_{i} \tilde{\chi}_{ijkt-1}, \qquad I^{F}_{jkt}(\lambda,\gamma) &\equiv \lambda \sum_{f} \tilde{\theta}_{jjkt-1} - \gamma \sum_{f} \tilde{\chi}_{ijkt-1}, \\ I^{B}_{jkt}(\lambda,\gamma) &\equiv \lambda \sum_{i} B_{ijt} \tilde{\theta}_{ijkt-1} - \gamma \sum_{i} B_{ijt} \tilde{\chi}_{ijkt-1}, \quad I^{\tau}_{jkt}(\lambda,\gamma) &\equiv \lambda \sum_{i} \tau_{ijt} \tilde{\theta}_{ijkt-1} - \gamma \sum_{i} \tau_{ijt} \tilde{\chi}_{ijkt-1}. \end{split}$$

The resulting first-stage equations for our estimator are as follows for  $\zeta^{ij} \equiv (\lambda^{ij}, \gamma^{ij})$ :<sup>39</sup>

$$\sum_{i} \tilde{\theta}_{ijkt} = \kappa_{kt} + \kappa_{jk} + I_{jkt}^{M} \left( \zeta^{11} \right) + I_{jkt}^{F} \left( \zeta^{12} \right) + I_{jkt}^{B} \left( \zeta^{13} \right) + I_{jkt}^{\tau} \left( \zeta^{14} \right) + \eta_{jkt}^{M}, \quad (D6)$$

$$\sum_{f} \tilde{\theta}_{jjkt} = \kappa_{kt} + \kappa_{jk} + I_{jkt}^{M}(\zeta^{21}) + I_{jkt}^{F}(\zeta^{22}) + I_{jkt}^{B}(\zeta^{23}) + I_{jkt}^{\tau}(\zeta^{24}) + \eta_{jkt}^{L}, \quad (D7)$$

$$\sum_{i} B_{ijt} \tilde{\theta}_{ijkt} = \kappa_{kt} + \kappa_{jk} + I_{jkt}^{M} \left( \zeta^{31} \right) + I_{jkt}^{F} \left( \zeta^{32} \right) + I_{jkt}^{B} \left( \zeta^{33} \right) + I_{jkt}^{\tau} \left( \zeta^{34} \right) + \eta_{jkt}^{B}, \tag{D8}$$

$$\sum_{i} \tau_{ijt} \tilde{\theta}_{ijkt} = \kappa_{kt} + \kappa_{jk} + I^{M}_{jkt} (\boldsymbol{\zeta}^{41}) + I^{F}_{jkt} (\boldsymbol{\zeta}^{42}) + I^{B}_{jkt} (\boldsymbol{\zeta}^{43}) + I^{\tau}_{jkt} (\boldsymbol{\zeta}^{44}) + \eta^{\tau}_{jkt} \,. \tag{D9}$$

With our base and extended models in hand, along with an IV strategy, we next turn to our estimate results and counterfactuals regarding the comparative advantage of firms.

## Appendix E

#### Average Firm-Level Comparative Advantage, by Industry

Table E1 shows the average comparative advantage of single-industry firms in industry k', for the industry in which they enjoy the highest average CA<sub>*jkt*</sub>.

<sup>&</sup>lt;sup>39</sup> In practice, sales within a firm-industry group are unlikely to be a balanced panel, as the extensive margin of a firm's industries is liable to change (we in fact model and estimate this with a logit model). Consequently, our one-period-lag strategy may lose some observations, but it reduces the number of parameters that must be estimated simultaneously.

