

# THE COMPARATIVE ADVANTAGE OF FIRMS

Johannes Boehm<sup>a</sup>, Swati Dhingra<sup>b</sup>, John Morrow<sup>c</sup>

April 17, 2019

**ABSTRACT.** Multiproduct firms dominate production, and their product turnover contributes substantially to aggregate growth. Theories propose that multiproduct firms grow by diversifying into products which need the same know-how or capabilities, but are less clear on what these capabilities are. Input-output tables show firms co-produce in industries that share intermediate inputs, suggesting input capabilities drive multiproduct production patterns. We provide evidence for this in Indian manufacturing: the similarity of a firm's input mix to an industry's input mix predicts entry into that industry. We identify the direction of causality from the removal of size-based entry barriers in input markets which made firms more likely to enter industries that were similar in input use to their initial input mix. We rationalize this finding with a model of industry choice and economies of scope to estimate the importance of input capabilities in determining comparative advantage. Complementarities driven by input capabilities make a firm on average 5% (and up to 15%) more likely to produce in an industry. Entry barriers in input markets constrained the comparative advantage of firms and were equivalent to a 10.5 percentage point 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.

<sup>a</sup>Sciences Po and CEP/LSE.

<sup>b</sup>Centre for Economic Performance, LSE, and CEPR.

<sup>c</sup>Kings College London, CEP/LSE, and CEPR. Corresponding Author: john.morrow@kcl.ac.uk.

## 1. INTRODUCTION

Theories of the firm, dating back to [Penrose \(1955\)](#), propose that successful product diversification is an engine of corporate growth. It enables firms to avoid the limits to growth imposed by the size of a single product market. Indeed, 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.<sup>1</sup> Recent work in international economics and industrial organization examines how many products firms make and the impact of economic changes on these choices. It emphasizes the importance of core products for firm growth,<sup>2</sup> but little is known about why these products are ‘core’.

Early theoretical work takes the view that a firm consists of a bundle of productive capabilities that can be used to produce a variety of products ([Marris 1964](#)). Different products require different knowhow or input capabilities, and firms differ in the capabilities they have. Capabilities are tied to the firm as they often cannot be bought ‘off the shelf’ ([Teece 1980](#); [Scherer 1982a](#); [Sutton 2012](#)). They are costly to acquire, so firms make products that share capabilities to benefit from economies of scope in acquiring them. Firms easily diversify 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](#)).

---

<sup>1</sup>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.

<sup>2</sup>[Bernard et al. 2010, 2011](#); [Eckel and Neary 2010](#); [Eckel et al. 2009](#); [Mayer et al. 2009](#); [Iacovone and Javorcik 2010](#).

With the increased availability of micro-data on firms and their product mix, evidence is emerging on the patterns of co-production by firms across industries. Using US data, [Bernard et al. \(2010\)](#) find that firms are much more likely to produce in certain pairs of industries. Many of these pairs suggest a possible role for input-based co-production within firms. 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, Fabricated Metal and Industrial Machinery. Similar patterns emerge in firm-level data from the United Kingdom and Belgium ([Hutchinson et al. 2010](#), [Bernard et al. 2018](#)).

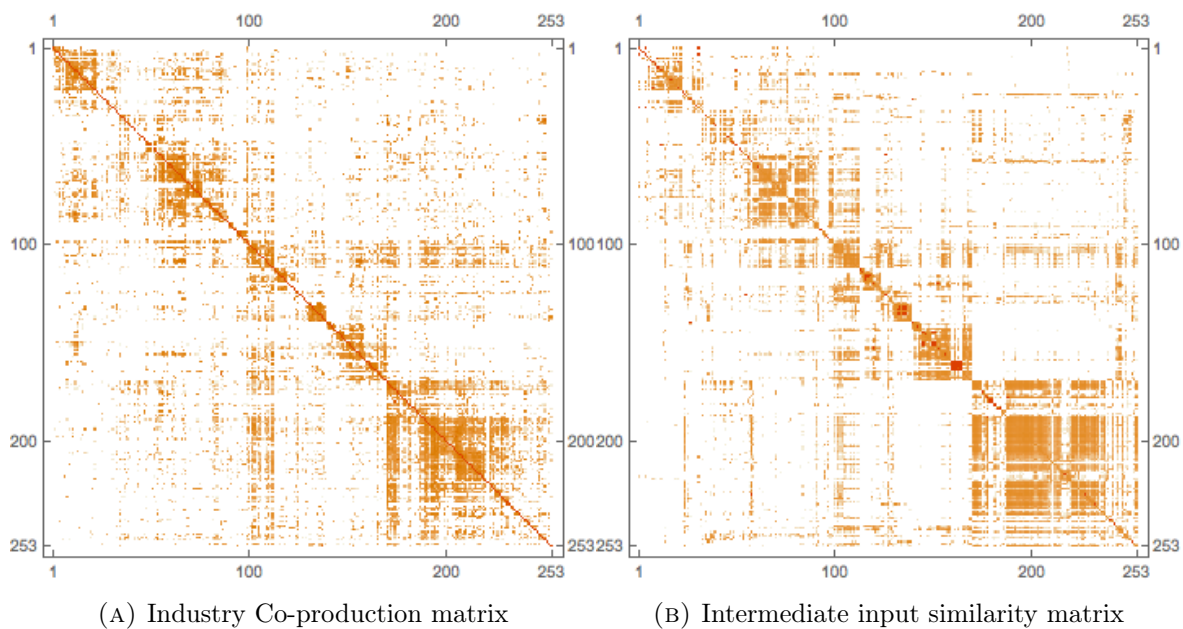


FIGURE 1.1. Co-production and Input Similarity

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. Darker values indicate larger numbers. Intermediate input shares (right matrix) are constructed from single-industry plants only. Plant-year observations are value-weighted. The correlation between values in the left and right matrices is 0.5.

Connecting the co-production patterns with shared input use, a first glance at plant-level data from India shows a striking pattern. Firms tend to co-produce in industries

that require similar intermediate inputs. Figure 1.1 shows the extent of co-production within plants (across 1 to 253 different industries; left panel) and a measure for input similarity between industry pairs on the right panel. Overall, there is a strong correlation between the extent of co-production of industries and the degree to which they share inputs.

