THE COMPARATIVE ADVANTAGE OF FIRMS

Johannes Boehm[a], Swati Dhingra[b], John Morrow[c]

July 16, 2021

ABSTRACT. Resource based theories propose that firms grow by diversifying into products which use common capabilities. We provide evidence for common input capabilities using a policy that removed entry barriers in input markets to show that the similarity of a firm's and industry's input mix determine firm production choices. We model industry choice and economies of scope from input capabilities. Estimating the model 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 co-production rates. Upstream entry barriers were equivalent to a 9.5% tariff on inputs.

JEL Codes: F11, L25, M2, O3.
Keywords: Multiproduct firms, firm capabilities, vertical input linkages, comparative advantage, economies of scope, size-based policies.

Acknowledgments. 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 for helpful comments. An earlier version of the paper was entitled "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 LSE - TISS - TATA Grant for financial assistance. Dhingra thanks the ERC for financial assistance under Starting Grant 760037. Shania Bhalotia and Yan Liang provided excellent research assistance.
[a]Sciences Po and CEP/LSE.
[b]Centre for Economic Performance, LSE, and CEPR.
[c]Kings College London, CEP/LSE, and CEPR. Corresponding Author: john.morrow@kcl.ac.uk.

1. INTRODUCTION

Multiproduct firms dominate production and export activity. They are much larger than their single product counterparts and their product turnover contributes substantially to aggregate output growth.[1] 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,[2] 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 co-production.

Early theoretical and empirical work recognises the sizable contribution of product diversification towards firm growth and aggregate productivity, and examines explanations for product diversification within firms (such as, Stigler (1951); Scherer (1982) and summarised in Chandler (1992); Montgomery (1994)). Explanations on the demand side include gaining market power through horizontal and vertical integration or internalising demand complementarities and network externalities across products (for example, Willig et al. (1991); Bernheim and Whinston (1990); Jovanovic and Gilbert (1993); Hoberg and Phillips (2016)).

On the supply side, agency-based theories of the firm suggest diversification is motivated by internal labour and capital markets of the firm. For example, managers choose to diversify to reduce their human capital risk, to gain rents from utilising free cash flows, or to obfuscate when their own division is doing badly (Amihud and Lev (1981); Jensen (1986); Morck et al. (1990)). This explained early trends of reduced firm valuations from diversification (for example, 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 size of a single product market. Entering new products requires resources, such as knowhow or inputs, that are costly to acquire and to transfer outside the

[1]For example, in the United States, multiproduct firms account for over 90 per cent of manufacturing output and multiproduct exporters account for over 95 per cent of exports. They are larger than single product firms in the same industry in terms of shipments (0.66 log points), employment (0.58), labour productivity (0.08) and TFP (0.02). About 89 per cent of multi-product firms vary their product mix within five years and these changes in the product mix make up a third of the increase in US manufacturing output (Bernard et al. 2007, 2010). In India, multiproduct firms (that produce in more than one of 262 different industries) account for 32 per cent of firms and 62 per cent of sales (as we discuss later). Among publicly listed firms, Goldberg et al. (2009) find multiproduct firms, that produce in more than one of 108 4-digit NIC industries, make up 47 per cent of firms and 80 per cent of sales. They are 107 per cent bigger in output than single-product firms within the same industry.

[2]Bernard et al. (2010, 2011); Eckel and Neary (2010); Mayer et al. (2014); Iacovone and Javorcik (2010).

firm. Firms therefore gain economies of scope by diversifying into products that require similar knowhow 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); 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.[3] For example, recent micro-data reveal that firms are much more likely to produce in certain pairs of industries (Bernard et al. (2010)). Many firms that make Fabricated Metal Products also make Industrial Machinery. A challenge however has been to move beyond systematic correlations in co-production and product characteristics to disentangling evidence for specific theories of the firm. In the example, systematic co-production of metal and machinery could arise because they share common inputs and technologies like metal and metalworking (as in the resource based view) or because consumers who order fabricated metal products also need industrial machinery (as in demand side theories) or still further because machinery and metal is 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.

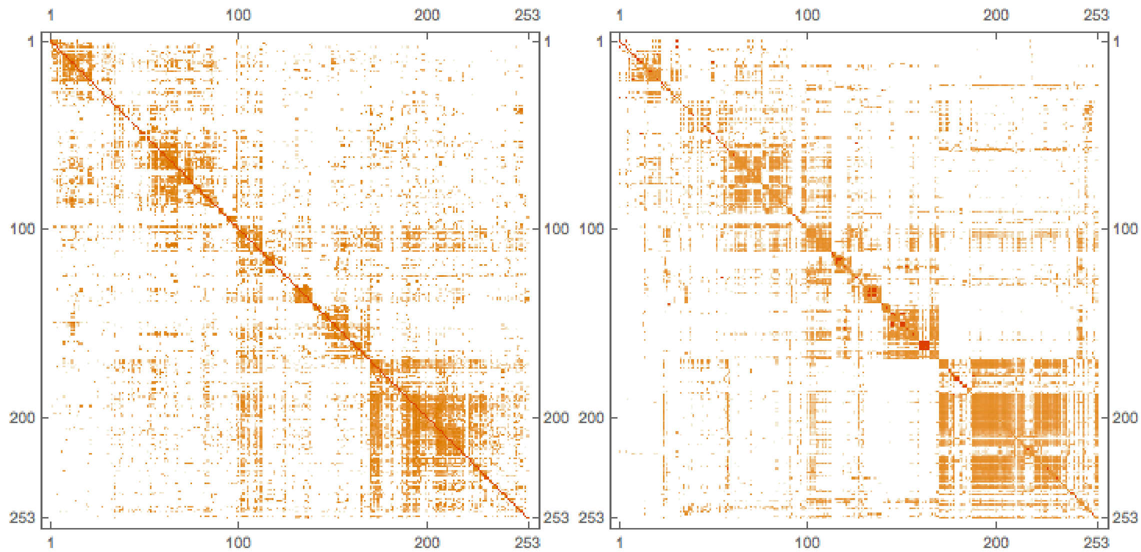
Using plant-level data from Indian manufacturing, this paper starts with the striking observation that product diversification within firms is systematically related to shared input use across these products. Figure 1.1 shows firms produce more in pairs of industries that require similar intermediate inputs. The left panel is the extent of co-production of industry pairs within plants (across 1 to 253 different industries) and the right panel is the input similarity between industry pairs.[4] This is also borne out in findings from the United States, where stark examples of industry pairs that are co-produced and that have similar input requirements include Textile and Apparel, Lumber and Paper, Primary Metal and Fabricated Metal (Bernard et al. (2010)).[5]

[3]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 (for example, Robins and Wiersema (1995); Bowen and Wiersema (2005); Bryce and Winter (2009); Fan and Lang (2000); Liu (2010); Rondi and Vannoni (2005)).

[4]The co-production 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.

[5]Similar patterns emerge in firm-level data from the United Kingdom and Belgium (Hutchinson et al. 2010, Bernard et al. 2018).

FIGURE 1.1. Co-production and Input Similarity



(A) Industry Co-production (B) Intermediate Input Similarity

The left matrix shows, for plants with primary sales in the row industry, the fraction of sales coming from products in the column industry. The right matrix shows the inner product between the row and column industry's intermediate input expenditure share vectors. Intermediate input shares (right matrix) are constructed from single-industry plants only. Darker colours indicate larger relative values within each sub-Figure.

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 nineties, 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 post-independence economic policy.[6] 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 which 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 Silk Textile production). In fact, even within the Cotton Apparel industry, a 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.

[6]The original aim of the reservation policy was employment generation through small scale units that were expected to be more labour intensive than larger firms (though Martin et al. (2017) show that the dismantling of this policy in fact generated relatively more employment).

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 knowhow (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 which 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, which 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 co-production. 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 co-production. Accounting for the different theories of the firm, we find that the dominant explanation in our context and policy episode is that firms specialise 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 co-production 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 which 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* industry.

The theory allows for love or hate for input variety and for Jevon's 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). A key insight of our framework is that economies of scope within multi-product firms imply 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 based on the 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, input-based comparative advantage makes single industry firms 5.2 percentage points 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, 2.3pp 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, size weighted premia range as high as

46.8pp, showing they are important for firm growth. Overall, input-based comparative advantage is quantitatively as important a determinant of firm entry into industries as time-invariant industry-pair determinants of co-production rates.

Related Literature. The results relate to the multiproduct firm literature, that 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.[7]

A large literature studies the role of access to inputs on firm productivity.[8] While we ask a different question, the focus on input supply is consistent with these studies. Specifically, Goldberg et al. (2009) 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. (2010) differentiates 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 anti-dumping measures by decreasing sales of products using these inputs. Similarly we find 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 Bloom et al. (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 which are similar in inputs, and that this channel contributes substantially to the variance in aggregate productivity. We

[7]See also Foster et al. (2008); Eckel and Neary (2010); Liu (2010); Dhingra (2013); Mayer et al. (2014) and Eckel et al. (2015) in the multiproduct literature and Hottman et al. (2016) and Bernard et al. (2021) in the firm heterogeneity literature.

[8]See, for example, Amiti and Konings (2007); Acemoglu et al. (2007); Kasahara and Rodrigue (2008); Kugler and Verhoogen (2009, 2012); Antras and Chor (2013); Halpern et al. (2015). In recent work, Lu et al. (2016) model the inherently dynamic process of accumulating input capabilities and its role in increasing firm productivity.

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 et al. (2007) and Hidalgo et al. (2007), which examine the product space of countries and the network structure of their products. They propose that products differ in the capabilities needed to make them and 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 non-monotonic 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 them, 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 (Martin et al. 2017; Garcia-Santana and Pijoan-Mas 2014; Galle 2015; Bollard et al. 2013). 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.[9] 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

[9]There are a growing number of studies relating linkages to productivity (see the handbook chapter by Combes and Gobillon (2015)). In particular, Lopez and Sudekum (2009) find that upstream, but not downstream, linkages are associated with higher productivity, perhaps in part due to the stronger effect of upstream linkages on product adoption that we find.

scale plants are equivalent to input tariffs of 9.5 per cent. Domestic policies, like size-based entry barriers, are well understood to be a non-tariff 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 in policies and correlation of policy changes with other shocks.[10] 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 which stress the importance of input linkages in amplifying micro shocks and policy effects.[11] 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 Rodriguez-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 due to 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 2 contains a description of the context, data and stylized facts. Section 3 shows the empirical relationship between input similarity and the industry mix of firms. Section 4 presents a model of capability choice and limit pricing suppliers, deriving structural estimation equations and an instrumentation strategy. Section 5 contains the results from estimation and quantification of input capabilities. Section 6 concludes.

2. Data and Stylized Facts

2.1. **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 of the Government of India. The ASI is the Indian government's main source of industrial statistics

[10]In their Handbook Chapter, Bown and Crowley (2016) summarize that "the existing literature and data sources are not sufficiently developed" to answer key questions like the extent to which domestic policies affect economic activity and how they compare with trade policy instruments.

[11]For example, Acemoglu et al. (2012), Di Giovanni et al. (2014), and early work by Jovanovic (1987) and Durlauf (1993).

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 of its history. In order to ensure consistency, we focus on the time frame of the fiscal years (April to March) 2000/01 to 2009/10.

The ASI has two unique aspects that make it particularly suitable for our analysis. Firstly, 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 reports 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.[12]

The three-digit industries are in 60 two-digit sectors. To give a sense of the level of detail in this classification, consider the sector "Cotton, Cotton yarn, and Fabrics" sector (ASIC 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 71). A comparison of the dimensionality of products with other sources and descriptive statistics are provided in Boehm et al. (2016, 2018).

Secondly, 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

[12] According to the ASI, the product classification is stratified into 2-digit sectors, 3-digit industries and 5-digit products.

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 multi-plant firm with sister plants that file separate survey returns. With that caveat in mind, we call the units of observation in our data "firms".

2.2. **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.

2.2.1. *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 for 62.2% of all sales. Firms that span three or more industries (11.2% of all observations) still account for more than 41% of total sales. This fact is well known and mirrors the results reported by Bernard et al. (2010) for the United States and by Goldberg et al. (2009) for the set of listed Indian firms.

TABLE 1. Frequency and Sales Shares of Multi-Industry Firms

		2-digits			3-digits		
		Obs	% Firms	% Sales	Obs	% Firms	% Sales
# of Industries	1	250028	81	50	208881	68	38
	2	43048	14	28	63997	21	23
	3	10113	3	12	22723	7	14
	4	2972	1	7	6843	2	8
	5	864	0	2	2835	1	6
	6	216	0	1	1198	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

Note: Observations are firm-years. Source: Authors' calculations from ASI data.

2.2.2. *Co-production Is Not Random.* We now turn to the question of which industries the firms produce in. Figure 1.2a in the Introduction shows two matrices. The left matrix shows the degree of co-production between industries. Each row contains 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. Hence, by construction, the diagonal contains the highest value in each row. There is substantial co-production across industries, as indicated by the off-diagonal dark areas. In particular, there is much co-production occurring within the metal product and machinery manufacturing sectors (the large shaded square on the bottom right), in the chemicals and pharmaceuticals industries (the industries with indices between 55 and 93), as well as within the textiles and apparel sectors (150 to 170). Firms from a diverse range of industries choose to have auxiliary outputs from the plastic and rubber industries (columns 100 to 112). These patterns are similar to the co-production documented by Bernard et al. (2010) for the United States.

The right panel of Figure 1.2a 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:

$$\overline{\text{IS}}_{kk'} = \sum_i \bar{\theta}_{ik} \bar{\theta}_{ik'}$$

where $\bar{\theta}_{ik}$ is the sum of expenditure of single-industry firms that only produce 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 industry 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 fibres 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.

3. The Input Mix and Comparative Advantage of Firms

Motivated by the strong positive relationship between co-production 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 4. 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.

3.1. **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, θ_j, with the vector of intermediate input expenditure shares of an industry k:

$$\text{inputSimilarity}_{jk}^{t} = \sum_{i=1}^{N} \theta_{ij}^{t} \overline{\theta}_{ik}$$

where i indexes the expenditure shares of spending on three-digit inputs and t denotes time. We construct the aggregate intermediate input shares $\bar{\theta}_{ik}$ by aggregating up the micro-data of single-industry plants that only produce in industry k. The input similarity measure ranges from zero, when firm j and sector k have no three-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 above in Section 2.2.2 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 θ_{ij}. The firm-industry input similarity measure is related to the measure of technological proximity of Bloom et al. (2013). Our model in Section 4 will provide a structural interpretation to the measure as the part of firm-level comparative advantage that comes from shared capabilities in intermediate input use.

3.2. **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).[13] Our baseline specification is a linear model for the probability of firm j adding industry k between time t and $t+1$:

$$\text{Add}_{jt}^{k} = \beta \cdot \text{inputSimilarity}_{jk}^{0} + \alpha_{jt} + \alpha_{k}^{t} + \alpha_{kk'}^{t} + \varepsilon_{jk}^{t} \tag{3.1}$$

Here, Add_{jt}^{k} is one if and only if firm j does not produce in industry k at time t, but does at time $t+1$; α_{jt} is a firm-time fixed effect which captures the average rate of adding industries for

[13]That said, the data on reported intermediate input use in the ASI is the expenditure on intermediate inputs that is being *consumed* in the current year. Hence, purchases of inventories should not show up in these variables.

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. α_k^t is an industry-time fixed effect which 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 fixed effects, $\alpha_{kk'}^t$, 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, ε_{jk}^t 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 (3.1), 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 impact 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 shocks (supply, demand, technology, infrastructure, etc.) that might affect industry co-production is accounted for and what remains are estimates of the direction of intra-industry 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.

TABLE 2. Industry Entry: Input Similarity

	Dependent variable: Add_{jkt}		
	(1)	(2)	(3)
$\text{InputSimilarity}^0_{jk}$	0.0391^{**}	0.0383^{**}	0.0281^{**}
	(0.00036)	(0.00036)	(0.00057)
Firm × Year FE α_{jt}	Yes	Yes	Yes
Industry × Year FE α_{kt}		Yes	
$k \times k' \times t$ FE $\alpha_{kk't}$			Yes
R^2	0.00969	0.0117	0.0575
Observations	52691029	52691029	52666907

Standard errors in parentheses, clustered at the firm-industry level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

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 cost shares is associated with a 0.38 percentage points (or 75%) higher entry probability.[14]

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.

3.3. **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.[15] 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.

[14]To put this number in perspective, a firm experiencing a dereservation of one of its inputs (see sub-section 3.3) 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-highest affected output industries. In distributional terms, a one standard deviation increase in input similarity is associated with a 174% higher industry entry rate.

[15]Development Commissioner, MSME, India (2018). Available at http://www.dcmsme.gov.in/publications/reserveditems/resvex.htm

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 post-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 due to the limitations imposed by SSI reservation. Consequently, the government was repeatedly advised to de-reserve products from the SSI list (Hussain, 1997). Over the course of the year 1997 to 2008, the government dereserved almost all products (see Table 3). The remaining 20 products were dereserved in 2015.

TABLE 3. Dereservation of Products, By Year

Year	1997	1999	2001	2002	2003	2004	2005	2006	2007	2008	2010	2015
# Products	15	9	15	51	75	85	108	180	212	107	1	20

Source: Government of India, Ministry of Micro, Small and Medium Enterprises

The definition of small scale industries, and therefore the scope of reservation, changed over the period during which the reservation was in place. In 1955, SSI was defined as establishments with fixed investments of less than Rs 500,000 which 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 Rs 10 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) 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 et al. (2017) find that the aggregate employment response is on average above 40%, output increased by about 30%, wages by 6%, and 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 et al. 2018).

In contrast to the existing literature, we use the dereservation as an unexpected change in the conditions that firms face on intermediate input 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 eight to twelve percent upon dereservation. (see 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.

3.4. **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 one if and only if firm j at some point uses a five-digit in the three-digit category i that has been dereserved during or before year t'. We then interact the similarity measures by these dereservation indicators as follows:

$$(\text{InputSimilarity-Dereservation})^{t}_{jkt'} = \sum_{i=1}^{N} \delta_{ijt'} \theta^{t}_{ij} \overline{\theta}_{ik} \tag{3.2}$$

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 3.1 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.[16] Later, the structural estimation provides a tariff equivalent for dereservation.

[16]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 Appendix C.2.