	Highest–Average	
Industry k	Comparative Advantage	Comparative
	mausuy (except k)	Auvantage
Dairy products	Live animals, chiefly for food	15.8*
Other jute and natural fiber	Fabrics and cloth of jute, coir, sisal,	13.1*
goods, n.e.c.	hemp, mista, etc.	
Fabrics and cloth of jute, coir,	Other jute and natural fiber goods,	12.3*
sisal, hemp, mista, etc.	n.e.c.	11 5+
Fibre of jute, coir, and other plants	Fabrics and cloth of jute, coir, sisal, hemp, mista, etc.	$11.7^{+}$
Cereals (incl. rice) and pulses, unmilled	Products of milling industries; malt and malted milk	11.6*
Products of milling industries; malt	Cereals (incl. rice) and pulses,	11.5*
Ginned cotton, cotton, and raw	Cotton yarn and fibre, including	10.2*
Cotton yarn and fibre, including	Ginned cotton, cotton, and raw	$10.0^{+}$
Vegetables oils and fats	Diesel products and by-products	9.8
Raw fibre of jute, coir, sisal, hemp,	Fabrics and cloth of jute, coir, sisal,	9.6
mista, etc.	hemp, mista, etc.	
Aluminium and aluminium	Aluminium and aluminium alloys	9.5*
alloys, unwrought	worked	
Leather apparel	Leather bags, cases, purse and other novelty items	9.2*
Fruit juices and vegetable juices	Edible fruit and nuts; edible	9.2
and syrup, pickles	vegetables and certain roots	
Craft paper and paper	Boards, paper boards	9.1*
for special use		
Leather bags, cases, purses and other novelty items	Leather apparel	9.0*
Boards, paper boards	Craft paper and paper for special use.	8.7
Chocolate, cocoa and cocoa	Sugar, molasses, khandsari, gur.	8.6
preparations and sugar	0, , ,0	
Edible fruits and nuts; edible	Fruit juices and vegetable juices &	8.5*
vegetables and certain roots	syrup, Pickles	
Aluminium and aluminium alloys worked	Aluminium and aluminium alloys, unwrought	8.2
Paper (uncoated) used for newsprint and for other	Craft paper and paper for special use	8.0
Big iron /formallow ata in	Matro railways and tramways and	7.0*
ng non/ terroanoys, etc., in	rolling stock	1.9"
Cotton apparel	Fur skins and articles thereof	77
Inorganic elements, excluding	Charcoal	74
hase metals rare gas	Charcoal	1.1
Miscellaneous leather manufac-	Leather bags cases purses and	73
tured items	other novelty items	
Copper and copper alloy, refined or not, unwrought	Copper and copper alloys, worked	7.0*

TABLE E1 Comparative Advantage of Single-Industry Firms, by Industry

NOTE.—Table shows the average comparative advantage  $\overline{CA}_{jkl}$  of single-industry plants in industry k', for the industry k where  $\overline{CA}_{jkl}$  is the highest. n.e.c. = not elsewhere classified. <sup>+</sup> p < .10. <sup>\*</sup> p < .05.

# 3096

## Appendix F

## **Core-Competency Premia with Factor Complementarity**

Industry Rank		NUMBER OF INDUSTRIES WITH POSITIVE SALES									
	1	2	3	4	5	6	7	8	9	10 +	
1	.024	.030	.030	.017	.013	.011	.010	.010	.007	.011	
2		.014	.012	.009	.009	.008	.008	.008	.006	.014	
3			.010	.008	.007	.007	.006	.006	.008	.009	
4				.007	.007	.006	.006	.005	.006	.009	
5					.006	.006	.006	.005	.006	.005	
6						.006	.006	.006	.006	.004	
7							.005	.005	.005	.004	
8								.005	.005	.005	
9									.005	.005	
10 +										.003	

 TABLE F1

 Core-Competency Sales Premium, with Primary Factors

 TABLE F2

 Core-Competency Sales Premium, with Primary Factors—Weighted

	Number of Industries with Positive Sales									
INDUSTRY RANK	1	2	3	4	5	6	7	8	9	10 +
1	.029	.038	.070	.085	.076	.077	.097	.168	.299	1.102
2		.003	.007	.022	.098	.198	.166	.205	.014	2.214
3			.001	.003	.004	.031	.012	.023	.144	.858
4				.001	.004	.025	.010	.014	.014	.116
5					.002	.005	.008	.006	.011	.029
6						.002	.004	.006	.004	.007
7							.001	.004	.003	.011
8								.001	.001	.004
9									.003	.003
10+										.001