The Figure suggests input linkages play a role in co-production patterns. For example, Leather Apparel and Footwear have the same major inputs, so firms tend to co-produce them. Despite the strong relationship between co-production and shared input use, there are other potential drivers of the observed co-production patterns. For example, consumers might demand new clothes together with new shoes. Multi-product firms could then internalize these demand complementarities and sell leather apparel along with footwear. However, if income growth makes consumers more likely to spend on leather items, demand for leather apparel and footwear could co-move at the macroeconomic level but not within firms. Disentangling these different explanations has been difficult because standard firm-level data records equilibrium product choices, and exogenous variation in demand or supply-side conditions is needed to identify the existence of specific linkages across products.<sup>3</sup>

This paper addresses the question of product choice microeconomically by focusing on plausibly exogenous variation in input supply from a policy change in the Indian manufacturing sector and by building on the literature on comparative advantage to define production patterns at the plant level. A large literature has shown that firms in developing countries are typically smaller, less productive and grow less (relative to firms in developed countries), and that supply-side bottlenecks, such as government policies on infrastructure and product market regulations, continue to constrain firm growth (Tybout 2000; Bloom et al. 2010). Building on these observations, this paper

---

<sup>3</sup>While we focus on economies of scope in input capabilities, in our reduced form we examine other possible firm-industry linkages before controlling for them with firm-industry fixed effects in our structural results.

considers input supply policies which enable identification of supply-side linkages that boost firm growth.

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 plants.<sup>4</sup> As the entry barriers were lifted, plants experienced better access to inputs. Plants intensively using these inputs were more likely to grow by diversifying into products also intensive in the use of these inputs. To concretize ideas, 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 plant that is relatively intensive in cotton becomes relatively more likely to move into Cotton Textile production.

The paper uses the policy change to operationalize comparative advantage at the plant level. According to comparative advantage theory, industries differ in the technology or the factors needed to produce them and countries differ in their technological prowess or factor endowments. Countries therefore produce relatively more in industries which they are more capable of producing in (through better technologies or greater reliance on the factors that countries are abundant in). Translating this from countries and technologies/factors to plants and inputs, this paper exhibits how better input supply enabled plants to raise production in their comparative advantage industries by more than the typical plant in those industries. As in the comparative advantage literature, industry differences are measured through input requirements, which are computed from the average shares of intermediate input use of single-industry plants. In our reduced form, plants' input capabilities are measured through their initial input intensities, which is computed from the initial shares of input use to capture

---

<sup>4</sup>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. The removal of entry barriers was driven primarily by the agenda of the Indian government to reform post-independence economic policy.

revealed comparative advantage. Comparative advantage then predicts plants would grow by diversifying into products that require an input mix similar to the plant's revealed input capabilities.

Input similarity is measured as the inner product of a plant's input shares and an industry's input shares to account for the correlation in input mix between plants and industries. Comparative advantage then predicts plants would grow by diversifying into products that require an input mix similar to the plant's revealed input capabilities. The results show input similarity makes it both more likely for a plant to add an industry and less likely for a plant to drop an industry from its product portfolio. The removal of entry barriers, which gave firms better access to input supplies, enables an examination of how policy interacts with input similarity to affect the likelihood of diversifying into similar industries. Input similarity makes it both more likely for a plant to add an industry and less likely for a plant to drop an industry from its product portfolio. This is related to product-level findings of [Schott \(2004\)](#), which shows that countries' within-product specialization reflects factor-based comparative advantage.

Having established a role for input linkages across industries, the paper provides a theoretical framework for input-based comparative advantage of firms. Starting from the primitive of industry-specific production functions, differences across firms arise from their idiosyncratic industry-productivities and endogenous decisions to invest in input capabilities. Firms acquire input capabilities by investing resources and deploying them across industries. Sharing input capabilities provides economies of scope which induces co-production in industries that are intensive in the use of the 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.

A key theoretical insight of our framework is that economies of scope within multi-product firms imply production choices and input capabilities are jointly determined. Unit costs across industries for multiproduct firms are interdependent on the relative demands a firm faces in the industries it operates in. The framework generates structural estimating equations that explain the portfolio of industries a firm adopts based on its extent of input similarity with each industry. Policy changes that improve access to inputs heighten these economies of scope and allow us to quantify their magnitude with parameter estimates.

A key econometric insight of our framework 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 guides estimation of common industry demand innovations to predict contemporaneous input similarity, which in turn determines product choice. The results show that input capabilities are quantitatively important in determining the production patterns of firms.

Quantitatively we find that on average, input-based comparative advantage makes single industry firms 5.2 per cent more likely to produce in an industry. This effect spreads across industries for multi-industry firms through economies of scope, but diffuses as input capabilities are not customized to any one industry. For instance, nine industry firms are from .8% to 1.4% more likely to produce in an industry (decreasing in sales rank). However, as multi-industry firms are larger across the board, the size-weighted premium from input capabilities ranges from .5% to 46.8%, showing that input-based comparative advantage has sizable impacts for firm growth.

We quantify entry barriers in terms of tariff rates that have equivalent effects on firm decisions to move into industries. On average, entry barriers from the policy to reserve products for small scale plants are equivalent to input tariffs of 10.5 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.<sup>5</sup> The Indian context overcomes these problems to reveal the constraints placed by domestic policy on firms and its comparison with trade policy.

**Related Literature.** The results relate to the multiproduct firm literature, which usually focuses on how many, not which, products firms make. We contribute to this literature by identifying the role of input linkages as a determinant of the core competencies of multiproduct firms.<sup>6</sup> A large literature studies the role of access to inputs on firm productivity.<sup>7</sup> While we ask a different question, the focus on input supply is consistent with these studies. Specifically, [Goldberg et al. \(2009\)](#) 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.<sup>8</sup> 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.

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

---

<sup>5</sup>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.

<sup>6</sup>See also [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. \(2019\)](#) in the firm heterogeneity literature.

<sup>7</sup>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.

<sup>8</sup>[Vandenbussche and Viegelaahn \(2014\)](#) also show that Indian firms move away from inputs facing domestic anti-dumping measures by decreasing sales of products using these inputs.



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 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 recent 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. It confirms the view of Hausmann and Hidalgo (2011) that a firm which has previously developed a transcontinental aircraft and a combustion engine is likely to have a lower cost of developing a regional jet aircraft, relative to a firm which has 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.<sup>9</sup> 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 misallocation 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.<sup>10</sup> Recent macroeconomic studies stress the importance of input linkages in amplifying micro shocks and policy effects.<sup>11</sup> The development literature emphasizes their role in aggregate productivity and volatility (Koren and Tenreyro 2013), and in motivating policies such as domestic content requirements that have interested governments across the developing world (Harrison and Rodriguez-Clare 2009). 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.