TABLE 4. Industry Entry: The Impact of Dereservation

	Dependent variable: Add_{jkt}			
	(1)	(2)	(3)	(4)
$\text{InputSimilarity}^0_{jk}$	0.0379^{**}	0.0371^{**}	0.0273^{**}	0.0268^{**}
	(0.00036)	(0.00036)	(0.00057)	(0.00058)
$\text{InputSimilarity-Dereservation}^0_{jkt}$	0.0429^{**}	0.0424^{**}	0.0203^{**}	0.0192^{**}
	(0.0025)	(0.0024)	(0.0024)	(0.0024)
$\text{InputSimilarity-Tariff}^0_{jkt}$				-0.0701^{**}
				(0.0095)
Firm $\times$ Year FE α_{jt}	Yes	Yes	Yes	Yes
Industry $\times$ Year FE α_{kt}		Yes		
$k \times k' \times t$ FE $\alpha_{kk't}$			Yes	Yes
R^2	0.00981	0.0118	0.0575	0.0576
Observations	52691029	52691029	52666907	52666907

Standard errors in parentheses, clustered at the firm-industry level.
$^+ p < 0.10$, $^* p < 0.05$, $^{**} p < 0.01$

3.5. **Case Study.** Dereservation reduced firm's 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 per cent 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, 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 USD 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 colouring, certain spices and flavoured oil. Following the dereservation, Akay expanded its product offerings to new products, which rely heavily on dereserved inputs, such as spiceuticals (spice-base health supplements) and various oleoresins (which are semi-solid spice oils such as capsicum oleoresin and cardamom oleoresin). Therefore, building on its earlier product portfolio, Akay has scaled up operations in products which use related dereserved inputs. Similar examples of moving towards 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.

3.6. **Alternative Theories of Product Diversification.** While the previous sub-sections control for a rich set of fixed effects to account for unobserved differences across firms, industries and industry-pairs over time, co-production could arise due to 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.

3.6.1. *Vertical Diversification Measures.* Firms could diversify up and down their value chain to gain vertical efficiency (for example, Stigler 1951; Chandler 1992). We use the constructed input-output shares $\bar{\theta}$ to measure whether a sector k is upstream or downstream from the firm's current product mix. Let σ_{zj}^{t} 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:

$$\text{upstream}_{jk}^{t} = \sum_{z=1}^{N} \sigma_{zj}^{t} \bar{\theta}_{kz}, \qquad \text{downstream}_{jk}^{t} = \sum_{z=1}^{N} \sigma_{zj}^{t} \bar{\theta}_{zk}. \tag{3.3}$$

where z runs over the set of three-digit industries. To make sense of these definitions, consider the following analogy: imagine a firm j with observed sales shares σ_j^t. Then given the firm's output mix σ_j^t and the industry's average input expenditures for these outputs, one would expect the expenditure share upstream of j on k to be upstream_{jk}^t. 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_{jk}^t is proportional to the expected expenditure share of downstream industry k on a firm with the output mix σ_j^t. 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 downstream car industry.

3.6.2. *Output Similarity Measures.* Firms might also enjoy other complementarities in outputs, by which firms who produce in one, 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:

$$\text{outputSimilarity}_{jk}^t = \sum_{i=1}^{N} \sigma_{ij}^t \overline{\sigma}_{ik},$$

where i runs over the set of three-digit industries. The vector $\overline{\sigma}_k$ 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 analogously to the input similarity measure in equation (3.2).

Output similarity summarizes similar distributions of sales, which would be implied by the horizontal diversification motives of demand-side theories (for example, Brander and Eaton 1984; Shaked and Sutton 1990; Willig et al. 1991; Jovanovic and Gilbert 1993; Dhingra 2013; Bernard et al. 2018). If firms diversify to internalise 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 co-production 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.[17] For our purposes, this means 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.

Output and input similarity would be positively correlated with each other when the latter is an important contributor to co-production, as suggested by Figures 1.2a and 1.2b. 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).

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

TABLE 5. Industry Entry: Robustness

	Dependent variable: Add_{jkt}				
	(1)	(2)	(3)	(4)	(5)
$\text{InputSimilarity}^0_{jk}$	0.0379**	0.0251**	0.0245**	0.0199**	0.0195**
	(0.00036)	(0.00044)	(0.00045)	(0.00057)	(0.00057)
$\text{InputSimilarity-Dereservation}^0_{jkt}$	0.0429**	0.0383**	0.0378**	0.0155**	0.0145**
	(0.0025)	(0.0024)	(0.0024)	(0.0023)	(0.0023)
$\text{OutputSimilarity}^0_{jk}$		0.0136**	0.0136**	0.100**	0.100**
		(0.00063)	(0.00063)	(0.0018)	(0.0018)
$\text{OutputSimilarity-Dereservation}^0_{jkt}$		0.0344**	0.0334**	0.0171**	0.0171**
		(0.0016)	(0.0016)	(0.0022)	(0.0022)
Upstream^0_{jk}		0.0335**	0.0315**	0.0291**	0.0291**
		(0.00092)	(0.00092)	(0.0030)	(0.0030)
Downstream^0_{jk}		-0.00826**	-0.00756**	-0.00351*	-0.00356*
		(0.00056)	(0.00056)	(0.0014)	(0.0014)
$\text{InputSimilarity-Tariff}^0_{jkt}$					-0.0640**
					(0.0095)
Firm × Year FE α_{jt}	Yes	Yes	Yes	Yes	Yes
Industry × Year FE α_{kt}			Yes		
$k \times k' \times t$ FE $\alpha_{kk't}$				Yes	Yes
R^2	0.00981	0.0122	0.0140	0.0646	0.0646
Observations	52691029	52691029	52691029	52666907	52666907

Standard errors in parentheses, clustered at the firm-industry level.
$^+ p < 0.10$, $^* p < 0.05$, $^{**} p < 0.01$

Table 5 shows the result of estimating equation (3.1) controlling for the output similarity variable, the dereservation-weighted version of it, and for 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. This is not entirely surprising, and could be suggestive of other complementarities, such as demand complementarities, motivating product diversification. Most importantly, however, the estimated coefficients of input similarity and dereservation-weighted input similarity remain positive and statistically significant across different specifications.

3.6.3. *Alternative Output Complementarity Measures.* A non-parametric way of capturing demand complementarities is to include a vector of output industry mix $\times$ 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 demand complementarities. The coefficient of the input similarity measures remain positive and statistically significant. It is slightly smaller than the baseline specification with $k \times k' \times t$ fixed effects (in Column (3) of Table 4), as might be expected because the average input similarity effect is subsumed in the new fixed effects.

TABLE 6. Industry Entry with Output-Mix Indicators

	Dependent variable: Add_{jkt}			
	(1)	(2)	(3)	(4)
$\text{InputSimilarity}^0_{jk}$	0.0145**	0.0143**	0.0147**	0.0145**
	(0.00058)	(0.00058)	(0.00058)	(0.00058)
$\text{InputSimilarity-Dereservation}^0_{jkt}$	0.0125**	0.0121**	0.0122**	0.0118**
	(0.0025)	(0.0025)	(0.0025)	(0.0025)
$\text{InputSimilarity-Tariff}^0_{jkt}$		-0.0268**		-0.0269**
		(0.010)		(0.010)
$\text{OutputSimilarity}^0_{jk}$			-0.0228**	-0.0228**
			(0.0038)	(0.0038)
$\text{OutputSimilarity-Dereservation}^0_{jkt}$			0.00659**	0.00660**
			(0.0025)	(0.0025)
Upstream^0_{jk}			0.00385	0.00388
			(0.0080)	(0.0080)
Downstream^0_{jk}			-0.0000240	-0.0000356
			(0.0025)	(0.0025)
Firm $\times$ Year FE α_{jt}	Yes	Yes	Yes	Yes
$k \times K(j) \times t$ FE $\alpha_{kK(j)t}$	Yes	Yes	Yes	Yes
R^2	0.123	0.123	0.123	0.123
Observations	47136891	47136891	47136891	47136891

Standard errors in parentheses, clustered at the firm-industry level.
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Table 6 confirms that input similarity is not confounded with demand-side 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.[18] It also reveals that firms move into industries similar in output to their main industry, but dissimilar in output to 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.

3.6.4. *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 based on 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 $\text{InputSimilarity}^0_{jk}$ 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 using log sales at time t instead of at the time of first observation. Therefore there is little evidence to support that weaker firms select into input-similar industries.

[18]See, for instance, Mayer et al. (2020) who use a similar approach to control for firm-export destination effects.

TABLE 7. Revealed Comparative Advantage – Sales Interactions

	Dependent variable: Add_{jkt}				
	(1)	(2)	(3)	(4)	(5)
$\text{InputSimilarity}^0_{jk}$	-0.0346**	-0.0348**	-0.0328**	-0.0396**	-0.0382**
	(0.0029)	(0.0028)	(0.0029)	(0.0029)	(0.0030)
$\text{InputSimilarity}^0_{jk} \times \text{logSales}^0_{jk}$	0.00416**	0.00412**	0.00398**	0.00385**	0.00375**
	(0.00017)	(0.00017)	(0.00017)	(0.00017)	(0.00018)
$\text{InputSimilarity-Dereservation}^0_{jkt}$	-0.0102	-0.00888	-0.00314	-0.0180	-0.0139
	(0.020)	(0.020)	(0.020)	(0.019)	(0.019)
$\text{InputSimilarity-Dereservation}^0_{jkt} \times \text{logSales}^0_{jk}$	0.00285*	0.00275*	0.00236*	0.00203+	0.00177+
	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0011)
$\text{InputSimilarity-Tariff}^0_{jkt}$			-0.156		-0.0898
			(0.10)		(0.094)
$\text{InputSimilarity-Tariff}^0_{jkt} \times \text{logSales}^0_{jk}$			0.00470		0.00238
			(0.0053)		(0.0049)
Firm × Year FE α_{jt}	Yes	Yes	Yes	Yes	Yes
Industry × Year FE α_{kt}		Yes	Yes		
$k \times k' \times t$ FE $\alpha_{kk't}$				Yes	Yes
R^2	0.0100	0.0120	0.0121	0.0577	0.0577
Observations	52691029	52691029	52691029	52666907	52666907

Standard errors in parentheses, clustered at the firm-industry level.
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

It may however be that firms that diversify into input-similar industries start to experience slower growth, which would again point towards weaker (low-growth) firms selecting into input similar industries. Table 8 studies the relationship between post-entry sales growth and input similarity, and Tables 25, 26, and 27 in the Appendix provide further checks with different sets of fixed effects to compare across firms by their entering and continuing industries. In particular, we examine the following specification:

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

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 three-year and five-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).[19] Table 8 includes industry-year fixed effects and therefore compares firms entering

[19]We choose this dependent variable because in any given year, smaller firms are only surveyed with about 20% probability. Given that we are conditioning on entry, sales in year $t+1$ are necessarily positive.

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 post-entry sales, and not a diversification discount as predicted by many agency-based theories.

TABLE 8. Post-entry Growth: Within Entering Industries

	Dep. var.: log $\left(\text{Avg. Sales}_{jk}^{t,\dots,t+3}\right)$			Dep. var.: log $\left(\text{Avg. Sales}_{jk}^{t,\dots,t+5}\right)$		
	(1)	(2)	(3)	(4)	(5)	(6)
InputSimilarity$^0_{jk}$	1.544**	1.489**	1.331**	1.740**	1.684**	1.526**
	(0.086)	(0.086)	(0.088)	(0.087)	(0.087)	(0.089)
InputSimilarity-Dereservation$^0_{jkt}$		1.431**	1.071**		1.467**	1.106**
		(0.33)	(0.33)		(0.33)	(0.33)
InputSimilarity-Tariff$^0_{jkt}$			-21.54**			-21.63**
			(1.32)			(1.34)
Industry $\times$ Year FE α_{kt}	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.181	0.181	0.186	0.188	0.189	0.194
Observations	55318	55318	55318	55318	55318	55318

Standard errors in parentheses, clustered at the firm-industry level.
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Notes: 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.

3.6.5. *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, like labour and capital, could also be shared across industries within the firm. This is explored in full detail in Section 5 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 post-entry sales growth specifications. We also show that results hold when focusing on (i) the set of large firms (100+ employees) that are sampled every year in the ASI; (ii) the set of firms that are single-plant firms; (iii) the sample when excluding industry-pairs (k, k') where there is never any co-production; (iv) the sample excluding few producing firms per industry-year; (v) the sample of firms that exclude those defined as wholesalers; and (vi) the sample of firms that are single product to begin with.

Finally, the results are also robust to changing the estimator from OLS to Logit to better account for the discrete nature of the dependent variable. We conclude that 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 firm level comparative advantage based on Input-Output mechanisms.

4. 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 input-specific 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 endogenises the flexible manufacturing hypothesis of Eaton and Schmitt (1994); Eckel and Neary (2010); Mayer et al. (2014), where unit costs of production rise as firms move away from their core competencies.[20]

The production model allows us to isolate upstream-downstream linkages and their role in multiproduct final good production. As we shall see, 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 Jevon'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 spiralling feedback

[20]This also rationalizes the finding of Fontagne et al. (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 due to country specific demand).

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 Jevon's Paradox, multiple supplier equilibria are possible, precisely as one would get with textbook 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.[21] 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 Panzar and Willig (1981) and Baumol (1977) 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 competence, which examination shows contains the Input Similarity measure

$$\underbrace{-\sum_i \left(\theta_{ijt}-\overline{\theta}_{ik}\right)^2}_{\text{Input Distance Penalty}} = 2 \underbrace{\sum_i \theta_{ijt}\overline{\theta}_{ik}}_{\text{Input Similarity}} - \underbrace{\sum_i \theta_{ijt}^2}_{\text{Hicks Neutral Capability Cost}} - \underbrace{\sum_i \overline{\theta}_{ik}^2}_{\text{Industry Effect}},$$

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

[21]The closest piece to our model of supplier behavior is de Blas and Russ (2015) who use Frechet cost draws and limit pricing for a discrete number of firms who sell to end consumers and our setting delivers analogues 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 Jevon's Paradox and multiple or non-existence of equilibria while still delivering a tractable model that can be readily estimated.

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 extent to which input-driven economies of scope explain the portfolios of multiproduct firms.

4.1. **Demand and Unit Costs.** In what follows, we take all prices in terms of a numeraire input or commodity (e.g. labor). 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[22]

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

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 return to scale Cobb-Douglas technology with industry input expenditure shares $\overline{\theta}_{ik}$ and industry productivity labeled φ_{jk}.[23] 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 φ_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

$$c_{jkt} \equiv \prod_i \left(S_{ijt}\psi_{it}/\overline{\theta}_{ik}\varphi_{jk}\right)^{\overline{\theta}_{ik}}.$$

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

4.2. **Unit Costs and Capabilities.** Inputs M_{ijk} at the industry level are a composite of quantities $m_{\iota ijkt}$ of varieties, indexed by ι. M_{ijk} is the CES aggregator of varieties of input i:

$$M_{ijkt}^{(\sigma-1)/\sigma} = \int_0^\infty m_{\iota ijkt}^{(\sigma-1)/\sigma} d\iota \tag{4.1}$$

where variety ι of input i has a price $s_{\iota it}$. Firms have *capabilities* of using inputs with prices $\left[\underline{c}_{ijt}, \infty\right)$ where $\underline{c}_{ijt}$ is chosen by the firm. Here lower $\underline{c}_{ijt}$ corresponds to both a greater variety of inputs

---

[22] As is well known, this structure can be microfounded with CES preferences over varieties at the industry level. How one chooses to aggregate across industries has implications for the patterns across $\{D_{kt}\}$ each period. We remain agnostic to allow for flexibility in the estimation.

[23] In keeping with this section's focus on input capabilities, φ_{jk} could be modeled as a Cobb-Douglas aggregation of firm-input productivities: $\varphi_{jk} = \prod_i A_{ij}^{\overline{\theta}_{ik}}$.

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 $\underline{c}_{ijt}$.

Suppliers can enter input market i by paying a fixed cost f_s and receive a cost draw b_ι with $\Pr(b_\iota \geq b) = (b/s_m)^{-\lambda}$ with $0 < \lambda < 1$. The resulting mass of entrants is N_{it}. Suppliers are monopolistically competitive across varieties ι, but within varieties inputs are perfect substitutes and so suppliers engage in limit pricing akin to Bernard et al. (2003). Supplier ι chooses price $s_{\iota it}$, supplying a quantity $m_{\iota ijt}$ to firm j and earns profits $\zeta_{\iota it}$ by selling to any interested downstream firms with the capability to purchase their variety. Since the minimum cost draw among N_{it} entrants is $\Pr(b_{\iota it} \geq b) = (b/s_m)^{-\Omega_{it}}$ 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.

In what follows, we will assume $N_{it} \geq \max\left\{1 + \frac{\sigma-2}{\lambda}, 1\right\}$ which is a continuous analogue of having at least two competitors, and is sufficient for limit pricing effects from entry. While supplier entry is endogenous, we will characterize when downstream demand from final good firms is sufficient to ensure this condition. In this setting, a firm's optimal choice of inputs can be summarized by the following Proposition (all proofs are in the Appendix):

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

(1) *The price index for input i for firm j is a function of capabilities $\underline{c}_{jit}$ and supplier entry N_{it} is*

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

where $\vartheta_{it} \equiv 1 + \frac{\sigma-1}{\lambda(N_{it}-1)+\sigma-1}\left(1 - ((\sigma-1)/\sigma)^{\lambda(N_{it}-1)+\sigma-1}\right)$.

(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),*

(b) *complements ($\sigma < 1$), decreasing varieties lowers costs (Hate for Variety).*

(3) *Unit costs c_{jkt} are given by*

$$c_{jkt} = \underbrace{\frac{1}{\varphi_{jk}}}_{\text{Firm}-\text{Industry }(jk)} \underbrace{\prod_i \left(\psi_{it}\left(\frac{\Omega_{it}\vartheta_{it}}{\Omega_{it}+(\sigma-1)}\right)^{1/(1-\sigma)} \frac{s_m^{\Omega_{it}/(1-\sigma)}}{\overline{\theta}_{ik}}\right)^{\overline{\theta}_{ik}}}_{\text{Supplier }(kt)} \underbrace{\prod_i \left(\underline{c}_{ijt}^{1-\Omega_{it}/(1-\sigma)}\right)^{\overline{\theta}_{ik}}}_{\text{Capability }(jkt)}.$$

We next discuss the interplay between capabilities and supplier entry, and the conditions under which Jevon's Paradox is avoided.

4.3. **Supplier Entry and Jevon's Paradox.** Since ρR_{jk} are expenditures by a firm j in industry k and $\overline{\theta}_{ik}$ 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* $\mathcal{L}_{jit} = \mathrm{E}\left[\min\left\{1/\sigma, 1 - b_1/b_2\right\} \mid b_1 \geq \underline{c}_{jit}\right]$ where b_1/b_2 is the ratio of the lowest to second lowest cost draw of suppliers of the same variety. While the Lerner index varies by firm, the aggregate Lerner index only depends on the intensity of supplier competition, even though each firm buys from a fraction $\left(s_m/\underline{c}_{jit}\right)^{\lambda N_{it}}$ of all suppliers. Compared to the usual role of the Lerner index $1/\sigma$ under CES preferences, this implies 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 industry revenues weighted by active suppliers into profits,*

$$\mathcal{L}_{it} \equiv \frac{\mathrm{E}\left[\zeta_{\imath it}\right]}{\sum_j \sum_k \rho\overline{\theta}_{ik}R_{jkt}\left(s_m/\underline{c}_{jit}\right)^{\lambda N_{it}}} \tag{4.2}$$

decreases in entry, approaches zero and is given by

$$\mathcal{L}_{it} = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{-\lambda(N_{it}-1)} + \frac{1}{\lambda\left(N_{it}-1\right)+1}\left(1-\left(\frac{\sigma}{\sigma-1}\right)^{-\lambda(N_{it}-1)-1}\right) \leq \frac{1}{\sigma-1}.$$

The aggregate Lerner index functions like an average markup that decreases with entry due to increased competition from limit pricing. The free entry condition is expected supplier profits per mass of entrants $\mathrm{E}\left[\zeta_{\imath it}\right]/N_{it}$ equal to the entry cost f_s, which with Equation (4.2) gives the entry condition:

$$\mathcal{L}_{it} = \frac{f_s}{\sum_j \sum_k \rho\overline{\theta}_{ik}R_{jkt}\left(s_m/\underline{c}_{jit}\right)^{\lambda N_{it}}/N_{it}}. \tag{4.3}$$

Equation (4.3) will obtain if the left and right 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, i.e. average weighted revenues $\sum_j \sum_k \rho\overline{\theta}_{ik}R_{jkt}\left(s_m/\underline{c}_{jit}\right)^{\lambda N_{it}}/N_{it}$ decreases.