### References

- Acemoglu, Daron, Pol Antràs, and Elhanan Helpman. 2007. "Contracts and Technology Adoption." A.E.R. 97 (3): 916–43.
- Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. 2012. "The Network Origins of Aggregate Fluctuations." *Econometrica* 80 (5): 1977–2016.
- Amihud, Yakov, and Baruch Lev. 1981. "Risk Reduction as a Managerial Motive for Conglomerate Mergers." *Bell J. Econ.* 12 (2): 605–17.
- Amirapu, Amrit, Michael Gechter, and Gabriel Smagghue. 2018. "Dynamic Effects of Product Market Regulation: Evidence from India." Working paper.
- Amiti, Mary, and Jozef Konings. 2007. "Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia." A.E.R. 97 (5): 1611–38.
- Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics:* An Empiricist's Companion. Princeton, NJ: Princeton Univ. Press.
- Antràs, Pol, and Davin Chor. 2013. "Organizing the Global Value Chain." *Econometrica* 81 (6): 2127–204.
- Atalay, Enghin, Ali Hortaçsu, Mary Jialin Li, and Chad Syverson. 2019. "How Wide Is the Firm Border?" *Q.J.E.* 134 (4): 1845–82.
- Aw, Bee Yan, and Yi Lee. 2009. "Product Choice and Market Competition: The Case of Multiproduct Electronic Plants in Taiwan." *Scandinavian J. Econ.* 111 (4): 711–40.
- Baumol, William J. 1977. "On the Proper Cost Tests for Natural Monopoly in a Multiproduct Industry." A.E.R. 67 (5): 809–22.
- Bear, D. V. T. 1965. "Inferior Inputs and the Theory of the Firm." *J.P.E.* 73 (3): 287–89.
- Bernard, Andrew B., Emily J. Blanchard, Ilke Van Beveren, and Hylke Vandenbussche. 2019. "Carry-Along Trade." *Rev. Econ. Studies* 86 (2): 526–38.
- Bernard, Andrew B., Emmanuel Dhyne, Glenn Magerman, Kalina Manova, and Andreas Moxnes. 2022. "The Origins of Firm Heterogeneity: A Production Network Approach." *I.P.E.* 130 (7): 1765–804.
- Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum. 2003. "Plants and Productivity in International Trade." A.E.R. 93 (4): 1268–90.
- Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott. 2007. "Firms in International Trade." *J. Econ. Perspectives* 21 (3): 105–30.
- 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.
- Bernheim, B. Douglas, and Michael D. Whinston. 1990. "Multimarket Contact and Collusive Behavior." *RAND J. Econ.* 21 (1): 1–26.
- Bloom, Nicholas, Aprajit Mahajan, David McKenzie, and John Roberts. 2010. "Why Do Firms in Developing Countries Have Low Productivity?" A.E.R. 100 (2): 619–23.
- Bloom, Nicholas, Mark Schankerman, and John Van Reenen. 2013. "Identifying Technology Spillovers and Product Market Rivalry." *Econometrica* 81 (4):1347– 93.
- Boehm, Johannes, Swati Dhingra, and John Morrow. 2016. "Swimming Upstream: Input-Output Linkages and the Direction of Product Adoption." Discussion Paper no. 1407, Centre for Economic Performance, London.
- ———. 2018. "Product Diversification in Indian Manufacturing." In *Developments in Global Sourcing*, edited by Wilhelm Kohler and Erdal Yalcin, 337–50. Cambridge, MA: MIT Press (for CESifo).
- Bollard, Albert, Peter J. Klenow, and Gunjan Sharma. 2013. "India's Mysterious Manufacturing Miracle." *Rev. Econ. Dynamics* 16 (1): 59–85.
- Bowen, Harry P., and Margarethe F. Wiersema. 2005. "Foreign-Based Competition and Corporate Diversification Strategy." *Strategic Management J.* 26 (12): 1153–71.
- Bown, Chad P., and Meredith A. Crowley. 2016. "The Empirical Landscape of Trade Policy." In *Handbook of Commercial Policy*, vol. 1A, edited by Kyle Bagwell and Robert W. Staiger, 3–108. Amsterdam: North-Holland.
- Brander, James A., and Jonathan Eaton. 1984. "Product Line Rivalry." A.E.R. 74 (3): 323–34.
- Broda, Christian, and David E. Weinstein. 2006. "Globalization and the Gains from Variety." *Q.J.E.* 121 (2): 541–85.
- Bryce, David J., and Sidney G. Winter. 2009. "A General Interindustry Relatedness Index." *Management Sci.* 55 (9): 1570–85.