---

<sup>9</sup>In early work, Scherer (1982b) estimates technology flows from data on the proportion of patents filed in origin industries used in destination industries and interindustry economic transfers drawn from the input-output tables to understand the slowdown in productivity growth in the US. Recent work has built on these findings to show a positive relationship between technological relatedness or input relatedness and various firm performance measures (Robins and Wiersema 1995; Bowen and Wiersema 2005; Bryce and Winter 2009; Fan and Lang 2000; Liu 2010; Rondi and Vannoni 2005). 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.

<sup>10</sup>There are a growing number of studies relating linkages to productivity (see the forthcoming handbook chapter by Combes and Gobillon 2014). 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.

<sup>11</sup>Example, Acemoglu et al. (2012), Di Giovanni et al. (2014), and early work by Jovanovic (1987) and Durlauf (1993).

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 the model, instrumentation strategy and the results from structural estimation and quantification of input capabilities. Section 5 concludes.

## 2. DATA AND STYLIZED FACTS

**2.1. Data Description.** We use annual data on manufacturing firms from the Indian Annual Survey of Industry (ASI), which is conducted by the Indian Ministry of Statistics and Programme Implementation. The ASI is the Indian government's main source of industrial statistics on the formal manufacturing sector, and consists of two parts: a census of all manufacturing plants that are larger than 100 employees, and a random sample of one fifth of all plants that employ between 20 and 100 workers (between 10 and 100 workers if the plant uses power). The ASI's sampling methodology and product classifications have changed several times over the course 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. Secondly, the same product codes are used to describe both inputs and outputs of plants. This enables us to treat inputs and outputs symmetrically.

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.<sup>12</sup>

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).

The unit of observation in our dataset is generally the plant, except if the firm owns other plants belonging to the same industry in the same state, in which case the unit of observation is the aggregate of those plants. For our purposes, 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 need not pick up other firm-wide, not just plant-wide, mechanisms, which could also be at play. 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.

---

<sup>12</sup>According to the ASI, the product classification is stratified into 2-digit sectors, 3-digit industries and 5-digit products.

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 |
| 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 questions which industries the firms are producing in. Figure 1.1a 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. Nevertheless, there is much 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 1.1a shows a matrix that captures the similarity of the row and column industries' mix of intermediate inputs. Each element  $(m, n)$  is the inner product of the industries' vector of intermediate input expenditure shares:

$$\bar{\text{IS}}_{mn} = \sum_i \bar{\theta}_{mi} \bar{\theta}_{ni}$$

where  $\bar{\theta}_{mi}$  is the sum of expenditure of single-industry firms that only produce  $m$  on intermediate inputs from  $i$ , divided by total expenditure of these firms on intermediate inputs. This measure captures the overlap in industry  $m$  and  $n$ 's intermediate input mixes.

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

We now turn to the determinants of firms' revealed comparative advantage – the extensive and intensive margins of the firms' product mix. 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 mixes explain 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, after a short case study at the end of this Section. The estimating equation in that model bears a close resemblance to the reduced-form regressions from this Section, but provides a structural interpretation of 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,  $\theta_j$ , with the vector of intermediate input expenditure shares of an industry  $k$ :

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

where  $i$  indexes the expenditure shares of spending on three-digit inputs and  $t$  denotes time. We construct the aggregate intermediate input shares  $\bar{\theta}_{ki}$  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  $\theta_{ij}$ . As an inner product of a vector of

firm and industry shares, our 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 of this inner product 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 movements in the industry space. To avoid the possibility that changes in the input mix predate an anticipated change in the industry 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). Our baseline specification is a linear model for the probability of firm  $j$  adding industry  $k$  between time  $t$  and  $t + 1$ :

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

Here,  $\text{Add}_{jt}^k$  is one if and only if firm  $j$  does not produce in industry  $k$  at time  $t$ , but does at time  $t + 1$ ;  $\alpha_{jt}$  is a firm  $\times$  time fixed effect which captures the average rate of adding industries for each firm-year, leaving the regression to identify only the *direction* of change in the industry mix and not changes in the number of industries that the firm operates in. We use the input similarity of firm  $j$  at time of the first observation (hence the superscript "0") to avoid endogeneity concerns that might arise from firms sourcing new inputs before they actually report the new outputs.<sup>13</sup>  $\alpha_k^t$  is an industry  $\times$  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 an industry-pair  $\times$  time fixed effect,  $\alpha_{kk'}^t$ , with an additional dimension of the firm's industry  $k'$  from which it derives the highest fraction

<sup>13</sup>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.



of revenue. These effects control for all shocks that might make all firms in industry  $k'$  more or less likely to start producing in industry  $k$ . Finally,  $\varepsilon_{jk}^t$  is an idiosyncratic error term at the firm-industry-time-level. 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 and second specification contain 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 control variables are also statistically significant (second column), but their inclusion does not change the estimated coefficient on input similarity by much. The third 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 fourth 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 the 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, the input similarity remains important even in this specification.

Our preferred specification is presented in column 3 of Table 2, which controls for annual rates of product adoption at the firm level in addition to annual supply and

TABLE 2. Industry Addition: Input Similarity and Vertical Relatedness

|   | Dependent variable: $Add_{jkt}$ |                       |                       |
|---|---------------------------------|-----------------------|-----------------------|
|   | (1)                             | (2)                   | (3)                   |
| InputSimilarity $^0_{jk}$                 | 0.0226**<br>(0.00021)           | 0.0222**<br>(0.00021) | 0.0163**<br>(0.00035) |
| Firm $\times$ Year FE $\alpha_{jt}$       | Yes                             | Yes                   | Yes                   |
| Industry $\times$ Year FE $\alpha_{kt}$   |                                 | Yes                   |                       |
| $k \times k' \times t$ FE $\alpha_{kk't}$ |                                 |                       | Yes                   |
| $R^2$                                     | 0.00834                         | 0.00972               | 0.0416                |
| Observations                              | 77745382                        | 77745382              | 77726154              |

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

<sup>+</sup>  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$

demand shocks that occur at the product level. Using the estimate from that specification, a one standard deviation increase in input similarity is associated with a 122% higher industry entry rate.

The results above constitute merely a set of correlations between firm characteristics and subsequent industry entry. To establish a causal channel, we now turn to exploiting a policy change that interacted with the firm's input mix to determine the direction of change in the industry mix.