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 *Jevon'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 spiralling 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 Jevon'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 takings downstream as measured by the Lerner index $1-\rho$ is greater than the upper bound $1/\sigma$ for all upstream Learner indexes, then the Paradox is avoided:[24]

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* $(1-\rho)\,\sigma \geq 1$ *which guarantees that downstream markups are high enough to absorb the upstream cost reductions, avoiding Jevon's Paradox.*

Returning to Equation (4.3) with Propositions 2 and 3 in hand, one can picture the aggregate Lerner index drawn across levels of supplier entry, dropping down from $1/\left(\sigma-1\right)$ towards an index of zero where limit pricing consumes all industry profits per unit mass of suppliers and average weighted revenues fall towards zero, sweeping the right hand side of Equation (4.3) unboundedly upwards. 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 $\mathrm{E}\left[\zeta_{\iota it}\right]$ (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 Jevon'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 (4.3)

[24] In fact, if the "No Jevon's Paradox Condition" $(1-\rho)\,\sigma \geq 1$ 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.

to hold, supplier entry must increase, potentially leading to multiple equilibria.[25] 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* $N_{it} = 1 + \frac{\sigma-2}{\lambda}$. *Supplier entry drives average expected profits monotonically down to zero, including demand feedbacks from lower input costs, provided* $\sigma > 1 + \lambda$ *and* $(1 - \rho)\sigma \geq 2$ *which further dampens Jevon's Paradox.*

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 further depend on demand structures (see Dhingra and Morrow 2019), and general results are not likely because of the non-monotonicities 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 (4.3) 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 (4.3) 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 side of Equation (4.3) and increasing the equilibrium level of supplier entry.[26] This yields the following Proposition.

Proposition 5. *Under the assumptions of Proposition 4:*

(1) *Decreases in supplier entry costs increase supplier entry.*
(2) *Decreases in tariffs increase supplier entry.*

4.4. **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

[25]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.

[26]This can be modelled as a decrease in iceberg transport costs or a first order stochastic shift downwards in the cost distribution since constant markups and limit pricing will pass this through into lower prices.

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.

We assume that all firms have an innate capability for inputs from industry i, $\underline{c}_{i0}$, and can adjust this capability due to demand and supply conditions subject to a Hicks neutral cost across production in all industries.[27] Letting $\underline{c}_{jt}$ denote the vector of acquired capabilities, the actual unit costs of a multiproduct firm are given by $\gamma\left(\underline{c}_{jt}\right)c_{jkt}$ in each industry, where

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

A firm can use its acquired capabilities across any number of products and re-optimizes by choosing $\underline{c}_{ijt}$ each period. In order to simplify the subsequent notation, we normalize $\underline{c}_{i0} = 1$.[28]

4.4.1. *Profits and Revenues.* The profit function of firm j at time t across all industries k is

$$\pi_{jt} = \sum_k \pi_{jkt} = \sum_k p_{jkt}q_{jkt} - \sum_k\sum_i \gamma\left(\underline{c}_{jt}\right) S_{it}M_{ijkt} = \sum_k \left(D_{kt}q_{jkt}^{\rho} - \gamma\left(\underline{c}_{jt}\right) c_{jkt}q_{jkt}\right).$$

A firm's profit maximizing capability and production choices considering product markets jointly are summarized in the following Proposition:

Proposition 6. *Assume $\lambda N_{it} > 1-\sigma$. For firm-input expenditure shares θ_{ijt}, the optimal capability choice is*

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

[27]The innate capability is assumed to be common for econometric reasons. It can be heterogeneous but will then need to be estimated with fixed effects beyond the combination of industry-time.

[28]This will not influence our estimating equations as it is an industry-time effect.

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

$$\ln R_{jkt} = \underbrace{\ln\left(\rho^{\frac{\rho}{1-\rho}} D_{kt}^{\frac{1}{1-\rho}}\right)}_{\text{Demand }(kt)} - \underbrace{\frac{\rho}{1-\rho}\sum_i \overline{\theta}_{ik} \ln \psi_{it} \vartheta_{it}^{\frac{1}{1-\sigma}} \left(1-\Theta_{it}^{-1}\right)^{\frac{1}{1-\sigma}} \frac{s_m^{\Theta_{it}-1}}{\overline{\theta}_{ik}}}_{\text{Supplier }(kt)}$$

$$(4.4) \qquad + \underbrace{\frac{\rho}{1-\rho} \ln \varphi_{jk}}_{\text{RCA }(jk)} + \underbrace{\underbrace{\frac{\rho}{2(1-\rho)} \sum_i \Theta_{it}^2 \overline{\theta}_{ik}^2}_{\text{Supplier}-\text{Tech }(kt)} - \underbrace{\frac{\rho}{2(1-\rho)} \sum_i \Theta_{it}^2 \left(\theta_{ijt} - \overline{\theta}_{ik}\right)^2}_{\text{Input Distance }(jkt)}}_{\text{Comparative Advantage }(jkt)}$$

with the dimension of variation listed below each term.

Since $\ln \underline{c}_{ijt} = -\Theta_{it}\theta_{ijt}$, 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 don't invest in capabilities for inputs they don't source, i.e. when $\theta_{ijt} = 0$. The addition to Equation (4.4) 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.[29] 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 Revealed Comparative Advantage (RCA) terms capture idiosyncratic advantages a firm has across industries which are static and can be estimated with Industry-Firm fixed effects, captured here with the interpretation of industry specific combinations of idiosyncratic input productivities.

The remaining Comparative Advantage term captures the dynamic re-deployment of input capability and is sensitive to the depth of input markets (through N_{it} in Θ_{it}). To interpret this term, 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

[29]Following Proposition 6, the effective mass of suppliers a firm chooses is

$$N_{it} \cdot \left(s_m / \underline{c}_{jit}\right)^{\lambda N_{it}} = N_{it} s_m^{\lambda N_{it}} e^{\lambda N_{it}(1+\lambda N_{it}/(\sigma-1))\theta_{ijt}}$$

Since the vector of expenditure shares θ_{jt} is a function of contemporaneous demand and supply conditions and comparative advantage, $\theta_{jt}\left(D_k, N_t, \phi_j\right)$, there is firm heterogeneity in effective suppliers and ranks of supplier purchases from capability choice stemming from the joint production and sourcing decision.

input shares θ_{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, i.e. $\Theta_{it} = \Theta$, the penalty takes the intuitive form of a coefficient times Euclidean distance squared, $\sum_i \left(\theta_{ijt} - \overline{\theta}_{ik}\right)^2$.

A final result stemming from profit maximizing behavior is how firms approximately update their core distance from changes in industry level demand $\{D_{kt}\}$, holding capabilities constant. We will 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 holding capabilities constant and optimally updating to respond to industry demand shocks through the relationship*

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

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

4.5. **Estimating Policy Effects.** Now consider an observable policy P that changes the depth of input markets of the form $\Omega_{it} = \lambda N_t = \Omega_{i0} + \alpha_P P_{it}$. Linearizing Equation (4.4) around the initial policy state Ω_{i0} and letting κ_x represent a fixed effect for characteristic x yields the following estimating equation:

$$\ln R_{jkt} = \kappa_{kt} + \kappa_{jk} + \underbrace{\frac{\rho}{1-\rho} \sum_i \left[\Theta_{i0}^2 + \frac{2\Theta_{i0}}{\sigma - 1} \alpha_P \left(P_{it} - P_{i0}\right)\right] \left(\theta_{ijt}\overline{\theta}_{ik} - \theta_{ijt}^2/2\right)}_{\text{Comparative Advantage } (jkt)}. \tag{4.5}$$

The theory implies that Θ_{it} has the same sign as $\sigma - 1$, so estimating $\alpha_P \cdot \Theta_{i0} / \left(\sigma - 1\right)$ allows for testing hypotheses about the sign of α_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) α_B within the three digit level (with B_{ijt} equal to 1 if a five digit product the firm ever uses is reserved in industry i and zero otherwise) and a linear effect α_τ of tariffs on entry for three digit tariffs τ_{ijt} (these are aggregated at the firm level from observed average firm level imports at the five digit level).

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

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

In light of the theory above, we can interpret these policy shifts as changing the depth of input markets with theory signing both α_B and α_τ to be negative, so that with no entry barriers and zero tariffs, $\Omega_{i0} = \overline{\Omega}$ is the 'maximal' market depth. Therefore Equation (4.5) approximates around a policy space of no entry barriers and no tariffs. This then implies the estimating equation

$$\ln R_{jkt} = \kappa_{kt} + \kappa_{jk} + \kappa_0 \sum_i \left(\theta_{ijt}\overline{\theta}_{ik} - \theta_{ijt}^2/2\right) + \kappa_1 \sum_i \left(\alpha_B B_{ijt} + \alpha_\tau \tau_{ijt}\right)\left(\theta_{ijt}\overline{\theta}_{ik} - \theta_{ijt}^2/2\right). \tag{4.6}$$

with $\kappa_0 = \Theta_{i0}^2 \rho / (1-\rho)$, $\kappa_1 = 2\Theta_{i0}\rho / (1-\rho)(\sigma - 1)$. This estimating equation says that (log) firm revenues depend on industry-time Demand and Supply effects κ_{kt}, firm-industry effects κ_{jk}, a distance effect of industry k from the firm's core competency κ_0 and policy effects which 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 $\alpha_B \kappa_1 / \alpha_\tau \kappa_1 = \alpha_B / \alpha_\tau$. Because of the selection issues involved, we estimate the extensive margin of production implied by Equation (4.6). Firms will produce in industry k exactly when $R_{jkt} > (1-\rho) f_{kt}$, so we estimate Equation (4.6) as a linear probability model for the outcome that observed revenues of the firm-industry are positive each period.[30] As we are estimating probabilities, we can think of how comparative advantage shifts the *production probability frontier* of firms.

4.6. **Structural Instrumentation.** In Equation (4.6), firm expenditure shares θ_{ijt} are a function of time varying input prices ψ_{it}, demand shocks D_{kt}, firm-industry productivities φ_{jk} and fixed technology $\overline{\theta}_{ik}$. Input price and demand shocks are estimated through industry-time fixed effects.

[30] This can be naturally extended to an extensive margin formulation with a logit type model, see Appendix. We implement this for the structural form as a robustness check but have difficulties with IV-Logit due to the high dimensional parameter space and well known sensitivity of that estimator.

Productivities are estimated through firm-industry fixed effects, expressed as Revealed Comparative Advantage. 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.[31]

There might be omitted variables from our structural equation that cause θ_{ijt} to change, which could bias our estimates of the role of capabilities. For example, demand or cost shocks at more disaggregated levels 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_{jkt} *at the firm level* would be positively correlated with input similarity through the composition of firm activity.
- Contemporaneous supply shocks ψ_{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 $\overline{\theta}_{ik}$).

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 instrumental variable 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 on to firm behaviour through theory. Recovering these common demand shocks allows us to predict changes in θ_{ijt} based on shifts in the within firm distribution of activity.[32] In

[31]One potential concern is that dereservation systematically changes technology $\overline{\theta}_{ik}$, 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 three digit category on $\overline{\theta}_{ik}$ 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.

[32]In doing so, we will hold the role of capabilities constant in the instrumentation stage to avoid non-linearity as the full expression for input similarity is recursive. Even assuming common input markets for all inputs ($\lambda N_{it} = \Omega$), the expression becomes

$$\sum_i \overline{\theta}_{ih}\theta_{ijt} = \frac{\sum_i \theta_{ih} \sum_k \overline{\theta}_{ik} D_{kt}^{1/(1-\rho)} \left(s_{kt} \underline{c}_0^{-(1+\Omega/(\sigma-1))^2 \sum_i \overline{\theta}_{ih}\theta_{ijt}} / \varphi_{jk} \right)^{-\rho/(1-\rho)}}{\sum_k D_{kt}^{1/(1-\rho)} \left(s_{kt} \underline{c}_0^{-(1+\Omega/(\sigma-1))^2 \sum_i \overline{\theta}_{ih}\theta_{ijt}} / \varphi_{jk} \right)^{-\rho/(1-\rho)}}$$

fact, examining the estimating Equation (4.5), what is needed is not instruments for each θ_{ijt}, but rather an instrument for terms of the form $\sum_i \left(\theta_{ijt}\overline{\theta}_{ik} - \theta_{ijt}^2/2\right)$and $\sum_i \left(P_{it} - P_{i0}\right)\left(\theta_{ijt}\overline{\theta}_{ik} - \theta_{ijt}^2/2\right)$. 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 revealed comparative advantage they contain) and then projecting them forward one period with a Bartik type instrument based on input expenditures from the Input-Output table.[33] In the case of a single instrument for terms of the form $\left(\theta_{ijt}\overline{\theta}_{ik} - \theta_{ijt}^2/2\right)$, the first stage of an IV strategy following from Proposition 7 is then:

$$\sum_i \left(\theta_{ijt}\overline{\theta}_{ik} - \theta_{ijt}^2/2\right) = \kappa_{kt} + \kappa_{jk} + \lambda \sum_i \left(\theta_{ijt}\overline{\theta}_{ik} - \theta_{ijt}^2/2\right) - \gamma_{kt} \sum_i \chi_{jkt-1}\left(\theta_{ijt}\overline{\theta}_{ik} - \theta_{ijt}^2/2\right). \tag{4.7}$$

Equation (4.7) is composed of three parts: the fixed effects found in the main structural equation for revenues, a lagged term for the endogenous sum $\sum_i \left(\theta_{ijt}\overline{\theta}_{ik} - \theta_{ijt}^2/2\right)$, and linear adjustment based on predicted input share changes from lagged revenue shares χ_{jkt-1} and contemporaneous industry level demand shocks γ_{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 which predicts the change in $\sum_i \left(\theta_{ijt}\overline{\theta}_{ik} - \theta_{ijt}^2/2\right)$ 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 (4.7), one for the shares and two for their two policy interactions. For this 2SLS estimator, we also need a system which includes all instruments in each first stage prediction equation. Accordingly, define both $\widetilde{\theta}_{ijkt} \equiv \left(\theta_{ijt}\overline{\theta}_{ik} - \theta_{ijt}^2/2\right)$ and $\widetilde{\chi}_{ijkt} \equiv \chi_{jkt}\left(\theta_{ijt}\overline{\theta}_{ik} - \theta_{ijt}^2/2\right)$ and the following sums for parameters λ

---

with $s_{kt} \equiv \prod_i \left(\psi_{it}\left(\Omega/\left(\Omega + (\sigma - 1)\right)\right)^{1/(1-\sigma)} s_m^{\Omega/(1-\sigma)}/\overline{\theta}_{ik}\right)^{\overline{\theta}_{ik}}$.

[33]While this instrumentation strategy works for contemporaneous shocks, if the omitted shocks are serially correlated, they would appear in the previous period terms which are the basis for the IV strategy and the exclusion restriction would fail. There are two potential approaches to this problem, either 1) explicitly modelling 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 non-linear in parameters estimator that presents computational issues. The latter suffers from a large loss in observations, which would select towards larger firms because of the nature of the data sample.

and the $K \times T$ vector γ:

$$I^M_{jkt+1} \equiv \lambda \sum_i \widetilde{\theta}_{ijkt} - \gamma \sum_i \widetilde{\chi}_{ijkt}, \;\; I^B_{jkt+1} \equiv \lambda \sum_i B_{ijt}\widetilde{\theta}_{ijkt} - \gamma \sum_i B_{ijt}\widetilde{\chi}_{ijkt}, \;\; I^\tau_{jkt+1} \equiv \lambda \sum_i \tau_{ijt}\widetilde{\theta}_{ijkt} - \gamma \sum_i \tau_{ijt}\widetilde{\chi}_{ijkt}.$$

The resulting first stage equations for our estimator are as follows:[34]

$$\sum_i \widetilde{\theta}_{ijkt} = \kappa_{kt} + \kappa_{jk} + I^M_{jkt}\left(\lambda^{11}, \gamma^{11}\right) + I^B_{jkt}\left(\lambda^{12}, \gamma^{12}\right) + I^\tau_{jkt}\left(\lambda^{13}, \gamma^{13}\right) + \eta_{jkt} \tag{4.8}$$

$$\sum_i B_{ijt}\widetilde{\theta}_{ijkt} = \kappa_{kt} + \kappa_{jk} + I^M_{jkt}\left(\lambda^{21}, \gamma^{21}\right) + I^B_{jkt}\left(\lambda^{22}, \gamma^{22}\right) + I^\tau_{jkt}\left(\lambda^{23}, \gamma^{23}\right) + \eta^B_{jkt} \tag{4.9}$$

$$\sum_i \tau_{ijt}\widetilde{\theta}_{ijkt} = \kappa_{kt} + \kappa_{jk} + I^M_{jkt}\left(\lambda^{31}, \gamma^{31}\right) + I^B_{jkt}\left(\lambda^{32}, \gamma^{32}\right) + I^\tau_{jkt}\left(\lambda^{33}, \gamma^{33}\right) + \eta^\tau_{jkt} \tag{4.10}$$

We implement the instrumental variable estimator of the structural coefficients in Equation (4.6) 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 γ_{kt} of the instruments in (4.8-4.10) and accordingly we do not report them. We correct for the well-known misspecification of the residual variance estimator in manual 2SLS (see Chapter 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.

5. 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 co-production rates, before ending with an extension of the structural estimates to primary factors.

5.1. **Structural Estimates.** Table 9 shows the OLS and IV estimates for the extensive margin version of Equation (4.6). The estimated coefficient on the deviation of the input similarity measure

[34] 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

is $\kappa_0 = 0.0086$ in the OLS, which rises to 0.1630 in the IV.[35] 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 of this. Both entry barriers and higher tariffs reduce the role of Input Distance since fewer suppliers disincentivise 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 percentage tariff on inputs.