3098

- Cameron, A. Colin, and Douglas L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." J. Human Resources 50 (2): 317–72.
- Chandler, Alfred D. 1992. "Organizational Capabilities and the Economic History of the Industrial Enterprise." *J. Econ. Perspectives* 6 (3): 79–100.
- Combes, Pierre-Philippe, and Laurent Gobillon. 2015. "The Empirics of Agglomeration Economies." In *Handbook of Regional and Urban Economics*, vol. 5, edited by Gilles Duranton, J. Vernon Henderson, and William C. Strange, 247–348. Amsterdam: North-Holland.
- Conley, Timothy G., and Bill Dupor. 2003. "A Spatial Analysis of Sectoral Complementarity." *J.P.E.* 111 (2): 311–52.
- Crozet, Matthieu, and Federico Trionfetti. 2013. "Firm-Level Comparative Advantage." J. Internat. Econ. 91 (2): 321–28.
- de Blas, Beatriz, and Katheryn N. Russ. 2015. "Understanding Markups in the Open Economy." *American Econ. J. Macroeconomics* 7 (2): 157–80.
- Dhingra, Swati. 2013. "Trading Away Wide Brands for Cheap Brands." A.E.R. 103 (6): 2554–84.
- Dhingra, Swati, and John Morrow. 2019. "Monopolistic Competition and Optimum Product Diversity under Firm Heterogeneity." J.P.E. 127 (1): 196–232.
- di Giovanni, Julian, Andrei A. Levchenko, and Isabelle Méjean. 2014. "Firms, Destinations, and Aggregate Fluctuations." *Econometrica* 82 (4): 1303–40.
- Ding, Xiang. 2019. "Intangible Economies of Scope: Micro Evidence and Macro Implications." Working paper.
- Durlauf, Steven N. 1993. "Nonergodic Economic Growth." *Rev. Econ. Studies* 60 (2): 349–66.
- Eaton, B. Curtis, and Nicolas Schmitt. 1994. "Flexible Manufacturing and Market Structure." A.E.R. 84 (4): 875–88.
- Eckel, Carsten, Leonardo Iacovone, Beata Javorcik, and J. Peter Neary. 2015. "Multi-product Firms at Home and Away: Cost- versus Quality-Based Competence." J. Internat. Econ. 95 (2): 216–32.
- Eckel, Carsten, and J. Peter Neary. 2010. "Multi-product Firms and Flexible Manufacturing in the Global Economy." *Rev. Econ. Studies* 77 (1): 188–217.
- Ethier, Wilfred J. 1982. "National and International Returns to Scale in the Modern Theory of International Trade." *A.E.R.* 72 (3): 389–405.
- Fan, Joseph P. H., and Larry H. P. Lang. 2000. "The Measurement of Relatedness: An Application to Corporate Diversification." *J. Bus.* 73 (4): 629–60.
- Flagge, Matthew, and Ritam Chaurey. 2014. "Firm-Product Linkages and the Evolution of Product Scope." Working paper.
- Fontagné, Lionel, Angelo Secchi, and Chiara Tomasi. 2018. "Exporters' Product Vectors across Markets." European Econ. Rev. 110:150–80.
- 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.
- Galle, Simon. 2015. "Competition, Financial Constraints and Misallocation: Plant-Level Evidence from Indian Manufacturing." Working paper, Univ. Calfornia, Berkeley.
- García-Santana, Manuel, and Josep Pijoan-Mas. 2014. "The Reservation Laws in India and the Misallocation of Production Factors." J. Monetary Econ. 66:193–209.
- 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.