**3.3. De-reservation of Products from Small-Scale Production.** Since the 1950s, India has given particular attention to the development of the small-scale industry (SSI) sector, which contributes almost 40% to gross industrial value-added and is the second largest employer after agriculture.<sup>14</sup> Starting in 1967, the government implemented a policy of reservation of certain products for exclusive manufacture by SSI firms. The stated aim of this policy was to ensure employment expansion, to achieve a more equitable distribution of income and “greater mobilization of private sector resources

<sup>14</sup>Development Commissioner, MSME, India (2018). Available at <http://dcmsme.gov.in/ssiindia/performance.htm#Employment>

of capital and skills” (Government of India, 2009). By the end of 1978, more than 800 products had been reserved; in 1996 it was more than a thousand.

By the early 1990s, the government realized that the reservation policy was inconsistent with the vast liberalization that had begun in the late 1980s and culminated in the new economic policy of 1991. According to the expert committee set up by the government to look into SSI policy, reservation did little to promote small enterprises and had negative consequences by keeping out large enterprises in these products. With free imports of most goods 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. De-reservation 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  
<http://www.dcmsme.gov.in/publications/reserveditems/revvex.htm>

The definition of small scale industries, and therefore the scope of reservation, changed over the time 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 de-reservation on output markets has been thoroughly studied in the literature. The consensus is that the de-reservation 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 de-reserved 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%. Notably, the firm’s response is heterogeneous: while small incumbents shrank, the larger ones expanded. 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 de-reservation 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 de-reserved markets. [Table 4](#) 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 de-reserved 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  $i$ . Unit values paid by firms using inputs from de-reserved markets drop by about eight to twelve percent upon de-reservation.<sup>15</sup> 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.

<sup>15</sup>The ASI unit value data is very noisy. We try to correct for known problems. In [Appendix B](#) we present results on subsamples that we believe to be particularly clean.

TABLE 4. Domestic input unit values after de-reservation

|                                   | Dependent variable: $\log p_{jit}$ |                      |
|-----------------------------------|------------------------------------|----------------------|
|                                   | (1)                                | (2)                  |
| $t \geq$ year $i$ was de-reserved | -0.128**<br>(0.014)                | -0.0864**<br>(0.015) |
| Year FE                           | Yes                                | Yes                  |
| Input Product FE                  | Yes                                |                      |
| Firm $\times$ Input Product FE    |                                    | Yes                  |
| $R^2$                             | 0.850                              | 0.955                |
| Observations                      | 957056                             | 547866               |

Standard errors in parentheses, clustered at the firm-year level.

+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$

**3.4. Input Similarity Weighted by De-reservation.** We use the de-reservation as an input-specific shock and weigh the input similarity measure according to de-reservation. We take the official lists of de-reserved items from the Ministry’s website and manually match them to 5-digit ASIC products. We then define  $\delta_{jit}$  to be one if and only if firm  $j$  at some point uses a five-digit in the three-digit category  $i$  that has been de-reserved during or before year  $t$ . We then interact the similarity measures by these de-reservation dummies as follows:

$$(3.2) \quad (\text{InputSimilarity-Dereservation})_{jkt}' = \sum_{i=1}^N \delta_{jit} \theta_{ij}' \bar{\theta}_{ki}$$

This measure ‘selects’ the portion of each input industries in the inner product that have been de-reserved.

We now study how the de-reservation interacts with firms’ idiosyncratic input mix in shaping their comparative advantage. We estimate the same specification as above (Equation 3.1, which explains firms’ additions to the industry mix), but with the input similarity weighted by de-reservation.

Table 5 shows the results. The estimated coefficient of the de-reservation-weighted input similarity coefficient is positive and statistically significant in all specifications:

when input  $i$  gets de-reserved, firms that have been using  $i$  intensively are more likely to add products that rely heavily on  $i$ . This holds both across industries (columns 1 to 3) and within industries (columns 4 and 5). Column 5 includes a tariff-change-weighted input similarity measure, analogous to the derservation-weighted input similarity.<sup>16</sup> When input  $i$  gets de-reserved or gets tariff reductions, firms that have been using  $i$  intensively are more likely to add products that rely heavily on  $i$ . Later, the structural estimation provides a tariff equivalent for de-reservation.

TABLE 5. Product Addition: The Impact of Dereservation

|  | Dependent variable: $\text{Add}_{jkt}$ |                       |                       |                       |
|--|--|-----------------------|-----------------------|-----------------------|
|  | (1)                                    | (2)                   | (3)                   | (4)                   |
| $\text{InputSimilarity}_{jk}^0$                | 0.0220**<br>(0.00021)                  | 0.0216**<br>(0.00021) | 0.0157**<br>(0.00035) | 0.0153**<br>(0.00035) |
| $\text{InputSimilarity-Dereservation}_{jkt}^0$ | 0.0227**<br>(0.0013)                   | 0.0228**<br>(0.0013)  | 0.0151**<br>(0.0013)  | 0.0143**<br>(0.0013)  |
| $\text{InputSimilarity-Tariff}_{jkt}^0$        |  |                       |                       | -0.0582**<br>(0.0054) |
| Firm $\times$ Year FE $\alpha_{jt}$            | Yes                                    | Yes                   | Yes                   | Yes                   |
| Industry $\times$ Year FE $\alpha_{kt}$        |  | Yes                   |                       |                       |
| $k \times k' \times t$ FE $\alpha_{kk't}$      |  |                       | Yes                   | Yes                   |
| $R^2$  | 0.00840                                | 0.00979               | 0.0417                | 0.0417                |
| Observations                                   | 77745382                               | 77745382              | 77726154              | 77726154              |

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

+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$

**3.5. Other controls.** Complementarities in the use of intermediate inputs might not be the only driver of co-production. Firms might also face demand-side complementarities, such that firms who produce one, or a certain set of industries, are able to obtain

<sup>16</sup>This is constructed by replacing the de-reservation indicator  $\delta_{jit}$  with the change in India's import tariffs  $\Delta\tau_{jit}$ . For the precise definition and data description, see Appendix C.2.

relatively higher prices on products from another industry.<sup>17</sup> To capture such complementarities, 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:

$$\text{outputSimilarity}_{jk}^t = \sum_{i=1}^N \sigma_{ji}^t \bar{\sigma}_{ki},$$

where  $i$  runs over the set of three-digit industries. The vector  $\sigma_j$  denotes the sales of firm  $j$  belonging to industry  $i$  at time  $t$ , divided by the total of  $j$ 's sales at time  $t$ . The vector  $\bar{\sigma}_k$  denotes the (size-weighted) average  $\sigma_{j'i}$  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 de-reservation dummies analogously to the input similarity measure in equation (3.2).