TABLE 9. Structural Estimates for Multi-Industry Sales Premium

		Positive Sales for Plant j in Industry k ($R_{jkt} > 0$)			
		(1)	(2)	(3)	(4)
Input Distance	$\sum_i \left(\theta_{ijt}\overline{\theta}_{ik} - \theta_{ijt}^2/2\right)$	0.0085**	0.0086**	0.1362**	0.1630**
		(0.0002)	(0.0002)	(0.0229)	(0.0226)
Input Distance	$\sum_i B_{ijt} \cdot \left(\theta_{ijt}\overline{\theta}_{ik} - \theta_{ijt}^2/2\right)$	-0.0004**	-0.0004**	-0.0016**	-0.0016**
Entry Barriers		(0.0001)	(0.0001)	(0.0004)	(0.0004)
Input Distance	$\sum_i \tau_{ijt} \cdot \left(\theta_{ijt}\overline{\theta}_{ik} - \theta_{ijt}^2/2\right)$		-0.0005		-0.0168**
Tariffs			(0.0003)		(0.0027)
κ_{jk}		Yes	Yes	Yes	Yes
κ_{kt}		Yes	Yes	Yes	Yes
Estimator		OLS	OLS	IV	IV
N		77,745,382	77,745,382	46,185,150	46,185,150
R^2		0.762	0.762	0.760	0.760

Standard errors in parentheses, clustered at the firm-industry level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

5.2. **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

$$CA_{jkt} \equiv \hat{\kappa}_0 \sum_i \overline{\theta}_{ik}\theta_{ijt} + \hat{\kappa}_1 \sum_i \left(\hat{\alpha}_B B_{ijt} + \hat{\alpha}_\tau \tau_{ijt}\right) \overline{\theta}_{ik}\theta_{ijt}, \tag{5.1}$$

[35]For the IV sample, the OLS coefficients for the RHS variables are similar: 0.0092 (0.0002), -0.0004 (0.0002) and -0.0001 (0.0004) respectively. Relevant summary statistics are in Table 17 of the Appendix.

where parameters with a hat denote our IV estimates of the parameters. Note that due to 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. On average across firms and industries, CA increases the production probability by 4.3 percentage points, and for more than 13 percentage points in the top tenth percentile. 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.

TABLE 10. Input-based Comparative Advantage by Industry Sales Rank

Industry rank	Obs	Mean	p_{10}	p_{90}
1	307,294	0.054	0.004	0.153
2	98,413	0.026	0.001	0.071
3	34,416	0.017	0.000	0.040
4	11,693	0.013	0.000	0.032
5	4,850	0.011	0.000	0.028
6	2,015	0.010	0.000	0.028
7	817	0.009	0.000	0.024
8	278	0.009	0.000	0.024
9	95	0.008	0.001	0.018
10+	38	0.005	0.000	0.010
Total	459,909	0.043	0.002	0.132

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 1.1, 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 8.5pp, whereas the single-industry firms in the (perhaps technologically more similar) industry of Soft drinks and mineral water (152) would on average only get a 0.6pp 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, are not vertically related. Consider the Leather Bags and Purses industry (441), which is not vertically related to both Leather footwear (443) and Plastic footwear (423). Given the Leather footwear industry's shared input use of leather with the Leather Bags and Purses industry, its premium is 6.8pp, whereas the Plastic footwear industry's premium is only 0.4pp. Table 28 in Appendix E states the average CA with the highest premium for 25 industries. It shows 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 11. Input-based Comparative Advantage for the Second Industry

Comparative Advantage in: *Fruit and vegetable juices (135)*	
Edible fruits & nuts, edible vegetables (121)	8.5pp
Soft drinks & mineral water (152)	0.6pp
Comparative Advantage in: *Animal Oils & Fats (115)*	
Other produce of animal origin (119)	5.3pp
Vegetable oils and fats (125)	1.1pp
Comparative Advantage in: *Leather Bags and Purses etc. (441)*	
Leather footwear (443)	6.8pp
Plastic footwear (423)	0.4pp

Note: The table shows the average firm-level comparative advantage among single-industry plants of two contrasting industries for the italicized industry. "Other produce of animal origin" covers mostly bone, horn, and meals thereof.

Table 12 further highlights the core competencies feature of input-based 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 5.2pp to the production probability. Firms that produce in two industries experience a 6pp premium on their core industry and about half of that, 2.9pp, 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. Core Competency Sales Premium from Input-Based Comparative Advantage

Industry rank	# of Industries With Positive Sales									
	1	2	3	4	5	6	7	8	9	10+
1	0.052	0.060	0.061	0.033	0.026	0.021	0.020	0.019	0.014	0.020
2		0.029	0.023	0.018	0.017	0.016	0.014	0.015	0.010	0.022
3			0.019	0.015	0.013	0.013	0.011	0.011	0.014	0.015
4				0.013	0.013	0.012	0.011	0.009	0.010	0.016
5					0.011	0.011	0.011	0.009	0.011	0.009
6						0.010	0.010	0.010	0.010	0.006
7							0.010	0.009	0.009	0.007
8								0.009	0.008	0.008
9									0.008	0.009
10+										0.005

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 typical single-product firms. The single-industry premium from acquiring capabilities is hardly changed at 5.5pp, compared to 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 46.8pp higher (size weighted) premium in its core industry compared to a 7.2pp 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 1pp. 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.

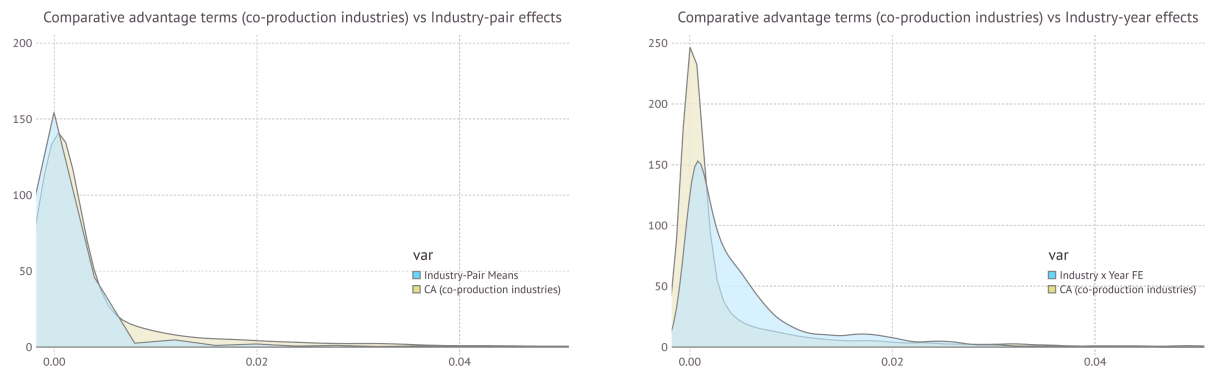
TABLE 13. Core Competency Sales Premium from Input-Based Comparative Advantage – Size-Weighted

Industry rank	# of Industries With Positive Sales (CA weighted by size)									
	1	2	3	4	5	6	7	8	9	10+
1	0.055	0.072	0.130	0.157	0.143	0.179	0.178	0.284	0.468	1.727
2		0.005	0.012	0.039	0.158	0.301	0.266	0.332	0.018	3.499
3			0.002	0.005	0.007	0.048	0.019	0.041	0.245	1.375
4				0.001	0.007	0.057	0.017	0.024	0.019	0.185
5					0.004	0.009	0.014	0.008	0.019	0.047
6						0.004	0.007	0.008	0.006	0.011
7							0.002	0.006	0.005	0.019
8								0.002	0.001	0.006
9									0.005	0.004
10+										0.002

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.

5.3. **Economic Significance.** To examine the economic importance of input-based comparative advantage, we compare the distribution of industry-level variation in co-production, which are 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 sub-section decomposes the margins of co-production within firms. It relates them to potential underlying drivers, such as industry-level demand and supply shocks, which are being explicitly modelled in the theory and are being picked up in the form of fixed effects in the empirical analysis of Sections 3 and 5.

The left panel of Figure 5.1 compares the distribution of CA, in yellow, to the distribution of (unconditional) co-production rates within industry-pair cells, in blue. Comparative advantage is estimated as in Equation (5.1) while the unconditional co-production rates capture, for example, demand complementarities such as left shoes being co-produced with right shoes. The distributions overlap substantially, showing that input-based comparative advantage shifts entry probabilities to an extent that is similar to time-invariant co-production rates within industry pairs. In other words, CA is as important in driving entry as summary statistics of observed co-production



(A) CA and Industry-Pair Entry Rates (B) CA and Industry-Year Entry Rates

FIGURE 5.1. Comparing Input-based Comparative Advantage with Other Drivers of Entry

rates across industries. Further, input-based comparative advantage explains much of the variation in entry probabilities across industry-pairs, which reduces the role that confounders could play in altering the contribution of CA towards entry into an industry.

The right panel compares the distribution of CA, in yellow, to the distribution of entry probabilities within industry-year cells (i.e. what an industry-year 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 in between the importance of industry-pair-level drivers and industry-year shocks.

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

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

$$\ln R_{jkt} = \kappa_{kt} + \kappa_{jk} + \sum_i \left[\kappa_0^I + \kappa_1^I \left(\alpha_B B_{ijt} + \alpha_\tau \tau_{ijt}\right)\right] \left(\theta_{ijt}\overline{\theta}_{ik} - \frac{\theta_{ijt}^2}{2}\right) + \sum_f \kappa_1^F \left(\theta_{fjt}\overline{\theta}_{fk} - \frac{\theta_{fjt}^2}{2}\right). \tag{5.2}$$

with $\kappa_0^I = \Theta_{i0}^2\rho/(1-\rho)$, $\kappa_1^I = 2\Theta_{i0}\rho/(1-\rho)(\sigma-1)$, $\kappa_1^F = 2\Theta_{f0}\rho/(1-\rho)(\sigma-1)$. 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 the Appendix.

Taking the extended model to data, Table 14 shows structural estimates accounting for primary factor - capital K and labour 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 co-production in high capital intensity industries. Tables 21 and 22 in the Appendix provide reduced form results with primary factors.

TABLE 14. Structural Estimates for Multi-Industry Sales Premium, with Primary Factors

		Positive Sales Dummy ($R_{jkt} > 0$)			
		(1)	(2)	(3)	(4)
Input Distance	$\sum_{i\notin\{K,L\}}\left(\theta_{ijt}\overline{\theta}_{ik} - \theta_{ijt}^2/2\right)$	0.0068***	0.0068***	0.1277***	0.1112***
		(0.0002)	(0.0002)	(0.0127)	(0.0110)
Factor Distance	$\sum_{f\in\{K,L\}}\left(\theta_{fjt}\overline{\theta}_{fk} - \theta_{fjt}^2/2\right)$	0.0084***	0.0084***	0.1094***	0.0600***
		(0.0003)	(0.0003)	(0.0122)	(0.0080)
Input Distance	$\sum_i B_{ijt}\cdot\left(\theta_{ijt}\overline{\theta}_{ik} - \theta_{ijt}^2/2\right)$	-0.0000	-0.0000	-0.0040***	-0.0036***
Entry Barriers		(0.0002)	(0.0002)	(0.0006)	(0.0006)
Input Distance	$\sum_i \tau_{ijt}\cdot\left(\theta_{ijt}\overline{\theta}_{ik} - \theta_{ijt}^2/2\right)$		0.0008		-0.0210***
Tariffs			(0.0005)		(0.0027)
κ_{jk}		Yes	Yes	Yes	Yes
κ_{kt}		Yes	Yes	Yes	Yes
Estimator		OLS	OLS	IV	IV
N		77,745,382	77,745,382	46,185,150	46,185,150
R^2		0.7622	0.7622	0.7598	0.7598

Standard errors in parentheses, clustered at the firm-industry level.
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

6. CONCLUSION

Even though multiproduct firms account for a disproportionately large share of economic activity, systematic theory and evidence examining product diversification is 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 formalises the resource-based view of the firm in ways that enables an assessment of its economic significance. We bring this theory to Indian manufacturing data to study the relevance of input capabilities in both 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 they are quantitatively as important as time-invariant co-production rates across industries.

A key theoretical insight of our framework is that economies of scope within multiproduct firms imply production choices and input capabilities are jointly determined. Production choices are interdependent on the relative demands 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 instrumental variable strategy which 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 single product 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.

References

ACEMOGLU, D., P. ANTRÀS, AND E. HELPMAN (2007): "Contracts and Technology Adoption," *American Economic Review*, 97, 916–943.

ACEMOGLU, D., V. M. CARVALHO, A. OZDAGLAR, AND A. TAHBAZ-SALEHI (2012): "The network origins of aggregate fluctuations," *Econometrica*, 80, 1977–2016.

AMIHUD, Y. AND B. LEV (1981): "Risk reduction as a managerial motive for conglomerate mergers," *The Bell Journal of Economics*, 605–617.

AMIRAPU, A., M. GECHTER, AND G. SMAGGHUE (2018): "Dynamic Effects of Product Market Regulation: Evidence from India," *Working Paper*.

AMITI, M. AND J. KONINGS (2007): "Trade liberalization, intermediate inputs, and productivity: Evidence from Indonesia," *The American Economic Review*, 1611–1638.

ANGRIST, J. D. AND J.-S. PISCHKE (2008): *Mostly harmless econometrics: An empiricist's companion*, Princeton University Press.

ANTRAS, P. AND D. CHOR (2013): "Organizing the global value chain," *Econometrica*, 81, 2127–2204.

ATALAY, E., A. H. SU, M. J. LI, AND C. SYVERSON (2019): "How Wide is the Firm Border?" *The Quarterly Journal of Economics*, 134, 1845–1882.

AW, B. Y. AND Y. LEE (2009): "Product Choice and Market Competition: The Case of Multiproduct Electronic Plants in Taiwan," *The Scandinavian Journal of Economics*, 111, 711–740.

BAUMOL, W. J. (1977): "On the proper cost tests for natural monopoly in a multiproduct industry," *The American Economic Review*, 67, 809–822.

BERNARD, A. B., E. J. BLANCHARD, I. VAN BEVEREN, AND H. VANDENBUSSCHE (2018): "Carry-Along Trade," *The Review of Economic Studies*, 0, 1–38.

BERNARD, A. B., E. DHYNE, G. MAGERMAN, K. MANOVA, AND A. MOXNES (2021): "The origins of firm heterogeneity: A production network approach," *forthcoming Journal of Political Economy*.

BERNARD, A. B., J. EATON, J. B. JENSEN, AND S. KORTUM (2003): "Plants and Productivity in International Trade," *The American Economic Review*, 93, 1268–1290.

BERNARD, A. B., J. B. JENSEN, S. J. REDDING, AND P. K. SCHOTT (2007): "Firms in International Trade," *The Journal of Economic Perspectives*, 21, 105–130.

BERNARD, A. B., S. J. REDDING, AND P. K. SCHOTT (2010): "Multiple-Product Firms and Product Switching," *The American Economic Review*, 100, 70–97.

——— (2011): "Multiproduct Firms and Trade Liberalization," *The Quarterly Journal of Economics*, 126, 1271–1318.

BERNHEIM, B. D. AND M. D. WHINSTON (1990): "Multimarket contact and collusive behavior," *The RAND Journal of Economics*, 1–26.

BLOOM, N., A. MAHAJAN, D. MCKENZIE, AND J. ROBERTS (2010): "Why Do Firms in Developing Countries Have Low Productivity?" *The American Economic Review*, 619–623.

BLOOM, N., M. SCHANKERMAN, AND J. VAN REENEN (2013): "Identifying technology spillovers and product market rivalry," *Econometrica*, 81, 1347–1393.

BOEHM, J., S. DHINGRA, AND J. MORROW (2016): "Swimming upstream: input-output linkages and the direction of product adoption," *CEP discussion paper*, 1407.

——— (2018): "Product Diversification in Indian Manufacturing," *Developments in Global Sourcing*, 337.

BOLLARD, A., P. J. KLENOW, AND G. SHARMA (2013): "India's Mysterious Manufacturing Miracle," *Review of Economic Dynamics*, 16, 59–85.

BOWEN, H. P. AND M. F. WIERSEMA (2005): "Foreign-based Competition and Corporate Diversification Strategy," *Strategic Management Journal*, 26, 1153–1171.

BOWN, C. P. AND M. A. CROWLEY (2016): "The Empirical Landscape of Trade Policy," in *Handbook of Commercial Policy*, Elsevier, vol. 1, 3–108.

BRANDER, J. A. AND J. EATON (1984): "Product Line Rivalry," *The American Economic Review*, 74, 323–334.

BRODA, C. AND D. E. WEINSTEIN (2006): "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 121, 541–585.

BRYCE, D. J. AND S. G. WINTER (2009): "A General Interindustry Relatedness Index," *Management Science*, 55, 1570–1585.

CAMERON, A. C. AND D. L. MILLER (2015): "A practitioner's guide to cluster-robust inference," *Journal of Human Resources*, 50, 317–372.

CHANDLER, A. D. (1992): "Organizational capabilities and the economic history of the industrial enterprise," *Journal of Economic Perspectives*, 6, 79–100.

COMBES, P.-P. AND L. GOBILLON (2015): "The Empirics of Agglomeration Economies," in *Handbook of Regional and Urban Economics*, Elsevier, vol. 5, 247–348.

CONLEY, T. G. AND B. DUPOR (2003): "A spatial analysis of sectoral complementarity," *Journal of Political Economy*, 111, 311–352.

CROZET, M. AND F. TRIONFETTI (2013): "Firm-level comparative advantage," *Journal of International Economics*, 91, 321–328.

DE BLAS, B. AND K. N. RUSS (2015): "Understanding Markups in the Open Economy," *American Economic Journal: Macroeconomics*, 7, 157–180.

DHINGRA, S. (2013): "Trading away wide brands for cheap brands," *The American Economic Review*, 103, 2554–2584.

DHINGRA, S. AND J. MORROW (2019): "Monopolistic competition and optimum product diversity under firm heterogeneity," *Journal of Political Economy*, 127, 196–232.

DI GIOVANNI, J., A. A. LEVCHENKO, AND I. MÉJEAN (2014): "Firms, destinations, and aggregate fluctuations," *Econometrica*, 82, 1303–1340.

DING, X. (2019): "Intangible Economies of Scope: Micro Evidence and Macro Implications," *Working Paper*.

DURLAUF, S. N. (1993): "Nonergodic economic growth," *The Review of Economic Studies*, 60, 349–366.

EATON, B. C. AND N. SCHMITT (1994): "Flexible manufacturing and market structure," *The American Economic Review*, 875–888.

ECKEL, C., L. IACOVONE, B. JAVORCIK, AND J. P. NEARY (2015): "Multi-product firms at home and away: Cost-versus quality-based competence," *Journal of International Economics*, 95, 216–232.

ECKEL, C. AND P. NEARY (2010): "Multi-product Firms and Flexible Manufacturing in the Global Economy," *Review of Economic Studies*, 77, 188–217.

ETHIER, W. (1982): "National and international returns to scale in the modern theory of international trade," *The American Economic Review*, 72, 389–405.

FAN, J. P. AND L. H. LANG (2000): "The Measurement of Relatedness: An Application to Corporate Diversification," *The Journal of Business*, 73, 629–660.

FLAGGE, M. AND R. CHAUREY (2014): "Firm-Product Linkages and the Evolution of Product Scope," *Working Paper*.