## COMPARATIVE ADVANTAGE OF FIRMS

- Government of India. 2009. "Micro, Small, and Medium Enterprises in India: An Overview." Report, Ministry of Micro, Small, and Medium Enterprises.
- Halpern, László, Miklós Koren, and Adam Szeidl. 2015. "Imported Inputs and Productivity." A.E.R. 105 (12): 3660–703.
- Harrison, Ann, and Andrés Rodríguez-Clare. 2010. "Trade, Foreign Investment, and Industrial Policy for Developing Countries." In *Handbook of Development Economics*, vol. 5, edited by Dani Rodrik and Mark Rosenzweig, 4039–214. Amsterdam: North-Holland.
- Hausmann, Ricardo, and César A. Hidalgo. 2011. "The Network Structure of Economic Output." *J. Econ. Growth* 16 (4): 309–42.
- Hausmann, Ricardo, Jason Hwang, and Dani Rodrik. 2007. "What You Export Matters." J. Econ. Growth 12 (1): 1–25.
- Hidalgo, César A., Bailey Klinger, Albert-László Barabási, and Ricardo Hausmann. 2007. "The Product Space Conditions the Development of Nations." *Science* 317 (5837): 482–87.
- Hoberg, Gerard, and Gordon Phillips. 2016. "Text-Based Network Industries and Endogenous Product Differentiation." *J.P.E.* 124 (5): 1423–65.
- Hottman, Colin J., Stephen J. Redding, and David E. Weinstein. 2016. "Quantifying the Sources of Firm Heterogeneity." *Q.J.E.* 131 (3): 1291–364.
- Hussain, Abid. 1997. *Report of the Expert Committee on Small Enterprises*. New Delhi: Nat. Council Appl. Econ. Res., Ministry of Industry, Government of India.
- Hutchinson, John, Jozef Konings, and Patrick Paul Walsh. 2010. "The Firm Size Distribution and Inter-industry Diversification." *Rev. Indus. Org.* 37 (2): 65–82.
- Iacovone, Leonardo, and Beata S. Javorcik. 2010. "Multi-product Exporters: Product Churning, Uncertainty and Export Discoveries." *Econ. J.* 120 (544): 481–99.
- Jensen, Michael C. 1986. "Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers." A.E.R. 76 (2): 323–29.
- Jovanovic, Boyan. 1987. "Micro Shocks and Aggregate Risk." Q.J.E. 102 (2): 395-410.
- Jovanovic, Boyan, and Richard J. Gilbert. 1993. "The Diversification of Production." Brookings Papers Econ. Activity Microeconomics 1993 (1): 197–247.
- Kasahara, Hiroyuki, and Joel Rodrigue. 2008. "Does the Use of Imported Intermediates Increase Productivity? Plant-Level Evidence." J. Development Econ. 87 (1): 106–18.
- Koren, Miklós, and Silvana Tenreyro. 2013. "Technological Diversification." *A.E.R.* 103 (1): 378–414.
- Kugler, Maurice, and Eric Verhoogen. 2009. "Plants and Imported Inputs: New Facts and an Interpretation." *A.E.R.* 99 (2): 501–7.
- Lichtenberg, Frank R. 1992. "Industrial De-diversification and Its Consequences for Productivity." J. Econ. Behavior and Org. 18 (3): 427–38.
- Liu, Runjuan. 2010. "Import Competition and Firm Refocusing." Canadian J. Econ./Revue canadienne d'économique 43 (2): 440–66.
- López, Ricardo A., and Jens Südekum. 2009. "Vertical Industry Relations, Spillovers, and Productivity: Evidence from Chilean Plants." J. Regional Sci. 49 (4): 721–47.
- Lu, Dan, Asier Mariscal, and Luis-Fernando Mejía. 2016. "How Firms Accumulate Inputs: Evidence from Import Switching." Working paper.
- Maksimovic, Vojislav, and Gordon M. Phillips. 2013. "Conglomerate Firms, Internal Capital Markets, and the Theory of the Firm." *Ann. Rev. Financial Econ.* 5:225–44.