While our input and output similarity measures focus on similar distributions of expenditures or sales, other important directions firms' product lines might move is up and down their value chain, for which we next define firm-specific variables. We again use the aggregate input-output shares  $\bar{\theta}$  to measure whether a sector  $k$  is upstream or downstream from the firm's current product mix. Accordingly, we define:

$$(3.3) \quad \text{upstream}_{jk}^t = \sum_{i=1}^N \sigma_{ji}^t \bar{\theta}_{ik}, \quad \text{downstream}_{jk}^t = \sum_{i=1}^N \sigma_{ji}^t \bar{\theta}_{ki}.$$

To make sense of these definitions, consider the following analogy: imagine a firm  $j$  where what is observed is the sales shares of the firm,  $\sigma_j$ , and the goal is to predict the expenditure shares of the firm knowing only the national input-output table. Then given the firm's output mix and the industry's average expenditures for the outputs, one would expect the expenditure share of  $j$  on  $k$  to be  $\text{upstream}_{jk}^t$ . Likewise,  $\text{downstream}_{jk}^t$

---

<sup>17</sup>See [Brander and Eaton \(1984\)](#); [Shaked and Sutton \(1990\)](#); [Bernard et al. \(2018\)](#) for a discussion of demand complementarities in the multiproduct firm literature.

is the expected expenditure share of industry  $k$  on firms that feature the same product mix as  $j$ .

TABLE 6. Product Addition: Robustness

|   | Dependent variable: $\text{Add}_{jkt}$ |                         |                         |                         |                         |
|---|--|-------------------------|-------------------------|-------------------------|-------------------------|
|   | (1)                                    | (2)                     | (3)                     | (4)                     | (5)                     |
| InputSimilarity $^0_{jk}$                 | 0.0220**<br>(0.00021)                  | 0.0146**<br>(0.00027)   | 0.0142**<br>(0.00027)   | 0.0111**<br>(0.00035)   | 0.0107**<br>(0.00035)   |
| InputSimilarity-Dereservation $^0_{jkt}$  | 0.0227**<br>(0.0013)                   | 0.0212**<br>(0.0013)    | 0.0212**<br>(0.0013)    | 0.0128**<br>(0.0013)    | 0.0121**<br>(0.0013)    |
| OutputSimilarity $^0_{jk}$                |  | 0.00860**<br>(0.00039)  | 0.00852**<br>(0.00039)  | 0.0599**<br>(0.0011)    | 0.0599**<br>(0.0011)    |
| OutputSimilarity-Dereservation $^0_{jkt}$ |  | 0.0160**<br>(0.00086)   | 0.0156**<br>(0.00086)   | 0.00622**<br>(0.0012)   | 0.00623**<br>(0.0012)   |
| Upstream $^0_{jk}$                        |  | 0.0197**<br>(0.00055)   | 0.0186**<br>(0.00055)   | 0.0160**<br>(0.0017)    | 0.0160**<br>(0.0017)    |
| Downstream $^0_{jk}$                      |  | -0.00526**<br>(0.00033) | -0.00479**<br>(0.00033) | -0.00238**<br>(0.00083) | -0.00244**<br>(0.00083) |
| InputSimilarity-Tariff $^0_{jkt}$         |  |                         |                         |                         | -0.0549**<br>(0.0054)   |
| Firm $\times$ Year FE $\alpha_{jt}$       | Yes                                    | Yes                     | Yes                     | Yes                     | Yes                     |
| Industry $\times$ Year FE $\alpha_{kt}$   |  |                         | Yes                     |                         |                         |
| $k \times k' \times t$ FE $\alpha_{kk't}$ |  |                         |                         | Yes                     | Yes                     |
| $R^2$                                     | 0.00840                                | 0.00980                 | 0.0110                  | 0.0459                  | 0.0459                  |
| Observations                              | 77745382                               | 77745382                | 77745382                | 77726154                | 77726154                |

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

+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$

Table 15 shows the result of estimating equation (3.1) controlling for the output similarity variable, the de-reservation-weighted version of it, and for the two vertical relatedness measures. The estimated coefficient output similarity is positive and significant, in particular in the specifications with  $k \times k' \times t$  fixed effects. This is not entirely surprising, since output similarity encompasses within it the supply-side complementarities that we try to measure using input similarity. Firms are also slightly more



likely to move upstream from their product mix, and slightly less likely to move downstream. Most importantly, however, the estimated coefficients of input similarity and de-reservation-weighted input similarity remain positive and statistically significant.

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. We also show that results hold when focusing on (i) the set of large firms (100+ employees) that get 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. The results are robust to changing the estimator from OLS to Logit to better account for the discrete nature of the dependent variable.

**3.6. Case Study.** De-reservation 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 de-reservation.

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 has been difficulty in getting high quality and reliable supply of spices, which tend to be supplied by 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 (other than Spice Oil and Oleo-resin 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.

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 de-reserved 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 de-reserved 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 de-reservation. Therefore, the case study confirms the findings from the reduced form evidence.

The next Section more deeply investigates these reduced form findings by building a structural model to better understand these findings and quantify 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 based on idiosyncratic firm-industry productivities (firm comparative advantage). 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 idiosyncratic 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. But as a firm keeps expanding its product range, its acquired capabilities get stretched further and the return to comparative advantage declines, as in models of core competencies. Policy changes that increase 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.

This framework generates structural estimating equations that explain the portfolio of products a firm produces and the impact that policy changes have on observed portfolios. The key insight here is that unit costs across industries for multiproduct firms are interdependent through the relative demands a firm faces because capabilities are chosen to maximize total profits, not minimize costs in any single industry. We then use the theory to motivate an instrumental variable (IV) strategy based on common industry-time demand shifts in the economy to isolate model mechanisms. This uses the combination of demand shifts and the Input-Output table to derive a structural ‘Bartik’ instrument from theory. Finally, we 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. Production, Demand and Revenues.** Firm  $j$  can produce in multiple industries, indexed by  $k$ . 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  $\bar{\theta}_{ik}$  and idiosyncratic industry productivity labeled  $\varphi_{jk}$ .<sup>18</sup> At input prices  $S_{ijt}\psi_{it}$ , the unit cost of firm  $j$  to produce in industry  $k$  at time  $t$ , is therefore