FONTAGNE, L., A. SECCHI, AND C. TOMASI (2018): "Exporters' product vectors across markets," *European Economic Review*, 110, 150–180.

FOSTER, L., J. HALTIWANGER, AND C. SYVERSON (2008): "Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?" *American Economic Review*, 98, 394–425.

GALLE, S. (2015): "Competition, Financial Constraints and Misallocation: Plant-Level Evidence from Indian Manufacturing," *UC Berkeley Working Paper*.

GARCIA-SANTANA, M. AND J. PIJOAN-MAS (2014): "The reservation laws in India and the misallocation of production factors," *Journal of Monetary Economics*, 66, 193–209.

GOLDBERG, P. K., A. K. KHANDELWAL, N. PAVCNIK, AND P. TOPALOVA (2009): "Multi-product Firms and Product Turnover in the Developing World: Evidence from India," *The Review of*

Economics and Statistics.

——— (2010): "Imported intermediate inputs and domestic product growth: Evidence from India," *The Quarterly Journal of Economics*, 125, 1727–1767.

GOVERNMENT OF INDIA (2009): "Micro, Small, and Medium Enterprises in India: An Overview," Tech. rep.

HALPERN, L., M. KOREN, AND A. SZEIDL (2015): "Imported inputs and productivity," *American Economic Review*, 105, 3660–3703.

HARRISON, A. AND A. RODRIGUEZ-CLARE (2010): "Trade, foreign investment, and industrial policy for developing countries," *Handbook of Development Economics*, 5, 4039–4214, publisher: Elsevier.

HAUSMANN, R. AND C. A. HIDALGO (2011): "The network structure of economic output," *Journal of Economic Growth*, 16, 309–342.

HAUSMANN, R., J. HWANG, AND D. RODRIK (2007): "What you export matters," *Journal of Economic Growth*, 12, 1–25.

HIDALGO, C. A., B. KLINGER, A. L. BARABASI, AND R. HAUSMANN (2007): "The product space conditions the development of nations," *Science*, 317, 482–487.

HOBERG, G. AND G. PHILLIPS (2016): "Text-based network industries and endogenous product differentiation," *Journal of Political Economy*, 124, 1423–1465.

HOTTMAN, C. J., S. J. REDDING, AND D. E. WEINSTEIN (2016): "Quantifying the sources of firm heterogeneity," *The Quarterly Journal of Economics*, 131, 1291–1364.

HUSSAIN, A. (1997): *Report of the expert committee on small enterprises*, Ministry of Industry, Government of India, by National Council of Applied Economic Research.

HUTCHINSON, J., J. KONINGS, AND P. P. WALSH (2010): "The firm size distribution and inter-industry diversification," *Review of Industrial Organization*, 37, 65–82.

IACOVONE, L. AND B. S. JAVORCIK (2010): "Multi-Product Exporters: Product Churning, Uncertainty and Export Discoveries," *The Economic Journal*, 120, 481–499.

JENSEN, M. C. (1986): "Agency costs of free cash flow, corporate finance, and takeovers," *The American Economic Review*, 76, 323–329.

JOVANOVIC, B. (1987): "Micro shocks and aggregate risk," *The Quarterly Journal of Economics*, 395–409.

JOVANOVIC, B. AND R. J. GILBERT (1993): "The diversification of production," *Brookings papers on economic activity. Microeconomics*, 1993, 197–247.

KASAHARA, H. AND J. RODRIGUE (2008): "Does the use of imported intermediates increase productivity? Plant-level evidence," *Journal of Development Economics*, 87, 106–118.

KOREN, M. AND S. TENREYRO (2013): "Technological diversification," *The American Economic Review*, 103, 378–414.

KUGLER, M. AND E. VERHOOGEN (2009): "Plants and Imported Inputs: New Facts and an Interpretation," *American Economic Review*, 99, 501–507.

——— (2012): "Prices, Plant Size, and Product Quality," *The Review of Economic Studies*, 79, 307–339.

LICHTENBERG, F. R. (1992): "Industrial de-diversification and its consequences for productivity," *Journal of Economic Behavior & Organization*, 18, 427–438.

LIU, R. (2010): "Import competition and firm refocusing," *Canadian Journal of Economics/Revue canadienne d'économique*, 43, 440–466.

LOPEZ, R. A. AND J. SUDEKUM (2009): "Vertical Industry Relations, Spillovers, and Productivity: Evidence from Chilean Plants," *Journal of Regional Science*, 49, 721–747.

LU, D., A. MARISCAL, AND L.-F. MEJIA (2016): "How Firms Accumulate Inputs: Evidence from Import Switching," *Working Paper*.

MAKSIMOVIC, V. AND G. M. PHILLIPS (2013): "Conglomerate firms, internal capital markets, and the theory of the firm," *Annu. Rev. Financ. Econ.*, 5, 225–244.

MARTIN, L. A., S. NATARAJ, AND A. E. HARRISON (2017): "In with the big, out with the small: Removing small-scale reservations in India," *American Economic Review*, 107, 354–86.

MAYER, T., M. J. MELITZ, AND G. I. P. OTTAVIANO (2014): "Market size, competition, and the product mix of exporters," *American Economic Review*, 104, 495–536.

——— (2020): "Product Mix and Firm Productivity Responses to Trade Competition," *Review of Economics and Statistics*, 59.

MONTGOMERY, C. A. (1994): "Corporate diversificaton," *Journal of Economic Perspectives*, 8, 163–178.

MORCK, R., A. SHLEIFER, AND R. W. VISHNY (1990): "Do managerial objectives drive bad acquisitions?" *The Journal of Finance*, 45, 31–48.

PANZAR, J. C. AND R. D. WILLIG (1981): "Economies of scope," *The American Economic Review*, 71, 268–272.

PENROSE, E. (1955): "Limits to the Growth and Size of Firms," *The American Economic Review*, 531–543.

RACHAPALLI, S. (2021): "Learning between Buyers and Sellers Along the Global Value Chain," *Working Paper*.

ROBINS, J. AND M. F. WIERSEMA (1995): "A Resource-based Approach to the Multibusiness Firm: Empirical Analysis of Portfolio Interrelationships and Corporate Financial Performance," *Strategic Management Journal*, 16, 277–299.

RONDI, L. AND D. VANNONI (2005): "Are EU leading firms returning to core business? Evidence on refocusing and relatedness in a period of market integration," *Review of Industrial Organization*, 27, 125–145.

SCHERER, F. M. (1982): "Inter-Industry Technology Flows and Productivity Growth," *The Review of Economics and Statistics*, 64, pp. 627–634.

SCHOTT, P. K. (2004): "Across-product versus within-product specialization in international trade," *The Quarterly Journal of Economics*, 119, 647–678.

SHAKED, A. AND J. SUTTON (1990): "Multiproduct firms and market structure," *The RAND Journal of Economics*, 45–62.

STIGLER, G. J. (1951): "The Division of Labor is Limited by the Extent of the Market," *Journal of Political Economy*, 59, 185–193.

SUTTON, J. AND D. TREFLER (2016): "Capabilities, wealth, and trade," *Journal of Political Economy*, 124, 826–878.

TEECE, D. J. (1982): "Towards an Economic Theory of the Multiproduct Firm," *Journal of Economic Behavior & Organization*, 3, 39–63.

TYBOUT, J. R. (2000): "Manufacturing firms in developing countries: How well do they do, and why?" *Journal of Economic literature*, 11–44.

VANDENBUSSCHE, H. AND C. VIEGELAHN (2018): "Input reallocation within multi-product firms," *Journal of International Economics*, 114, 63–79.

WILLIG, R. D., S. C. SALOP, AND F. M. SCHERER (1991): "Merger analysis, industrial organization theory, and merger guidelines," *Brookings Papers on Economic Activity. Microeconomics*, 1991, 281–332.

Appendix A. Descriptive Statistics

Table 15. Summary Statistics

	count	mean	sd	min	max
Industry Add Dummy	52,691,029	0.0011	0.03	0.00	1
InputSimilarity$^0_{jk}$	52,691,029	0.0110	0.05	0.00	1
InputSimilarity-Dereservation$^0_{jkt}$	52,691,029	0.0005	0.01	0.00	1
InputSimilarity-Tariff$^0_{jkt}$	52,691,029	-0.0001	0.00	-0.33	0
OutputSimilarity$^0_{jk}$	52,691,029	0.0038	0.05	0.00	1
OutputSimilarity-Dereservation$^0_{jkt}$	52,691,029	0.0005	0.02	0.00	1
Upstream$^0_{jk}$	52,691,029	0.0035	0.03	0.00	1
Downstream$^0_{jk}$	52,691,029	0.0058	0.04	0.00	1

Table 16. Correlation Matrix of Similarity Indices

	IS$^0_{jk}$	OS$^0_{jk}$	IS-DR$^0_{jkt}$	OS-DR$^0_{jkt}$	Up$^0_{jk}$	Down$^0_{jk}$
InputSimilarity$^0_{jk}$	1.00					
OutputSimilarity$^0_{jk}$	0.37	1.00				
InputSimilarity-Dereservation$^0_{jkt}$	0.17	0.06	1.00			
OutputSimilarity-Dereservation$^0_{jkt}$	0.09	0.37	0.10	1.00		
Upstream$^0_{jk}$	0.46	0.54	0.05	0.10	1.00	
Downstream$^0_{jk}$	0.54	0.45	0.06	0.08	0.50	1.00

TABLE 17. Structural Summary Statistics

	count	mean	sd	min	max
Indicator $R_{jkt} > 0$ (OLS)	77,745,382	0.0059	0.08	0.00	1
$\sum_i \left(\overline{\theta}_{ik}\theta_{ijt} - \theta^2_{ijt}/2\right)$	77,745,382	-0.3585	0.15	-0.50	1
$\sum_i B_{it} \cdot \left(\overline{\theta}_{ik}\theta_{ijt} - \theta^2_{ijt}/2\right)$	77,745,382	-0.0142	0.07	-0.50	0
$\sum_i \tau_{it} \cdot \left(\overline{\theta}_{ik}\theta_{ijt} - \theta^2_{ijt}/2\right)$	77,745,382	-0.0067	0.02	-1.05	0
Indicator $R_{jkt} > 0$ (IV)	46,185,150	0.0060	0.08	0.00	1
$\sum_i \left(\overline{\theta}_{ik}\theta_{ijt} - \theta^2_{ijt}/2\right)$	46,185,150	-0.3477	0.15	-0.50	1
$\sum_i B_{it} \cdot \left(\overline{\theta}_{ik}\theta_{ijt} - \theta^2_{ijt}/2\right)$	46,185,150	-0.0128	0.06	-0.50	0
$\sum_i \tau_{it} \cdot \left(\overline{\theta}_{ik}\theta_{ijt} - \theta^2_{ijt}/2\right)$	46,185,150	-0.0076	0.03	-0.83	0
$\sum_i \left(\overline{\theta}_{ik}\theta_{ijt-1} - \theta^2_{ijt-1}/2\right)$	46,185,150	-0.3374	0.17	-1.46	1
$\sum_i B_{it} \cdot \left(\overline{\theta}_{ik}\theta_{ijt-1} - \theta^2_{ijt-1}/2\right)$	46,185,150	-0.0130	0.07	-1.00	0
$\sum_i \tau_{it} \cdot \left(\overline{\theta}_{ik}\theta_{ijt-1} - \theta^2_{ijt-1}/2\right)$	46,185,150	-0.0061	0.02	-0.91	0
$\sum_i \chi_{jkt-1} \left(\overline{\theta}_{ik} - \theta_{ijt-1}\right)^2$	46,185,150	0.0000	0.00	0.00	1
$\sum_i B_{it} \cdot \chi_{jkt-1} \left(\overline{\theta}_{ik} - \theta_{ijt-1}\right)^2$	46,185,150	0.0011	0.02	0.00	3
$\sum_i \tau_{it} \cdot \chi_{jkt-1} \left(\overline{\theta}_{ik} - \theta_{ijt-1}\right)^2$	46,185,150	0.0001	0.01	0.00	1

APPENDIX B. ROBUSTNESS OF ESTIMATES AND FURTHER RESULTS

B.1. **Robustness of Industry Add, Drop and Sales Regressions.** Table 18 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 which never have any co-production 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). Tables 19 shows how the probability to drop an industry from the industry mix is shaped by input similarity. Table 20 shows how log sales are correlated with input similarity with a wide range of input-output linkage controls.

TABLE 18. Revealed Comparative Advantage – Robustness

	Dependent variable: Add_{jkt}			
	(1)	(2)	(3)	(4)
$\text{InputSimilarity}^0_{jk}$	0.0199**	0.0196**	0.0251**	0.0360**
	(0.00057)	(0.00062)	(0.00087)	(0.0011)
$\text{InputSimilarity-Dereservation}^0_{jkt}$	0.0155**	0.0146**	0.0110**	0.0214**
	(0.0023)	(0.0027)	(0.0028)	(0.0042)
$\text{OutputSimilarity}^0_{jk}$	0.100**	0.0970**	0.119**	0.0858**
	(0.0018)	(0.0020)	(0.0025)	(0.0019)
$\text{OutputSimilarity-Dereservation}^0_{jkt}$	0.0171**	0.0178**	0.0179**	0.0147**
	(0.0022)	(0.0025)	(0.0030)	(0.0021)
Upstream^0_{jk}	0.0291**	0.0219**	0.0358**	0.0186**
	(0.0030)	(0.0032)	(0.0041)	(0.0035)
Downstream^0_{jk}	-0.00351*	-0.00443**	-0.00384*	-0.0148**
	(0.0014)	(0.0016)	(0.0020)	(0.0029)
Sample	Full	Single-plant firms	Census plants	Co-production industries
Firm × Year FE	Yes	Yes	Yes	Yes
$k \times k' \times t$ FE	Yes	Yes	Yes	Yes
R^2	0.0646	0.0660	0.0848	0.0965
Observations	52666907	43120945	27076486	5165511

Standard errors in parentheses, clustered at the firm-industry level.
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

TABLE 19. Industry Drop Regressions

	Dependent variable: Drop_{jkt}		
	(1)	(2)	(3)
$\text{InputSimilarity}^0_{jkt}$	0.0128^{+}	-0.195^{**}	-0.139^{**}
	(0.0078)	(0.013)	(0.014)
$\text{InputSimilarityKL}^0_{jk}$	-0.780^{**}	-0.961^{**}	-0.663^{**}
	(0.100)	(0.15)	(0.18)
$\text{InputSimilarity-Dereservation}^0_{jkt}$	-0.400^{**}	-0.133^{+}	-0.144^{+}
	(0.066)	(0.070)	(0.087)
$\text{OutputSimilarity}^0_{jkt}$	-0.249^{**}	-0.263^{**}	-0.213^{**}
	(0.0040)	(0.0049)	(0.0074)
$\text{OutputSimilarity-Dereservation}^0_{jkt}$	-0.131^{**}	-0.101^{**}	-0.0721^{**}
	(0.0095)	(0.011)	(0.014)
Firm $\times$ Year FE α_{jt}	Yes	Yes	Yes
Industry $\times$ Year FE α_{kt}		Yes	
$k \times k' \times t$ FE $\alpha_{kk't}$			Yes
R^2	0.536	0.572	0.656
Observations	159001	158920	134861

Standard errors in parentheses, clustered at the firm-industry level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

TABLE 20. Intensive Margin of Sales

	Dependent variable: $\log \text{Sales}_{jkt}$		
	(1)	(2)	(3)
$\text{InputSimilarity}^0_{jkt}$	0.526**	1.043**	0.508**
	(0.036)	(0.052)	(0.051)
$\text{InputSimilarityKL}^0_{jk}$	1.851**	4.891**	4.916**
	(0.51)	(0.73)	(0.72)
$\text{InputSimilarity-Dereservation}^0_{jkt}$	1.250**	0.672**	0.866**
	(0.26)	(0.25)	(0.24)
$\text{OutputSimilarity}^0_{jkt}$	4.110**	3.475**	1.507**
	(0.019)	(0.022)	(0.024)
$\text{OutputSimilarity-Dereservation}^0_{jkt}$	-0.381**	-0.548**	-0.262**
	(0.039)	(0.044)	(0.041)
Firm × Year FE α_{jt}	Yes	Yes	Yes
Industry × Year FE α_{kt}		Yes	
$k \times k' \times t$ FE $\alpha_{kk't}$			Yes
R^2	0.802	0.832	0.911
Observations	251026	250963	220613

Standard errors in parentheses, clustered at the firm-industry level.
$^+ p < 0.10$, $^* p < 0.05$, $^{**} p < 0.01$

B.2. **Robustness of Industry Add Regressions to the Inclusion of Factors.** Tables 21 and 22 include analogous measures of factor similarity (capital and labor) in the reduced form, showing similar results.

TABLE 21. Industry Entry Correlations, with Primary Factors

	Dependent variable: Add_{jkt}					
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{InputSimilarity}^0_{jk}$	0.0543**	0.0543**	0.0530**	0.0529**	0.0377**	0.0376**
	(0.00055)	(0.00055)	(0.00055)	(0.00055)	(0.00089)	(0.00089)
$\text{InputSimilarityKL}^0_{jk}$		0.00322**		0.00767**		0.00449**
		(0.00020)		(0.00026)		(0.00030)
Constant	0.000815**	0.000713**	0.000824**	0.000582**	0.000931**	0.000789**
	(0.0000052)	(0.0000077)	(0.0000052)	(0.0000093)	(0.0000075)	(0.000012)
Firm × Year FE α_{jt}	Yes	Yes	Yes	Yes	Yes	Yes
Industry × Year FE α_{kt}			Yes	Yes		
$k \times k' \times t$ FE $\alpha_{kk't}$					Yes	Yes
R^2	0.00879	0.00880	0.0109	0.0110	0.0572	0.0572
Observations	46387505	46387505	46387505	46387505	46360567	46360567

Standard errors in parentheses, clustered at the firm-industry level.
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

TABLE 22. Industry Entry, with Primary Factors

	Dependent variable: Add_{jkt}			
	(1)	(2)	(3)	(4)
$\text{InputSimilarity}^0_{jk}$	0.0525**	0.0511**	0.0365**	0.0357**
	(0.00055)	(0.00055)	(0.00089)	(0.00089)
$\text{InputSimilarityKL}^0_{jk}$	0.00324**	0.00768**	0.00449**	0.00449**
	(0.00020)	(0.00026)	(0.00030)	(0.00030)
$\text{InputSimilarity-Dereservation}^0_{jkt}$	0.0713**	0.0705**	0.0339**	0.0325**
	(0.0043)	(0.0042)	(0.0040)	(0.0040)
$\text{InputSimilarity-Tariff}^0_{jkt}$				-0.0997**
				(0.014)
Constant	0.000704**	0.000574**	0.000787**	0.000787**
	(0.0000077)	(0.0000093)	(0.000012)	(0.000012)
Firm × Year FE α_{jt}	Yes	Yes	Yes	Yes
Industry × Year FE α_{kt}		Yes		
$k \times k' \times t$ FE $\alpha_{kk't}$			Yes	Yes
R^2	0.00893	0.0111	0.0573	0.0573
Observations	46387505	46387505	46360567	46360567

Standard errors in parentheses, clustered at the firm-industry level.
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

B.3. **Robustness of Industry Add Regressions to Logit.** Table 23 shows the results of the logit estimation of the industry addition regressions, corresponding to the baseline specifications of Table 4.