3100

- 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.

- Montgomery, Cynthia A. 1994. "Corporate Diversification." J. Econ. Perspectives 8 (3): 163–78.
- Morck, Randall, Andrei Shleifer, and Robert W. Vishny. 1990. "Do Managerial Objectives Drive Bad Acquisitions?" *J. Finance* 45 (1): 31–48.
- Panzar, John C., and Robert D. Willig. 1981. "Economies of Scope." A.E.R. 71 (2): 268–72.
- Penrose, Edith. 1955. "Limits to the Growth and Size of Firms." A.E.R. 45 (2): 531–43.
- Portes, Richard D. 1968. "Input Demand Functions for the Profit-Constrained Sales-Maximizer: Income Effects in the Theory of the Firm." *Economica* 35 (139): 233– 48.
- Rachapalli, Swapnika. 2021. "Learning between Buyers and Sellers along the Global Value Chain." Working paper.
- Robins, James, and Margarethe F. Wiersema. 1995. "A Resource-Based Approach to the Multibusiness Firm: Empirical Analysis of Portfolio Interrelationships and Corporate Financial Performance." *Strategic Management J.* 16 (4): 277–99.
- Rondi, Laura, and Davide Vannoni. 2005. "Are EU Leading Firms Returning to Core Business? Evidence on Refocusing and Relatedness in a Period of Market Integration." *Rev. Indus. Org.* 27 (2): 125–45.
- Scherer, Frederic M. 1982. "Inter-industry Technology Flows and Productivity Growth." *Rev. Econ. and Statis.* 64 (4): 627–34.
- Schott, Peter K. 2004. "Across-Product versus Within-Product Specialization in International Trade." Q.J.E. 119 (2): 647–78.
- Shaked, Avner, and John Sutton. 1990. "Multiproduct Firms and Market Structure." RAND J. Econ. 21 (1): 45–62.
- Stigler, George J. 1951. "The Division of Labor Is Limited by the Extent of the Market." J.P.E. 59 (3): 185–93.
- Sutton, John, and Daniel Trefler. 2016. "Capabilities, Wealth, and Trade." J.P.E. 124 (3): 826–78.
- Teece, David J. 1982. "Towards an Economic Theory of the Multiproduct Firm." *J. Econ. Behavior and Org.* 3 (1): 39–63.
- Tybout, James R. 2000. "Manufacturing Firms in Developing Countries: How Well Do They Do, and Why?" *J. Econ. Literature* 38 (1): 11–44.
- Vandenbussche, Hylke, and Christian Viegelahn. 2018. "Input Reallocation within Multi-product Firms." J. Internat. Econ. 114:63–79.
- Wernerfelt, Birger. 1984. "A Resource-Based View of the Firm." Strategic Management J. 5 (2): 171–80.
- Willig, Robert D., Steven C. Salop, and Frederic M. Scherer. 1991. "Merger Analysis, Industrial Organization Theory, and Merger Guidelines." *Brookings Papers Econ. Activity Microeconomics* 1991:281–332.