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

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

**4.1.1. Producing Input Industry Aggregates.** Inputs  $M_{ijk}$  at the industry level are a composite of quantities  $m_{i\iota jkt}$  of varieties, indexed by  $\iota$ .  $M_{ijk}$  is the CES aggregator of varieties of input  $i$ :

$$M_{ijk}^{(\sigma-1)/\sigma} = \int_0^\infty m_{i\iota jkt}^{(\sigma-1)/\sigma} dh$$

where variety  $\iota$  of input  $i$  has a price  $s_{i\iota t}$  which follows a Pareto distribution with  $\Pr(s_{i\iota t} \geq s) = (s/s_m)^{-\Omega_{it}}$ . This naturally lends itself to the following interpretation: suppose a unit mass of suppliers sell at prices distributed  $\Pr(s_{i\iota} \geq s) = (s/s_m)^{-\lambda}$  and that for market  $i$  in period  $t$ , there is a mass  $N_{it}$  of suppliers. Then the distribution of the minimum price for each variety is Pareto with  $\Omega_{it} = \lambda N_{it}$ . A rise in the mass of input suppliers therefore increases the chances of getting lower price suppliers.

Firms have capabilities of using inputs with prices  $[\underline{c}_{ijt}, \infty)$  where  $\underline{c}_{ijt}$  is chosen by the firm. Here lower  $\underline{c}_{ijt}$  corresponds to both a greater variety of inputs and lower average prices. This can be interpreted as firms screening their input suppliers by choosing a

<sup>18</sup>In keeping with this section's focus on input capabilities,  $\varphi_{jk}$  could be modeled as  $\varphi_{jk} = \prod_i A_{ij}^{\bar{\theta}_{ik}}$ . However, we remain agnostic in keeping with the multiple channels of comparative advantage explored in the reduced form above.

lower cost cutoff for suppliers that they meet. Firms minimize costs to produce  $M_{ijkt}$  conditional on  $\underline{c}_{ijt}$ . In this setting, a firm's optimal choice of inputs can be summarized by the following Proposition (all proofs may be found in the Appendix):

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

(1) *The cost index for inputs from industry  $i$  for firm  $j$  at time  $t$  are*

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

(2) *Since  $d \ln S_{ijt} / d \ln \underline{c}_{ijt} = 1 + \Omega_{it} / (\sigma - 1)$ , 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*

$$c_{jkt} = \underbrace{\frac{1}{\varphi_{jk}}}_{\text{Idiosyncratic}} \underbrace{\prod_i \left( \psi_{it} \left( \frac{\Omega_{it}}{\Omega_{it} + (\sigma - 1)} \right)^{1/(1-\sigma)} \frac{S_m^{\Omega_{it}/(1-\sigma)}}{\bar{\theta}_{ik}} \right)^{\bar{\theta}_{ik}}}_{\text{Supplier}} \underbrace{\prod_i \left( \underline{c}_{ijt}^{1-\Omega_{it}/(1-\sigma)} \right)^{\bar{\theta}_{ik}}}_{\text{Capability}}.$$

4.1.2. *Endogenous Capabilities.* As derived above, unit costs for a given industry  $k$  are a function of input capabilities  $\underline{c}_{ijt}$  which we now endogenize. The key insight here is that unit costs across industries for multiproduct firms are interdependent on all the relative demands a firm faces because capabilities are chosen to maximize total profits, not minimize costs in any single industry. Thus this setting 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.

We assume that all firms have an innate capability for inputs from industry  $i$ ,  $\underline{c}_{i0}$ , and can adjust this capability (perhaps due to demand and supply conditions) subject to a Hicks neutral cost across production in all industries.<sup>19</sup> This can be interpreted as

<sup>19</sup>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.

scarce plant capacities being stretched towards improving some inputs and away from others. Letting  $\underline{c}_{jt}$  denote the vector of acquired capabilities, the actual unit costs of a multiproduct firm are given by  $\gamma(\underline{c}_{jt}) c_{jkt}$  in each industry, where

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

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  $\underline{c}_{i0} = 1$ .<sup>20</sup>

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

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

where  $p_{jkt}$  are prices,  $q_{jkt}$  are quantities and  $D_{kt}$  is an industry-time demand shifter. Then the profit function of firm  $j$  at time  $t$  across all industries  $k$  is

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

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

**Proposition 2.** *For firm-input expenditure shares  $\theta_{ijt}$ , the optimal capability choice is*

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

<sup>20</sup>This will not influence our estimating equations as it is an industry-time effect.

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

$$(4.1) \quad \ln R_{jkt} = \underbrace{\ln \frac{1-\rho}{\rho} \left( \rho^{\frac{\rho}{1-\rho}} D_{kt}^{\frac{1}{1-\rho}} \right)}_{\text{Demand (kt)}} - \underbrace{\frac{\rho}{1-\rho} \sum_i \bar{\theta}_{ik} \ln \psi_{it} (1 - \Theta_{it}^{-1})^{\frac{1}{1-\sigma}} \frac{s_m^{\Theta_{it}-1}}{\bar{\theta}_{ik}}}_{\text{Supplier (kt)}} + \underbrace{\frac{\rho}{1-\rho} \ln \varphi_{jk}}_{\text{RCA (jk)}} + \underbrace{\frac{\rho}{1-\rho} \sum_i \Theta_{it}^2 (\bar{\theta}_{ik} \theta_{ijt} - \theta_{ijt}^2/2)}_{\text{Firm Capability (jkt)}}$$

with the dimension of variation listed below each term.

The addition to Equation (4.1) of Firm Capability is beyond standard models and yields dynamic comparative advantage through capability adjustment. 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 Firm Capability term captures the dynamic re-deployment of capability across input productivities and is sensitive to the depth of input markets (through  $\Omega_{it}$  in  $\Theta_{it}$ ).<sup>21</sup>

<sup>21</sup>An alternative approach to inducing comparative advantage based on input linkages across industries could be through common firm-input productivities and production functions where input intensities change with a reduction in input prices. This is not however sufficient to generate input-based comparative advantage. For example, under CES production in both nests, higher firm-input productivities imply higher revenues (under love for variety) in industries using the input. But it does not imply that a firm with high input productivity has a *comparative* advantage in products using that input because comparative advantage depends on the *distributions of all* firm-input productivities. The underlying reasoning for this is similar to a many good-many factor comparative advantage model, for which Costinot (2009) shows that further “restrictions on the full distribution” of factor endowment ratios (firm-input productivities in our setting) is needed to get strong predictions akin to standard 2x2 comparative advantage models. Consequently, generating comparative advantage requires adding jointness to the firm problem, done here through capability choice. The framework can be extended to CES production, but this does not provide many more testable insights because input price and quantity data are needed to separately identify firm-input productivities from the elasticity of substitution under CES production. Further, identifying firm-input productivities would require restricting

Economies of scope arise in this model because firms can use their acquired capabilities across industries. The returns to acquired capabilities however decreases 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. This endogenizes 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 (defined as the industry in which the firm has the highest  $\varphi_{jk}$ ).