TABLE 23. Revealed Comparative Advantage – Robustness

	Dependent variable: Add_{jkt}					
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{InputSimilarity}^0_{jk}$	6.665***	5.209***	6.585***	8.204***	5.136***	5.095***
	(0.031)	(0.070)	(0.032)	(0.047)	(0.071)	(0.072)
$\text{InputSimilarity-Dereservation}^0_{jkt}$			2.134***	2.448***	1.455***	1.388***
			(0.172)	(0.198)	(0.247)	(0.247)
$\text{InputSimilarity-Tariff}^0_{jkt}$						-3.731**
						(0.899)
Plant $\times$ Year FE α_{jt}	Yes	Yes	Yes	Yes	Yes	Yes
Industry $\times$Year FE α_{kt}				Yes		
$k \times k' \times t$ FE		Yes			Yes	Yes
Estimator	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

Standard errors in parentheses, clustered at the firm-industry level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B.4. **Robustness of Unit Value Regressions.** Table 24 shows results of a regression of log unit values of domestically sourced intermediate inputs (by 5-digit input category i) on a dummy that is one 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 de-reservation 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 onwards, 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 24 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).

TABLE 24. Domestic Input Unit Values After Dereservation – Robustness

	Dependent variable: log p_{jit}			
	(1)	(2)	(3)	(4)
$t \geq$ year i was de-reserved	-0.130**	-0.0864**	-0.0477**	-0.0635**
	(0.014)	(0.015)	(0.012)	(0.014)
	All	All	Safe	Safe
Year FE	Yes	Yes	Yes	Yes
Input Product FE	Yes		Yes	
Firm $\times$ Input Product FE		Yes		Yes
R^2	0.850	0.955	0.880	0.966
Observations	957056	547866	789791	453948

Standard errors in parentheses, clustered at the firm-year level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

B.5. **Diversification Discount - Robustness.** The dependent variable in the following Tables is log average sales of the firm in the industry it entered over a three-year and five-year horizon, where the average is taken across all years where we observe the firm. Table 25 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 are associated with higher sales performance. Table 26 also includes industry-year fixed effects and finds similar results. Finally, Table 27 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 post-entry performance.

TABLE 25. Post-entry Growth: Within Continuing Industries

	Dep. var.: log $\left(\text{Avg. Sales}_{jk}^{t,\ldots,t+3}\right)$			Dep. var.: log $\left(\text{Avg. Sales}_{jk}^{t,\ldots,t+5}\right)$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{InputSimilarity}_{jk}^{0}$	0.971**	0.896**	0.761**	1.160**	1.084**	0.949**
	(0.082)	(0.083)	(0.084)	(0.083)	(0.084)	(0.085)
$\text{InputSimilarity-Dereservation}_{jkt}^{0}$		2.199**	1.831**		2.225**	1.856**
		(0.33)	(0.33)		(0.33)	(0.33)
$\text{InputSimilarity-Tariff}_{jkt}^{0}$			-20.08**			-20.13**
			(1.34)			(1.36)
Firm Industry $\times$ Year FE $\alpha_{k't}$	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.159	0.159	0.164	0.166	0.167	0.172
Observations	55296	55296	55296	55296	55296	55296

Standard errors in parentheses, clustered at the firm-industry level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$
Notes: 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.

TABLE 26. Post-entry Growth: Within Continuing and Entering Industries

	Dep. var.: log $\left(\text{Avg. Sales}_{jk}^{t,\ldots,t+3}\right)$			Dep. var.: log $\left(\text{Avg. Sales}_{jk}^{t,\ldots,t+5}\right)$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{InputSimilarity}_{jk}^{0}$	1.292**	1.230**	1.077**	1.485**	1.423**	1.268**
	(0.088)	(0.088)	(0.089)	(0.089)	(0.090)	(0.091)
$\text{InputSimilarity-Dereservation}_{jkt}^{0}$		1.538**	1.231**		1.557**	1.249**
		(0.34)	(0.33)		(0.34)	(0.33)
$\text{InputSimilarity-Tariff}_{jkt}^{0}$			-19.66**			-19.70**
			(1.27)			(1.28)
Industry $k\times$ Year FE α_{kt}	Yes	Yes	Yes	Yes	Yes	Yes
Firm Industry $k'\times$ Year FE $\alpha_{k't}$	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.259	0.260	0.264	0.266	0.267	0.271
Observations	55163	55163	55163	55163	55163	55163

Standard errors in parentheses, clustered at the firm-industry level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$
Notes: 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.

TABLE 27. Post-entry Growth: Within Industry-Pairs

	Dep. var.: log $\left(\text{Avg. Sales}_{jk}^{t,\dots,t+3}\right)$			Dep. var.: log $\left(\text{Avg. Sales}_{jk}^{t,\dots,t+5}\right)$		
	(1)	(2)	(3)	(4)	(5)	(6)
InputSimilarity$^0_{jk}$	0.219^{+}	0.155	0.0125	0.334^{**}	0.273^{*}	0.132
	(0.12)	(0.12)	(0.13)	(0.13)	(0.13)	(0.13)
InputSimilarity-Dereservation$^0_{jkt}$		1.600^{**}	1.439^{**}		1.533^{**}	1.372^{**}
		(0.39)	(0.39)		(0.39)	(0.39)
InputSimilarity-Tariff$^0_{jkt}$			-17.46^{**}			-17.37^{**}
			(1.48)			(1.48)
$k \times k' \times t$ FE $\alpha_{kk't}$	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.445	0.445	0.448	0.450	0.450	0.453
Observations	39695	39695	39695	39695	39695	39695

Standard errors in parentheses, clustered at the firm-industry level.
$^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Notes: 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.

APPENDIX C. DATA APPENDIX

C.1. **Data sources.**

C.1.1. *Manufacturing plant data:* Our manufacturing plant data is the "detailed unit level data with factory identifier" of the Indian *Annual Survey of Industries* (ASI), years 2000/01 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`.

C.1.2. *Tariff data:* The Indian import tariff data comes from UNCTAD-TRAINS (accessed 05/14/2016 through WITS: `http://wits.worldbank.org/`).

C.1.3. *Dereservation data:* Notices of dereservation of products from the website of the Development Commissioner, Ministry of Micro, Small, and Medium Enterprises.[36] We manually concord the product codes to 5-digit ASIC codes based on the text description of the dereserved items.

C.2. **Variable definitions.**

- *Add dummies* Add$_{jkt}$: one 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.

[36] `http://www.dcmsme.gov.in/publications/reserveditems/resvex.htm` (accessed December 2014)

- *Drop dummies* Drop_{jkt} : one 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*$_{jkt}$: j's total sales of products in 3-digit industry k at time t, including exports.
- *Plant expenditure shares* θ_{ijt} : 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* σ^t_{kj}, χ_{jkt} : 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 dummy* δ_{ijt} and B_{ijt}: one if and only if there is a 5-digit input in the 3-digit basket i that has been dereserved during or prior to t and shows up at some point in j's basket of intermediate inputs. In Section 4, the reservation dummy B_{ijt} is one 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* $\Delta\tau_{ijt}$: 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 codes reported by TRAINS to ASIC codes via the the ASIC 2009/10 – NPCMS concordance published by MO-SPI, and the CPC–HS concordance published by UNSTATS (the first five digits of NPCMS are CPC v2.0 codes). Tariffs are effective applied tariffs where available, and MFN tariffs

otherwise. We focus on non-agricultural tariffs to avoid endogeneity concerns with agricultural tariffs, which often vary due to policy responses to domestic economic conditions that can affect firm sales directly. In Section 4, τ_{ijt} is defined analogously as the level of that tariff.

- *Input Similarity* $\text{InputSimilarity}^t_{jk}$ (where N is the number of 3-digit industries):

$$\text{InputSimilarity}^t_{jk} \equiv \sum_{i=1}^{N} \theta_{ijt}\bar{\theta}_{ik}$$

- *Output Similarity* $\text{OutputSimilarity}^t_{jk}$:

$$\text{OutputSimilarity}^t_{jk} \equiv \sum_{i=1}^{N} \sigma_{ijt}\bar{\sigma}_{ik}$$

- *Input Similarity weighted by policy changes:*

$$\text{InputSimilarity-Dereservation}^t_{jk} \equiv \sum_{i=1}^{N} \delta_{ijt}\theta_{ijt}\bar{\theta}_{ik}, \qquad \text{InputSimilarity-Tariff}^t_{jk} \equiv \sum_{i=1}^{N} \Delta\tau_{ijt}\theta_{ijt}\bar{\theta}_{ik}$$

- *Output Similarity weighted by a policy change:*

$$\text{OutputSimilarity-Dereservation}^t_{jk} \equiv \sum_{i=1}^{N} \delta_{ijt}\sigma_{ijt}\bar{\sigma}_{ik}, \qquad \text{OutputSimilarity-Tariff}^t_{jk} \equiv \sum_{i=1}^{N} \Delta\tau_{ijt}\sigma_{ijt}\bar{\sigma}_{ik}$$

- *Upstream and Downstream:*

$$\text{Upstream}^t_{jk} = \sum_{i=1}^{N} \sigma^t_{ji}\bar{\theta}_{ik}, \qquad \text{Downstream}^t_{jk} = \sum_{i=1}^{N} \sigma^t_{ji}\bar{\theta}_{ki}.$$

C.3. **Sample definition.** Our sample consists of all plant-year observations between 2000/01 and 2009/10 that report being in operation and that report both physical intermediate inputs and outputs.

Appendix D. Theory Appendix

D.1. **Firm Input Choice.**

Proof of Proposition 1.

Proof. 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 $s_{\iota i t}$ charged to firms with be either the monopolistically competitive markup over b_1 or the limit price b_2, implying $s_{\iota i t} = \min\left\{\sigma b_1 / \left(\sigma - 1\right), b_2\right\}$.

Letting G denote the cdf of supplier cost draws and suppressing i, j and t subscripts, a general positive moment of S_{ijt}^{α} is

$$
\begin{aligned}
\mathrm{E}\left[S^{\alpha}\right] &= \mathrm{E}\left[S^{\alpha} \mid \sigma b_1/\left(\sigma-1\right) \leq b_2\right] + \mathrm{E}\left[S^{\alpha} \mid \sigma b_1/\left(\sigma-1\right) > b_2\right] \\
&= \int_{\underline{c}}^{\infty}\left(\frac{\sigma}{\sigma-1}b_1\right)^{\alpha}\int_{\frac{\sigma}{\sigma-1}b_1}^{\infty}\frac{N!b_2^{-\lambda(N-2)}s_m^{\lambda(N-2)}}{(N-2)!}dG\left(b_2\right)dG\left(b_1\right) \\
&+ \int_{\underline{c}}^{\infty}\int_{b_1}^{\frac{\sigma}{\sigma-1}b_1} b_2^{\alpha}\frac{N!b_2^{-\lambda(N-2)}s_m^{\lambda(N-2)}}{(N-2)!}dG\left(b_2\right)dG\left(b_1\right) \\
&= \int_{\underline{c}}^{\infty}\left(\frac{\sigma}{\sigma-1}b_1\right)^{\alpha}\int_{\frac{\sigma}{\sigma-1}b_1}^{\infty}\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\left(b_1\right) \\
&+ \int_{\underline{c}}^{\infty}\int_{b_1}^{\frac{\sigma}{\sigma-1}b_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\left(b_1\right) \\
&= \int_{\underline{c}}^{\infty}\left(\frac{\sigma}{\sigma-1}b_1\right)^{\alpha}\int_{\frac{\sigma}{\sigma-1}b_1}^{\infty}\frac{N!\lambda b_2^{-\lambda(N-1)-1}s_m^{\lambda(N-1)}}{(N-2)!}db_2 dG\left(b_1\right) \\
&+ \int_{\underline{c}}^{\infty}\int_{b_1}^{\frac{\sigma}{\sigma-1}b_1}\frac{N!\lambda b_2^{\alpha-\lambda(N-1)-1}s_m^{\lambda(N-1)}}{(N-2)!}db_2 dG\left(b_1\right) \\
&= \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\left(b_1\right) \\
&+ \int_{\underline{c}}^{\infty} N s_m^{\lambda(N-1)}\frac{\lambda\left(N-1\right)}{\lambda\left(N-1\right)-\alpha}\left[1-\left(\frac{\sigma}{\sigma-1}\right)^{\alpha-\lambda(N-1)}\right] b_1^{\alpha-\lambda(N-1)-1}dG\left(b_1\right) \\
&= \int_{\underline{c}}^{\infty}\lambda N s_m^{\lambda N}\left[\left(\frac{\sigma}{\sigma-1}\right)^{\alpha-\lambda(N-1)} + \frac{\lambda\left(N-1\right)}{\lambda\left(N-1\right)-\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\left(N-1\right)}{\lambda\left(N-1\right)-\alpha}\left[1-\left(\frac{\sigma}{\sigma-1}\right)^{\alpha-\lambda(N-1)}\right]\right]\underline{c}^{\alpha-\lambda N} \\
&= \frac{\lambda N}{\lambda N-\alpha}s_m^{\lambda N}\underline{c}^{-\lambda N+\alpha}\left(1-\frac{\alpha}{\lambda\left(N-1\right)-\alpha}\left(1-\left(\frac{\sigma}{\sigma-1}\right)^{-\lambda(N-1)+\alpha}\right)\right),
\end{aligned}
$$

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

□

Proof of Proposition 2.

Proof. 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 $\underline{c}_{jit}$, letting G denote the cdf of supplier cost draws is therefore

$$\begin{aligned}
\int_{\underline{c}_{jit}}^{\infty}\int_{\frac{\sigma}{\sigma-1}b_1}^{\infty}\frac{1}{\sigma}\frac{N_{it}!b_2^{-\lambda(N_{it}-2)}s_m^{\lambda(N_{it}-2)}}{(N_{it}-2)!}dG(b_2)\,dG(b_1) &= \int_{\underline{c}_{jit}}^{\infty}\int_{\frac{\sigma}{\sigma-1}b_1}^{\infty}\frac{1}{\sigma}\left(\frac{\lambda s_m^{\lambda}}{b_2^{\lambda+1}}\right)\frac{N_{it}!b_2^{-\lambda(N_{it}-2)}s_m^{\lambda(N_{it}-2)}}{(N_{it}-2)!}db_2 dG(b_1)\\
&= \int_{\underline{c}_{jit}}^{\infty}\frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{-\lambda(N_{it}-1)}\frac{N_{it}!b_1^{-\lambda(N_{it}-1)}s_m^{\lambda(N_{it}-1)}}{(N_{it}-1)!}\left(\frac{\lambda s_m^{\lambda}}{b_1^{\lambda+1}}\right)db_1\\
&= \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{-\lambda(N_{it}-1)}\underline{c}_{jit}^{-\lambda N_{it}}s_m^{\lambda N_{it}}.
\end{aligned}$$

The probability of limit pricing is therefore $\Pr(\text{LP}) = \left(1-\left(\frac{\sigma}{\sigma-1}\right)^{-\lambda(N_{it}-1)}\right)\underline{c}_{jit}^{-\lambda N_{it}}s_m^{\lambda N_{it}}$ while the probability of not producing (with a Lerner index of zero) is $1-\underline{c}_{jit}^{-\lambda N_{it}}s_m^{\lambda N_{it}}$. The aggregate Lerner index under limit pricing is

$$\begin{aligned}
&\Pr(\text{LP}) - \int_{\underline{c}_{jit}}^{\infty}b_1\int_{b_1}^{\frac{\sigma}{\sigma-1}b_1}b_2^{-1}\frac{N_{it}!b_2^{-\lambda(N_{it}-2)}s_m^{\lambda(N_{it}-2)}}{(N_{it}-2)!}dG(b_2)\,dG(b_1)\\
&= \Pr(\text{LP}) - \int_{\underline{c}_{jit}}^{\infty}b_1\int_{b_1}^{\frac{\sigma}{\sigma-1}b_1}\frac{\lambda N_{it}!b_2^{-\lambda(N_{it}-1)-2}s_m^{\lambda(N_{it}-1)}}{(N_{it}-2)!}db_2 dG(b_1)\\
&= \Pr(\text{LP}) - \int_{\underline{c}_{jit}}^{\infty}b_1\frac{\lambda N_{it}!b_1^{-\lambda(N_{it}-1)-1}s_m^{\lambda(N_{it}-1)}}{(\lambda(N_{it}-1)+1)(N_{it}-2)!}\left(1-\left(\frac{\sigma-1}{\sigma}\right)^{\lambda(N_{it}-1)+1}\right)\left(\frac{\lambda b_1^{-\lambda-1}}{s_m^{-\lambda}}\right)db_1\\
&= \Pr(\text{LP}) - \int_{\underline{c}}^{\infty}\frac{\lambda^2 N_{it}!b_1^{-\lambda N-1}s_m^{\lambda N_{it}}}{(\lambda(N_{it}-1)+1)(N_{it}-2)!}\left(1-\left(\frac{\sigma-1}{\sigma}\right)^{\lambda(N_{it}-1)+1}\right)db_1\\
&= \frac{1}{\lambda(N_{it}-1)+1}\left(1-\left(\frac{\sigma-1}{\sigma}\right)^{\lambda(N_{it}-1)+1}\right)\underline{c}_{jit}^{-\lambda N_{it}}s_m^{\lambda N_{it}}.
\end{aligned}$$

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

We define $\Lambda \equiv \lambda(N-1)+1 \geq 1-\lambda > 0$ 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 $\frac{d\ln\mathcal{L}}{dN}$ 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 $\Lambda\geq\sigma-1$ which holds for $N\geq 1+\frac{\sigma-2}{\lambda}$, the second term is negative. The upper bound for $\mathcal{L}$ comes from evaluation at $N=1+\frac{\sigma-2}{\lambda}$. □

Proof of Proposition 2.