**4.2. Estimating Policy Effects.** Now consider an observable policy  $P$  that changes the depth of input markets of the form  $\Omega_{it} = \Omega_{i0} + \alpha_P P_{it}$ . Linearizing Equation (D.1) around the initial policy state  $\Omega_{i0}$  and letting  $\kappa_x$  represent a fixed effect for characteristic  $x$  yields the following estimating equation:

$$(4.2) \quad \ln R_{jkt} = \kappa_{kt} + \kappa_{jk} + \underbrace{\frac{\rho}{1-\rho} \sum_i \left[ \Theta_{i0}^2 + \frac{2\Theta_{i0}}{(\sigma-1)} \alpha_P (P_{it} - P_{i0}) \right]}_{\text{Firm Capability Change (jkt)}} \left( \bar{\theta}_{ik} \theta_{ijt} - \frac{\theta_{ijt}^2}{2} \right).$$

The theory above signs  $\Theta_{it}$  as the same sign as  $\sigma - 1$ , so estimating  $\alpha_P \cdot 2\Theta_{i0}/(\sigma - 1)$  gives the same sign as  $\alpha_P$  and allows for testing hypotheses about  $\alpha_P$ .

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

---

estimation to multiproduct firms which produce in all industries, which is to say zero observations. We therefore work with a simpler production function - Cobb Douglas technology across inputs with a nested CES across input varieties.



For ease of estimation, we will impose  $\Omega_{i0} = \bar{\Omega}$ , so that

$$\Omega_{it} = \bar{\Omega} + \alpha_B B_{it} + \alpha_\tau (1 + \tau_{it}).$$

In light of the theory above, we can interpret these policy shifts as changing the depth of input markets with theory signing both  $\alpha_B$  and  $\alpha_\tau$  to be negative, so that with no entry barriers and zero tariffs,  $\Omega_{i0} = \bar{\Omega}$  is the ‘maximal’ market depth. Therefore Equation (4.2) approximates around a policy space of no entry barriers and no tariffs. This then implies the estimating equation

$$(4.3) \quad \ln R_{jkt} = \kappa_0 \sum_i \left( \bar{\theta}_{ik} \theta_{ijt} - \frac{\theta_{ijt}^2}{2} \right) + \kappa_1 \sum_i (\alpha_B B_{it} + \alpha_\tau \tau_{it}) \left( \bar{\theta}_{ik} \theta_{ijt} - \frac{\theta_{ijt}^2}{2} \right) \\ + \kappa_{kt} + \kappa_{jk}.$$

with  $\kappa_0 = \Theta_{i0}^2 \rho / (1 - \rho)$ ,  $\kappa_1 = 2\Theta_{i0} \rho / (1 - \rho) (\sigma - 1)$ . 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 (4.3). Firms will produce in industry  $k$  exactly when  $R_{jkt} > (1 - \rho) f_{kt}$ , so we estimate Equation (4.3) as a linear probability model for the outcome that observed revenues of the firm-industry are positive each period.<sup>22</sup> As we are estimating probabilities, we can think of how comparative advantage shifts the *production probability frontier* of firms.

**4.3. Structural Instrumentation.** In Equation (4.2), firm expenditure shares  $\theta_{ijt}$  are a function of fixed technology  $\bar{\theta}_{ik}$ , time varying input prices  $\psi_{it}$ , demand shocks  $D_{kt}$  and idiosyncratic productivities  $\varphi_{jk}$ . Input price and demand shocks are estimated through industry-time fixed effects. Idiosyncratic productivities are estimated through firm-industry fixed effects, expressed as Revealed Comparative Advantage. Technology

---

<sup>22</sup>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.

is estimated with a large number of observations, so the risk of measurement error contaminating  $\bar{\theta}_{ik}$  is small, and similarly for demand and input shocks.

One potential concern is that dereservation systematically changes technology  $\bar{\theta}_{ik}$ , 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  $\bar{\theta}_{ik}$  however have a mean of -.009 with a standard deviation of .017, which is to say about zero in significance and magnitude.<sup>23</sup>

There might be omitted variables from our structural equation that cause  $\theta_{ijt}$  to change, which could bias our estimates of the role of capabilities. For example, demand or cost shocks at 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:

- Demand shocks  $D_{jkt}$  at the firm level would be positively correlated with input similarity through the composition of firm activity.
- Input price shocks  $\psi_{ijkt}$  at the firm level would be negatively correlated with input similarity through the composition of firm activity away from industries intensive in using input  $i$  (high  $\bar{\theta}_{ik}$ ).

We therefore propose a novel instrument based on our structural equations. 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  $\theta_{ijt}$  based on shifts in the

---

<sup>23</sup>Since the percentage of reserved inputs is generally much less than 100%, the implied changes are negligible. See Figure B.1 of the Appendix for the histogram of estimated coefficients.

within firm distribution of activity.<sup>24</sup> In fact, examining the estimating Equation (4.2), what is needed is not instruments for each  $\theta_{ijt}$  rather an instrument for terms of the form  $\sum_i (\bar{\theta}_{ik}\theta_{ijt} - \theta_{ijt}^2/2)$  and  $\sum_i (P_{it} - P_{i0}) (\bar{\theta}_{ik}\theta_{ijt} - \theta_{ijt}^2/2)$ . Changes in these terms can be approximated, holding capabilities constant, as summarized in the following Proposition:

**Proposition 3.** *An input similarity approximation for an instrumental variable first stage regression, holding capabilities constant based on demand shocks is*

$$\sum_i (\bar{\theta}_{ik}\theta_{ijt} - \theta_{ijt}^2/2) \approx \lambda \sum_i (\bar{\theta}_{ik}\theta_{ijt-1} - \theta_{ijt-1}^2/2) + \gamma_{kt} \sum_i \chi_{jkt-1} (\bar{\theta}_{ik} - \theta_{ijt-1})^2$$

where  $\chi_{jkt}$  are firm revenue shares for a firm in year  $t$ . The coefficients are as follows:

- $\lambda$  should equal one,
- $\gamma_{kt}$  is a demand innovation term  $(D_{kt}/D_{kt-1} - 1) / (1 - \rho)$ .