Proof. We define $\Lambda\equiv\lambda\left(N-1\right)+1$, suppress i and t subscripts. For $\overline{R}\equiv\sum_j\sum_k\rho\overline{\theta}_kR_{jk}\left(s_m/\underline{c}_j\right)^{\lambda N}/N$, noting $\pi_{jk}=\left(1-\rho\right)R_{jk}$ and holding capabilities fixed that

$$\frac{\mathrm{d}\ln\overline{\mathrm{R}}}{\mathrm{dN}}=\frac{\sum_j\sum_k\overline{\theta}_k\cdot d\pi_{jk}\left(s_m/\underline{c}_j\right)^{\lambda N}/dN}{\sum_j\sum_k\overline{\theta}_k\pi_{jk}\left(s_m/\underline{c}_j\right)^{\lambda N}}-\frac{1}{N}=\frac{\sum_j\sum_k\overline{\theta}_k\left(s_m/\underline{c}_j\right)^{\lambda N}\cdot d\pi_{jk}/dN}{\sum_j\sum_k\overline{\theta}_k\pi_{jk}\left(s_m/\underline{c}_j\right)^{\lambda N}}+\frac{\sum_j\sum_k\overline{\theta}_k\pi_{jk}\cdot d\left(s_m/\underline{c}_j\right)^{\lambda N}/dN}{\sum_j\sum_k\overline{\theta}_k\pi_{jk}\left(s_m/\underline{c}_j\right)^{\lambda N}}.$$

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

$$\left.\frac{d\ln\pi_{jk}}{dN}\right|_{\underline{c}_{jt}\text{ fixed}}=-\frac{\rho}{1-\rho}\left.\frac{d\ln c_{jkt}}{dN}\right|_{\underline{c}_{jt}\text{ fixed}}=\frac{\rho}{1-\rho}\sum_i\frac{\overline{\theta}_{ik}}{\sigma-1}\frac{d\ln\frac{\Omega_{it}\vartheta_{it}}{\Omega_{it}+(\sigma-1)}}{dN}$$

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

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

$\lambda\left(N-1\right)+\sigma-1>0$ is assumed and implies that $\left(\frac{\sigma-1}{\sigma}\right)^{\lambda(N-1)+\sigma-1}<1$, which implies

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

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

$$\sum_j \sum_k \overline{\theta}_k \frac{d\pi_{jk}}{dN} \left(\frac{s_m}{\underline{c}_j}\right)^{\lambda N}$$
$$= \sum_j \sum_k \overline{\theta}_k \pi_{jk} \frac{\rho}{1-\rho} \frac{1}{\sigma-1} \left(\frac{1}{N} + \frac{\lambda}{\Phi} \frac{\left(1+\frac{\Phi}{\Phi+\lambda}\right) \frac{(\sigma-1)}{\Phi} \left(1 - \left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right) - (\sigma-1) \left(\frac{\sigma-1}{\sigma}\right)^{\Phi} \ln \frac{\sigma}{\sigma-1} - \frac{\Phi}{\Phi+\lambda}}{1 - \frac{\sigma-1}{\Phi}\left(1 - \left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right)} \right) \left(\frac{s_m}{\underline{c}_j}\right)^{\lambda N}$$
$$= \frac{\rho}{1-\rho} \frac{1}{\sigma-1} \left(\frac{1}{N} + \frac{\lambda}{\Phi} \frac{\left(1+\frac{\Phi}{\Phi+\lambda}\right) \frac{(\sigma-1)}{\Phi} \left(1 - \left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right) - (\sigma-1) \left(\frac{\sigma-1}{\sigma}\right)^{\Phi} \ln \frac{\sigma}{\sigma-1} - \frac{\Phi}{\Phi+\lambda}}{1 - \frac{\sigma-1}{\Phi}\left(1 - \left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right)} \right) \sum_j \sum_k \overline{\theta}_k \pi_{jk} \left(\frac{s_m}{\underline{c}_j}\right)^{\lambda N}.$$

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

$$W(N) \equiv \frac{\sum_j \sum_k \overline{\theta}_k \frac{\pi_{jk} d\left(s_m/\underline{c}_j\right)^{\lambda N}}{dN}}{\sum_j \sum_k \overline{\theta}_k \pi_{jk} \left(s_m/\underline{c}_j\right)^{\lambda N}} = \frac{\sum_j \sum_k \overline{\theta}_k \pi_{jk} \left(s_m/\underline{c}_j\right)^{\lambda N} \lambda \ln \frac{s_m}{\underline{c}_j}}{\sum_j \sum_k \overline{\theta}_k \pi_{jk} \left(s_m/\underline{c}_j\right)^{\lambda N}} \leq 0.$$

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

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

In what follows, the denominator $1 - \frac{\sigma-1}{\Phi}\left(1 - \left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right)$ is positive so long as $N \geq 1$ which is assumed. Since

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

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

Proof of Proposition 4.3.

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

$$\frac{d \ln \mathrm{E}[\zeta_\iota]}{dN} = \frac{d \ln \mathcal{L}}{dN} + \frac{\sum_j \sum_k \overline{\theta}_k \cdot d\pi_{jk} \left(s_m/\underline{c}_j\right)^{\lambda N} / dN}{\sum_j \sum_k \overline{\theta}_k \pi_{jk} \left(s_m/\underline{c}_j\right)^{\lambda N}} - \frac{1}{N}.$$

Also, direct inspection shows that

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

$$=\lambda\ln\frac{\sigma}{\sigma-1}\cdot\lim_{N\longrightarrow\infty}\frac{\frac{N}{\Lambda}-\frac{N}{\sigma-1}}{\frac{1}{\sigma-1}}=-\infty,$$

so from the proof of Proposition 3, we have

$$\lim_{N\longrightarrow\infty}\frac{d\ln\mathrm{E}\left[\zeta_\iota\right]}{dN}<\lim_{N\longrightarrow\infty}\frac{d\ln\mathcal{L}}{d\ln N}+\frac{\rho}{1-\rho}\frac{1-\lambda}{\sigma-1}-1=-\infty$$

which implies $\lim_{N\longrightarrow\infty}\mathrm{E}\left[\zeta_\iota\right]=0$. We also have from the proof of Proposition 3 that

$$\frac{d\ln\mathrm{E}\left[\zeta_\iota\right]}{dN}<\frac{d\ln\mathcal{L}}{d\ln N}+\frac{\rho}{1-\rho}\frac{1}{\sigma-1}\left(1+\frac{\lambda N}{\Phi}\frac{\left(1+\frac{\Phi}{\Phi+\lambda}\right)\frac{(\sigma-1)}{\Phi}\left(1-\left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right)-(\sigma-1)\left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\ln\frac{\sigma}{\sigma-1}-\frac{\Phi}{\Phi+\lambda}}{1-\frac{\sigma-1}{\Phi}\left(1-\left(\frac{\sigma-1}{\sigma}\right)^{\Phi}\right)}\right)-1.$$

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

(D.1)

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

Dropping the $\ln\frac{\sigma}{\sigma-1}$ terms in Equation (D.1) 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 $1+\frac{\sigma-2}{\lambda}$. Substituting this in to the last term we have

$$\frac{2}{\frac{\lambda+\sigma-2}{\sigma-1}-\frac{\lambda}{\sigma-2+\sigma-1}} \leq \frac{2}{\frac{\lambda+\sigma-2}{\sigma-1}-\frac{\lambda}{\sigma-1}} = 2\frac{\sigma-1}{\sigma-2}$$

so $2(\sigma-1)/(\sigma-2)$ is an upper bound for $B(N)$ for $N \geq 1+\frac{\sigma-2}{\lambda}$. We can conclude that $\frac{(1-\rho)(\sigma-1)}{\rho} \geq 2\frac{\sigma-1}{\sigma-2}$ implies expected profits are decreasing in entry which holds for $(1-\rho)\sigma \geq 2$. □

Proof of Proposition 1.

Proof. Firms solve

$$\min_{m_{ijkt}} \int_{\underline{c}_{ijt}}^{\infty} s_{\iota it} m_{\iota ijkt} dG_{it}(\iota) \text{ subject to } \left(\int_{\underline{c}_{ijt}}^{\infty} m_{\iota ijkt}^{(\sigma-1)/\sigma} dG_{it}(\iota)\right)^{\sigma/(\sigma-1)} \geq M_{ijkt}.$$

A natural question is why not frame this as a free endpoint problem with a choice of input varieties $\left[\underline{c}_{ijt}, \overline{c}_{ijt}\right]$. 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[37]

$$m_{\iota ijkt}^{(\sigma-1)/\sigma} = M_{ijkt}^{(\sigma-1)/\sigma}\left(\frac{\sigma}{\sigma-1}\frac{s_{\iota it}}{\eta}\right)^{1-\sigma} \text{ where } \eta_{it} = \left(-\int_{\infty}^{\underline{c}_{ijt}}\left(\frac{\sigma}{\sigma-1}s\right)^{1-\sigma} dG_{it}(s)\right)^{1/(1-\sigma)}.$$

Under these distributional assumptions, we have

$$\eta_{it} = \frac{\sigma}{\sigma-1}\left(\frac{\Omega_{it}}{\Omega_{it}+(\sigma-1)} s_m^{\Omega_{it}} \underline{c}_{ijt}^{1-\sigma-\Omega_{it}}\right)^{1/(1-\sigma)}$$

under the condition $\Omega_{it} > 1-\sigma$, η_{it} is finite and the input choice is non-degenerate.[38] Defining the cost index of input i as S_{ijt} we have minimum costs of $S_{ijt}M_{ijkt}$ where

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

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

[38]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 $\kappa \longrightarrow 0$, costs go to zero.

and therefore

$$d\ln S_{ijt}/d\ln \underline{c}_{ijt} = 1 + \Omega_{it}/\left(\sigma - 1\right).$$

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. □

Proof of Proposition 6.

Proof. 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 $p_{jkt} = c_{jkt}/\rho$ in the first maximization step, so the profit accruing from each industry is

$$\pi_{jkt} = \left(1/\rho - 1\right)\gamma\left(\underline{c}_{jt}\right)c_{jkt}q_{jkt} = \left(1/\rho - 1\right)\left(\rho D_{kt}\right)^{1/(1-\rho)}/\left(\gamma\left(\underline{c}_{jt}\right)c_{jkt}\right)^{\rho/(1-\rho)}. \tag{D.2}$$

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

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

it follows that the first order condition for profit maximization

$$\frac{d\pi_{jt}}{d\underline{c}_{ijt}} = \sum_k \frac{\pi_{jkt}}{\underline{c}_{ijt}}\frac{d\ln \pi_{jkt}}{d\ln \underline{c}_{ijt}} = -\frac{\rho}{1-\rho}\sum_k \frac{\pi_{jkt}}{\underline{c}_{ijt}}\left[\ln \underline{c}_{ijt} - \ln \underline{c}_{i0} + \overline{\theta}_{ik}\left(1 - \Omega_{it}/\left(1-\sigma\right)\right)\right] = 0. \tag{D.3}$$

Using the fact that $\rho\pi_{jkt}/\left(1-\rho\right) = \gamma\left(\underline{c}_{jt}\right)c_{jkt}q_{jkt}$, Equation (D.3) implies that for firm-input expenditure shares of θ_{ijt}, the optimal capability choice satisfies

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

Substitution into Equation (D.2) and further expansion shows that revenues R_{jkt} take the above form. □

D.2. **Extensive Product Margin.** Equation (4.6) can be modified to consider the extensive product margin choice of firms. Assume firms face a fixed cost $\left(1-\rho\right)f_{kt}$ to produce in an industry k each period, so produce when profits $\pi_{jkt} = \left(1-\rho\right)R_{jkt} > \left(1-\rho\right)f_{kt}$. From Equation (4.5), with identical coefficients and fixed effects similar to Equation (4.6) and error terms with $-\epsilon_{jkt}$ logistic, firms operate in industry k when either of the following equations is positive:

$$\ln \frac{R_{jkt}}{f_{kt}} = \kappa_{kt} + \kappa_{jk} - \kappa_0 \sum_i \left(\theta_{ijt} - \overline{\theta}_{ik}\right)^2 + \kappa_1 \sum_i \left(\alpha_B B_{ijt} + \alpha_\tau \Delta\tau_{ijt}\right) \left(\theta_{ijt} - \overline{\theta}_{ik}\right)^2 + \epsilon_{jkt}, \tag{D.4}$$

Equation (D.4) can be estimated to recover the tariff equivalent of dereservation on the extensive margin of industry adoption.

D.3. **Input Similarity Equation.**

Proposition 7.

Proof. Let $\{D_{kt}\}$ 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

$$\theta_{ijt} = \frac{\sum_k \overline{\theta}_{ik} C_{jk}}{C_j} = \frac{\sum_k \overline{\theta}_{ik} D_{kt}^{1/(1-\rho)} c_{jkt}^{-\rho/(1-\rho)}}{\sum_k D_{kt}^{1/(1-\rho)} c_{jkt}^{-\rho/(1-\rho)}}. \tag{D.5}$$

Holding $\underline{c}_{ijt}$ 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

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

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

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

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

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

adding $\overline{\theta}_{ik}^2$ and rearranging gives the result. □

D.4. **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_{jkt} in industry k at time t, firm j combines inputs from industry i, M_{ijkt}, and factor services from factor f, F_{fjkt}, using a constant return to scale Cobb-Douglas technology with industry input expenditure shares

$\overline{\theta}_{ik}, \overline{\theta}_{fk}$ and idiosyncratic industry productivity labeled φ_{jk}. At input prices $S_{ijt}\psi_{it}$ and factor service prices $W_{fjt}\psi_{ft}$, the unit cost of firm j to produce in industry k at time t, is therefore

$$c_{jkt} \equiv \prod_i \left(S_{ijt}\psi_{it}/\overline{\theta}_{ik}\varphi_{jk}\right)^{\overline{\theta}_{ik}} \cdot \prod_f \left(W_{fjt}\psi_{ft}/\overline{\theta}_{fk}\varphi_{jk}\right)^{\overline{\theta}_{fk}}.$$

Thus c_{jkt} is a vector of unit costs which 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 σ for inputs and σ_f for factor services where varieties follow a Pareto distribution with $\Pr\left(s_{\iota it} \geq s\right) = \left(s/s_m\right)^{-\Omega_{it}}$ for inputs and $\Pr\left(w_{\iota ft} \geq w\right) = \left(w/w_m\right)^{-\Omega_{ft}}$ for factors with $\Omega_{ft} = \lambda N_{ft}$ and N_{ft} is the mass of suppliers as modelled above but for factor services. Firms have capabilities of using inputs with prices $\left[\underline{c}_{ijt}, \infty\right)$ for inputs and $\left[\underline{c}_{fjt}, \infty\right)$ for factors where $\underline{c}_{ijt}$ and $\underline{c}_{fjt}$ are chosen by the firm. The analogous version of Proposition 1 go through with corresponding and symmetric terms for factors. Letting $\underline{c}_{jt}$ denote the vector of acquired capabilities, the actual unit costs of a multiproduct firm are given by $\gamma\left(\underline{c}_{jt}\right) c_{jkt}$ in each industry are assumed to follow

$$\gamma\left(\underline{c}_{jt}\right) \equiv \exp\left\{\sum_i \left(\ln \underline{c}_{ijt}\right)^2/2 + \sum_f \left(\ln \underline{c}_{fjt}\right)^2/2\right\}.$$

A firm can use its acquired capabilities across any number of products and re-optimizes 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

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

as above. A firm's profit maximizing capability and production choices considering product markets jointly are analogous to Proposition 6, in particular for $\Theta_{it} \equiv 1 + \Omega_{it}/\left(\sigma - 1\right)$ and $\Theta_{ft} \equiv 1 +$

$\Omega_{ft}/\left(\sigma_f-1\right)$, firm revenues are given by (where $\vartheta_{ft}\equiv 1+\frac{\sigma-1}{\lambda\left(N_{ft}-1\right)+\sigma-1}\left(1-\left(\left(\sigma-1\right)/\sigma\right)^{\lambda\left(N_{ft}-1\right)+\sigma-1}\right)$):

$$
\begin{aligned}
\ln R_{jkt} = & \underbrace{\frac{\rho}{1-\rho}\ln\varphi_{jk}}_{\text{RCA }(jk)} + \underbrace{\ln\left(\rho^{\frac{\rho}{1-\rho}}D_{kt}^{\frac{1}{1-\rho}}\right)}_{\text{Demand }(kt)} \\
& \underbrace{-\frac{\rho}{1-\rho}\sum_i\overline{\theta}_{ik}\ln\psi_{it}\vartheta_{ft}^{\frac{1}{1-\sigma}}\left(1-\Theta_{it}^{-1}\right)^{\frac{1}{1-\sigma}}\frac{s_m^{1-\Theta_{it}}}{\overline{\theta}_{ik}}}_{\text{Input Supplier }(kt)} \underbrace{-\frac{\rho}{1-\rho}\sum_f\overline{\theta}_{fk}\ln\psi_{ft}\vartheta_{ft}^{\frac{1}{1-\sigma}}\left(1-\Theta_{ft}^{-1}\right)^{\frac{1}{1-\sigma_f}}\frac{w_m^{1-\Theta_{ft}}}{\overline{\theta}_{fk}}}_{\text{Factor Supplier }(kt)} \\
& \underbrace{\underbrace{+\frac{\rho}{2\left(1-\rho\right)}\left[\sum_i\Theta_{it}^2\overline{\theta}_{ik}^2+\sum_f\Theta_{ft}^2\overline{\theta}_{fk}^2\right]}_{\text{Supplier}-\text{Tech }(kt)} \underbrace{-\frac{\rho}{2\left(1-\rho\right)}\left[\sum_i\Theta_{it}^2\left(\theta_{ijt}-\overline{\theta}_{ik}\right)^2+\sum_f\Theta_{ft}^2\left(\theta_{fjt}-\overline{\theta}_{fk}\right)^2\right]}_{\text{Core Competency }(jkt)}}_{\text{Comparative Advantage }(jkt)}
\end{aligned}
$$

Linearizing Ω_{it} and Ω_{ft} around the initial policy state $\Omega_{i0}=\overline{\Omega}$ and $\Omega_{f0}=\overline{\Omega}_f$ and letting κ_x represent a fixed effect for characteristic x yields Equation (5.2). The IV estimator is analogous to the one above, for $\widetilde{\theta}_{fjkt}\equiv\left(\theta_{fjt}\overline{\theta}_{fk}-\theta_{fjt}^2/2\right)$ and $\widetilde{\chi}_{fjkt}\equiv\chi_{jkt}\left(\theta_{fjt}\overline{\theta}_{fk}-\theta_{fjt}^2/2\right)$ we define

$$
\begin{aligned}
I_{jkt}^M\left(\lambda,\gamma\right) &\equiv \lambda\sum_i\widetilde{\theta}_{ijkt-1}-\gamma\sum_i\widetilde{\chi}_{ijkt-1}, & I_{jkt}^F\left(\lambda,\gamma\right) &\equiv \lambda\sum_f\widetilde{\theta}_{fjkt-1}-\gamma\sum_f\widetilde{\chi}_{fjkt-1}, \\
I_{jkt}^B\left(\lambda,\gamma\right) &\equiv \lambda\sum_i B_{ijt}\widetilde{\theta}_{ijkt-1}-\gamma\sum_i B_{ijt}\widetilde{\chi}_{ijkt-1}, & I_{jkt}^\tau\left(\lambda,\gamma\right) &\equiv \lambda\sum_i\tau_{ijt}\widetilde{\theta}_{ijkt-1}-\gamma\sum_i\tau_{ijt}\widetilde{\chi}_{ijkt-1}.
\end{aligned}
$$

The resulting first stage equations for our estimator are as follows for $\zeta^{ij}\equiv\left(\lambda^{ij},\gamma^{ij}\right)$:[39]

$$
\text{(D.6)}\qquad \sum_i\widetilde{\theta}_{ijkt}=\kappa_{kt}+\kappa_{jk}+I_{jkt}^M\left(\zeta^{11}\right)+I_{jkt}^F\left(\zeta^{12}\right)+I_{jkt}^B\left(\zeta^{13}\right)+I_{jkt}^\tau\left(\zeta^{14}\right)+\eta_{jkt}^M,
$$

$$
\text{(D.7)}\qquad \sum_f\widetilde{\theta}_{fjkt}=\kappa_{kt}+\kappa_{jk}+I_{jkt}^M\left(\zeta^{21}\right)+I_{jkt}^F\left(\zeta^{22}\right)+I_{jkt}^B\left(\zeta^{23}\right)+I_{jkt}^\tau\left(\zeta^{24}\right)+\eta_{jkt}^L,
$$

$$
\text{(D.8)}\qquad \sum_i B_{ijt}\widetilde{\theta}_{ijkt}=\kappa_{kt}+\kappa_{jk}+I_{jkt}^M\left(\zeta^{31}\right)+I_{jkt}^F\left(\zeta^{32}\right)+I_{jkt}^B\left(\zeta^{33}\right)+I_{jkt}^\tau\left(\zeta^{34}\right)+\eta_{jkt}^B,
$$

$$
\text{(D.9)}\qquad \sum_i\tau_{ijt}\widetilde{\theta}_{ijkt}=\kappa_{kt}+\kappa_{jk}+I_{jkt}^M\left(\zeta^{41}\right)+I_{jkt}^F\left(\zeta^{42}\right)+I_{jkt}^B\left(\zeta^{43}\right)+I_{jkt}^\tau\left(\zeta^{44}\right)+\eta_{jkt}^\tau.
$$

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

[39]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

APPENDIX E. AVERAGE FIRM-LEVEL COMPARATIVE ADVANTAGE, BY INDUSTRY

Table 28 shows the average comparative advantage of single-industry firms in industry k', for the industry in which they enjoy the highest average CA_{jkt}.

TABLE 28. Comparative Advantage of Single-industry Firms, by Industry

Industry k'	Highest average comparative advantage industry (except k')	Comp Adv
Dairy products	Live animals, chiefly for food	15.8**
Other jute and natural fibre goods, n.e.c.	Fabrics & cloth of jute, coir, sisal, hemp, mista etc.	13.1**
Fabrics & cloth of jute, coir, sisal, hemp, mista etc.	Other jute and natural fibre goods, n.e.c.	12.3**
Fibre of jute, coir, and other plants	Fabrics & cloth of jute, coir, sisal, hemp, mista etc.	11.7*
Cereals (incl. rice) and pulses, unmilled	Products of milling industries; malt & malted milk	11.6**
Products of milling industries; malt & malted milk	Cereals (incl. rice) and pulses, unmilled	11.5*
Ginned cotton, cotton, and raw cotton waste	Cotton yarn and fibre, incl. cotton thread	10.2**
Cotton yarn and fibre, incl. cotton thread	Ginned cotton, cotton, and raw cotton waste	10.0*
Vegetables oils & fats	Diesel products & by-products.	9.8
Raw fibre of jute, coir, sisal, hemp, mista etc	Fabrics & cloth of jute, coir, sisal, hemp, mista etc.	9.6
Aluminium and aluminium alloys, unwrought	Aluminium and aluminium alloys worked	9.5**
Leather apparel	Leather bags, cases, purse & other novelty items	9.2**
Fruit juices and vegetable juices & syrup, pickles	Edible fruits & nuts; edible vegetables and certain roots	9.2
Craft paper and paper for special use	Boards, paper boards	9.1**
Leather bags, cases, purse & other novelty items	Leather apparel	9.0**
Boards, paper boards	Craft paper and Paper for special use.	8.7
Chocolate, cocoa & cocoa preparations and sugar	Sugar, Mollasses, Khandsari, Gur.	8.6
Edible fruits & nuts; edible vegetables and certain roots	Fruit juices and vegetable juices & syrup, Pickles	8.5**
Aluminium and aluminium alloys worked	Aluminium and Aluminium alloys, unwrought	8.2
Paper (uncoated) used for newsprint and for other special purposes	Craft paper and paper for special use	8.0
Pig Iron/Ferro alloys etc. in primary form	Metro railways and tramways and rolling stock	7.9**
Cotton apparel	Fur skins and articles thereof	7.7
Inorganic elements, excl. base metals, rare gas	Charcoal	7.4
Misc. leather manufactured items	Leather bags, cases, purse & other novelty items	7.3
Copper & copper alloy, refined or not, unwrought	Copper and copper alloys, worked	7.0**

Note: Table shows the average comparative advantage $\overline{CA}_{jkt}$ of single-industry plants in industry k', for the industry k where $\overline{CA}_{jkt}$ is the highest. $^{**}p < 0.05,^{*}\ p < 0.10$.

APPENDIX F. CORE COMPETENCY PREMIA WITH FACTOR COMPLEMENTARITY

TABLE 29. Core Competency Sales Premium, with Primary Factors

Industry rank	Number of Industries With Positive Sales									
	1	2	3	4	5	6	7	8	9	10+
1	0.024	0.030	0.030	0.017	0.013	0.011	0.010	0.010	0.007	0.011
2		0.014	0.012	0.009	0.009	0.008	0.008	0.008	0.006	0.014
3			0.010	0.008	0.007	0.007	0.006	0.006	0.008	0.009
4				0.007	0.007	0.006	0.006	0.005	0.006	0.009
5					0.006	0.006	0.006	0.005	0.006	0.005
6						0.006	0.006	0.006	0.006	0.004
7							0.005	0.005	0.005	0.004
8								0.005	0.005	0.005
9									0.005	0.005
10+										0.003

TABLE 30. Core Competency Sales Premium, with Primary Factors – Weighted

Industry rank	Number of Industries With Positive Sales									
	1	2	3	4	5	6	7	8	9	10+
1	0.029	0.038	0.070	0.085	0.076	0.077	0.097	0.168	0.299	1.102
2		0.003	0.007	0.022	0.098	0.198	0.166	0.205	0.014	2.214
3			0.001	0.003	0.004	0.031	0.012	0.023	0.144	0.858
4				0.001	0.004	0.025	0.010	0.014	0.014	0.116
5					0.002	0.005	0.008	0.006	0.011	0.029
6						0.002	0.004	0.006	0.004	0.007
7							0.001	0.004	0.003	0.011
8								0.001	0.001	0.004
9									0.003	0.003
10+										0.001

APPENDIX G. ONLINE APPENDIX

G.1. **Input Unit Values and Dereservation.** Table 31 shows results of a regression of log unit values of domestically sourced intermediate inputs (by 5-digit input category i) on dereservation and tariff changes.

G.2. **Additional Robustness Checks. Clustering at the firm level:** Tables 32 and 33 below, which are the main reduced-form regression table and its robustness table from the paper, are re-estimated with standard errors clustered at the firm level.

Single-industry firms: Table 34 below shows the benchmark reduced-form regressions using only single-industry firm-year observations (producers that are currently producing in one three-digit industry only).

TABLE 31. Domestic Input Unit Values After Dereservation

	Dependent variable: log p_{jit}	
	(1)	(2)
$t \geq$ year i was de-reserved	-0.139**	-0.0915**
	(0.014)	(0.015)
log InputTariff$_{it}$	-0.0706**	-0.0344**
	(0.0051)	(0.0053)
Year FE	Yes	Yes
Input Product FE	Yes	
Firm $\times$ Input Product FE		Yes
R^2	0.847	0.954
Observations	861991	491953

Standard errors in parentheses, clustered at the firm-year level.
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

TABLE 32. Industry Entry with Dereservation, Clustering at firm level

	Dependent variable: Add$_{jkt}$			
	(1)	(2)	(3)	(4)
InputSimilarity$^0_{jk}$	0.0379**	0.0371**	0.0273**	0.0268**
	(0.00041)	(0.00041)	(0.00062)	(0.00062)
InputSimilarity-Dereservation$^0_{jkt}$	0.0429**	0.0424**	0.0203**	0.0192**
	(0.0027)	(0.0027)	(0.0025)	(0.0025)
InputSimilarity-Tariff$^0_{jkt}$				-0.0701**
				(0.010)
Firm $\times$ Year FE α_{jt}	Yes	Yes	Yes	Yes
Industry $\times$ Year FE α_{kt}		Yes		
$k \times k' \times t$ FE $\alpha_{kk't}$			Yes	Yes
R^2	0.00981	0.0118	0.0575	0.0576
Observations	52691029	52691029	52666907	52666907

Standard errors in parentheses, clustered at the firm level.
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

Industries with Many Firms: There are few industries with just a few producers. 90% of all industry-years have more than 8 producers (75% more than 26 producers). More than 70% of producers in those industries with less than 8 producers are multi-industry firms. Table 35 shows the main reduced-form regressions on the subsample of industry-years with more than 8 producers. Results are very similar to those from the baseline specifications.

TABLE 33. Industry Entry with Dereservation, Clustering at firm level - Robustness

	Dependent variable: Add_{jkt}				
	(1)	(2)	(3)	(4)	(5)
$\text{InputSimilarity}^0_{jk}$	0.0379**	0.0251**	0.0245**	0.0199**	0.0195**
	(0.00041)	(0.00049)	(0.00050)	(0.00060)	(0.00061)
$\text{InputSimilarity-Dereservation}^0_{jkt}$	0.0429**	0.0383**	0.0378**	0.0155**	0.0145**
	(0.0027)	(0.0027)	(0.0026)	(0.0025)	(0.0025)
$\text{OutputSimilarity}^0_{jk}$		0.0136**	0.0136**	0.100**	0.100**
		(0.00060)	(0.00061)	(0.0018)	(0.0018)
$\text{OutputSimilarity-Dereservation}^0_{jkt}$		0.0344**	0.0334**	0.0171**	0.0171**
		(0.0016)	(0.0016)	(0.0022)	(0.0022)
Upstream^0_{jk}		0.0335**	0.0315**	0.0291**	0.0291**
		(0.00092)	(0.00092)	(0.0029)	(0.0029)
Downstream^0_{jk}		-0.00826**	-0.00756**	-0.00351*	-0.00356*
		(0.00062)	(0.00062)	(0.0015)	(0.0015)
$\text{InputSimilarity-Tariff}^0_{jkt}$					-0.0640**
					(0.0100)
Firm $\times$ Year FE α_{jt}	Yes	Yes	Yes	Yes	Yes
Industry $\times$ Year FE α_{kt}			Yes		
$k \times k' \times t$ FE $\alpha_{kk't}$				Yes	Yes
R^2	0.00981	0.0122	0.0140	0.0646	0.0646
Observations	52691029	52691029	52691029	52666907	52666907

Standard errors in parentheses, clustered at the firm level.
$^+ p < 0.10$, $^* p < 0.05$, $^{**} p < 0.01$

Wholesalers Excluded: The ASI contains a survey question, G11, which asks for the "sale value of goods sold in the same condition as purchased". That value is missing for about 12% of observations. Among those with nonmissing observations, it is zero for about 66% of observations, below one percent of *manufacturing* gross output for 78% of observations, and below 5 percent for 84% of observations. In Table 36 below, we show the regressions from the reduced-form section of the paper for observations that report a G11 of less than one percent (column (2)), less than five percent (column (3)), and less than ten percent (column (4)) of manufacturing gross output. The results are almost the same as for the full sample.

G.3. **Estimated Technology Changes from Dereservation.** Figure G.1 provides a histogram of coefficients from regressing the average single product firm 3-digit expenditure shares $\overline{\theta}_{ikt}$ each period on fixed effects for each input-industry and whether a within industry i has been dereserved at time t. While the estimates are on average slightly negative with a mean of -0.01 and

TABLE 34. Industry Entry – Single-industry Firms Only

	Dependent variable: Add_{jkt}					
	(1)	(2)	(3)	(4)	(5)	(6)
$\text{InputSimilarity}^0_{jk}$	0.0371**	0.0246**	0.0367**	0.0244**	0.0294**	0.0223**
	(0.00046)	(0.00055)	(0.00046)	(0.00056)	(0.00074)	(0.00072)
$\text{InputSimilarity-Dereservation}^0_{jkt}$	0.0358**	0.0320**	0.0355**	0.0317**	0.0155**	0.0109**
	(0.0030)	(0.0029)	(0.0030)	(0.0029)	(0.0029)	(0.0028)
$\text{OutputSimilarity}^0_{jk}$		0.0139**		0.0137**		0.120**
		(0.00072)		(0.00072)		(0.0025)
$\text{OutputSimilarity-Dereservation}^0_{jkt}$		0.0286**		0.0279**		0.0104**
		(0.0018)		(0.0018)		(0.0026)
Upstream^0_{jk}		0.0333**		0.0321**		0.0347**
		(0.0011)		(0.0011)		(0.0041)
Downstream^0_{jk}		-0.0103**		-0.00976**		-0.00434*
		(0.00069)		(0.00069)		(0.0021)
$\text{InputSimilarity-Tariff}^0_{jkt}$					-0.0488**	-0.0420**
					(0.012)	(0.012)
Firm × Year FE α_{jt}	Yes	Yes	Yes	Yes	Yes	Yes
Industry × Year FE α_{kt}			Yes	Yes		
$k \times k' \times t$ FE $\alpha_{kk't}$					Yes	Yes
R^2	0.0100	0.0131	0.0117	0.0146	0.0783	0.0900
Observations	35318097	35318097	35318097	35318097	35286189	35286189

Standard errors in parentheses, clustered at the firm-industry level.
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

TABLE 35. Industry Entry with Dereservation – Industry-Years with > 8 producers

	Dependent variable: Add_{jkt}			
	(1)	(2)	(3)	(4)
$\text{InputSimilarity}^0_{jk}$	0.0443**	0.0434**	0.0344**	0.0338**
	(0.00042)	(0.00042)	(0.00072)	(0.00073)
$\text{InputSimilarity-Dereservation}^0_{jkt}$	0.0521**	0.0516**	0.0243**	0.0230**
	(0.0029)	(0.0029)	(0.0028)	(0.0028)
$\text{InputSimilarity-Tariff}^0_{jkt}$				-0.0825**
				(0.011)
Firm × Year FE α_{jt}	Yes	Yes	Yes	Yes
Industry × Year FE α_{kt}		Yes		
$k \times k' \times t$ FE $\alpha_{kk't}$			Yes	Yes
R^2	0.0113	0.0131	0.0584	0.0584
Observations	44366233	44366233	44345156	44345156

Standard errors in parentheses, clustered at the firm-industry level.
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

TABLE 36. Industry Entry, with No Wholesalers

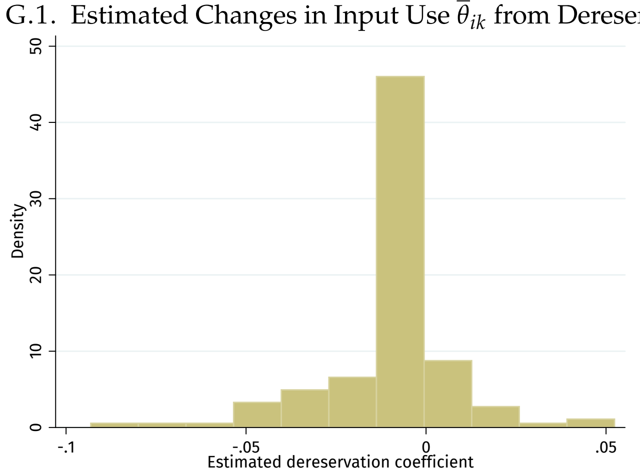
	Dependent variable: Add_{jkt}			
	(1)	(2)	(3)	(4)
$\text{InputSimilarity}^0_{jk}$	0.0195**	0.0197**	0.0198**	0.0196**
	(0.00057)	(0.00065)	(0.00063)	(0.00062)
$\text{InputSimilarity-Dereservation}^0_{jkt}$	0.0145**	0.0131**	0.0138**	0.0140**
	(0.0023)	(0.0026)	(0.0025)	(0.0025)
$\text{InputSimilarity-Tariff}^0_{jkt}$	-0.0640**	-0.0494**	-0.0523**	-0.0564**
	(0.0095)	(0.011)	(0.010)	(0.010)
$\text{OutputSimilarity}^0_{jk}$	0.100**	0.102**	0.101**	0.101**
	(0.0018)	(0.0021)	(0.0020)	(0.0020)
$\text{OutputSimilarity-Dereservation}^0_{jkt}$	0.0171**	0.0152**	0.0159**	0.0165**
	(0.0022)	(0.0026)	(0.0025)	(0.0025)
Upstream^0_{jk}	0.0291**	0.0314**	0.0314**	0.0313**
	(0.0030)	(0.0034)	(0.0033)	(0.0033)
Downstream^0_{jk}	-0.00356*	-0.00268	-0.00294+	-0.00284+
	(0.0014)	(0.0017)	(0.0016)	(0.0016)
Sample		$\chi < 0.01$	$\chi < 0.05$	$\chi < 0.10$
Firm × Year FE α_{jt}	Yes	Yes	Yes	Yes
Industry × Year FE α_{kt}				
$k \times k' \times t$ FE $\alpha_{kk't}$	Yes	Yes	Yes	Yes
R^2	0.0646	0.0692	0.0676	0.0669
Observations	52666907	36046972	39388763	41037045

Standard errors in parentheses, clustered at the firm-industry level.
$^+ p < 0.10$, $^* p < 0.05$, $^{**} p < 0.01$

Notes: χ is the fraction of sales from wholesaling activity (G11) in manufacturing gross output (total sales in the J block).

standard deviation of 0.018, indicating the average movement is very small and the distribution of changes are hard to distinguish from zero.

FIGURE G.1. Estimated Changes in Input Use $\overline{\theta}_{ik}$ from Dereservation



G.4. **Structural Robustness.** In parallel with the reduced form results, Table 37 presents structural estimation results controlling for $k \times k' \times t$ fixed effects. This is overcontrolling relative to the theory and we lose some precision but it reaffirms that the IV results still survive in the full specification.

TABLE 37. Structural results with (k, k', t) fixed effects

	Dependent variable: $\mathbf{1}(Sales_{jkt} > 0)$			
	(1)	(2)	(3)	(4)
$\sum_i (\theta_{ijt}\overline{\theta}_{ik} - \overline{\theta}_{ik}^2/2)$	0.0037**	0.0038**	0.0185	0.033*
	(0.0001)	(0.0001)	(0.0128)	(0.0134)
$\sum_i B_{it} \cdot (\theta_{ijt}\overline{\theta}_{ik} - \overline{\theta}_{ik}^2/2)$	-0.0001	-0.0001	-0.0004	-0.0006$^+$
	(0.0001)	(0.0001)	(0.0004)	(0.0003)
$\sum_i \tau_{it} \cdot (\theta_{ijt}\overline{\theta}_{ik} - \overline{\theta}_{ik}^2/2)$		-0.0003		-0.0039*
		(0.0003)		(0.0018)
κ_{jk}	Yes	Yes	Yes	Yes
κ_{jt}	Yes	Yes	Yes	Yes
$\kappa_{kk't}$	Yes	Yes	Yes	Yes
Estimator	OLS	OLS	IV	IV
R^2	0.862	0.862	0.857	0.857
Observations	77,745,382	77,745,382	46,185,150	46,185,150

Standard errors in parentheses, clustered at the firm-industry level.
$^+$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$