The Proposition above 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 input similarity 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. In the case of a single instrument for terms of the form  $\bar{\theta}_{ik}\theta_{ijt} - \theta_{ijt}^2/2$ , the first stage of an IV strategy following

<sup>24</sup>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 ( $\Omega_{it} = \Omega$ ), the expression becomes

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

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

from the Proposition is then:

$$(4.4) \quad \sum_i \left( \bar{\theta}_{ik} \theta_{ijt} - \frac{\theta_{ijt}^2}{2} \right) = \lambda \sum_i \left( \bar{\theta}_{ik} \theta_{ijt-1} - \frac{\theta_{ijt-1}^2}{2} \right) + \gamma_{kt} \sum_i \chi_{jkt-1} (\bar{\theta}_{ik} - \theta_{ijt-1})^2 \\ + \kappa_{kt} + \kappa_{jk}.$$

Equation (4.4) is composed of three parts: the fixed effects found in the main structural equation for revenues, a lagged term for the endogenous sum  $\sum_i (\bar{\theta}_{ik} \theta_{ijt-1} - \theta_{ijt-1}^2/2)$ , and linear adjustment based on predicted input share changes from lagged revenue shares  $\chi_{jkt-1}$  and contemporaneous industry level demand shocks  $\gamma_{kt}$ . This last term is essentially a (lagged) sales weighted ‘technological distance’ measure of the firm away from an industry  $k$  times the magnitude of the demand innovation which predicts the change in  $\sum_i (\bar{\theta}_{ik} \theta_{ijt} - \theta_{ijt}^2/2)$  between periods.

However, as we need to instrument for both changes in input shares and these input shares interacted with two policy changes, we need three instruments of the type in Equation (4.4), 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.<sup>25</sup> Accordingly, define both  $\tilde{\theta}_{ijkt} \equiv \bar{\theta}_{ik} \theta_{ijt} - \theta_{ijt}^2/2$  and  $\tilde{\chi}_{jkt} \equiv \chi_{jkt} (\bar{\theta}_{ik} - \theta_{ijt})^2$  and the following sums for  $\lambda$  and the KxT vector  $\gamma$ :

$$I_{jkt}(\lambda, \gamma) \equiv \lambda \sum_i \tilde{\theta}_{ijkt-1} + \gamma_{kt} \sum_i \tilde{\chi}_{jkt-1}, \\ I_{jkt}^B(\lambda, \gamma) \equiv \lambda \sum_i B_{it} \tilde{\theta}_{ijkt-1} + \gamma_{kt} \sum_i B_{it} \tilde{\chi}_{jkt-1}, \\ I_{jkt}^\tau(\lambda, \gamma) \equiv \lambda \sum_i \tau_{it} \tilde{\theta}_{ijkt-1} + \gamma_{kt} \sum_i \tau_{it} \tilde{\chi}_{jkt-1}.$$

<sup>25</sup>The underlying assumption here is no serial correlation in idiosyncratic demand and supply shocks. If this is thought to hold, longer lags can be taken to decrease any potential bias, at the cost of observations.

The resulting first stage equations for our estimator are as follows:<sup>26</sup>

$$(4.5) \quad \sum_i \tilde{\theta}_{ijkt} = \kappa_{kt} + \kappa_{jk} + I_{jkt}(\lambda^{11}, \gamma^{11}) + I_{jkt}^B(\lambda^{12}, \gamma^{12}) + I_{jkt}^\tau(\lambda^{13}, \gamma^{13}) + \eta_{jkt}$$

$$(4.6) \quad \sum_i B_{it} \tilde{\theta}_{ijkt} = \kappa_{kt} + \kappa_{jk} + I_{jkt}(\lambda^{21}, \gamma^{21}) + I_{jkt}^B(\lambda^{22}, \gamma^{22}) + I_{jkt}^\tau(\lambda^{23}, \gamma^{23}) + \eta_{jkt}^B$$

$$(4.7) \quad \sum_i \tau_{it} \tilde{\theta}_{ijkt} = \kappa_{kt} + \kappa_{jk} + I_{jkt}(\lambda^{31}, \gamma^{31}) + I_{jkt}^B(\lambda^{32}, \gamma^{32}) + I_{jkt}^\tau(\lambda^{33}, \gamma^{33}) + \eta_{jkt}^\tau$$

We implement the instrumental variable estimator of the structural coefficients in Equation (4.3) as a manual 2SLS estimator, which allows us to calculate the fitted values of the first stage without having to recover the high number of demand innovation coefficients  $\gamma_{kt}$  of the instruments in (4.5-4.7) 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.

**4.4. Results and the Economic Relevance of Input Capabilities.** Table 7 shows the OLS and IV estimates for the extensive margin version of Equation (4.3).<sup>27</sup> The estimated coefficient on the deviation of the input similarity measure is  $\kappa_0 = .009$  in the OLS, which rises to .14 in the IV. The policy coefficient of interest for the entry barriers is  $\kappa_1 \alpha_B = -.0004$  in the OLS which increases in magnitude to  $-.002$  in the IV. Comparing this with the coefficient on tariffs interacted with the input similarity deviation,  $\kappa_1 \alpha_\tau = -.017$ , the effect of entry barriers is a tenth of this. The tariff equivalent of dereservation is then  $\alpha_B / \alpha_\tau = .0168 / .0016 = 10.5$ . Entry barriers from

<sup>26</sup>In practice, sales within a firm-industry group are unlikely to be a balanced panel as the extensive margin of a firm's industries is liable to change (we in fact model and estimate this with a logit model). Consequently, our one period lag strategy may lose some observations but it reduces the number of parameters that must be estimated simultaneously

<sup>27</sup>Relevant summary statistics are in Table 14 of the Appendix.

TABLE 7. Structural Estimates for Multiproduct Sales Premium

|   | Positive Sales for Plant $j$ in Industry $k$ ( $R_{jkt} > 0$ ) |                        |                        |                        |
|---|--|------------------------|------------------------|------------------------|
|   | (1)  | (2)                    | (3)                    | (4)                    |
| $\sum_i (\bar{\theta}_{ik}\theta_{ijt} - \theta_{ijt}^2/2)$                 | 0.0086***<br>(0.0002)  | 0.0086***<br>(0.0002)  | 0.1362***<br>(0.0229)  | 0.1630**<br>(0.0226)   |
| $\sum_i B_{it} \cdot (\bar{\theta}_{ik}\theta_{ijt} - \theta_{ijt}^2/2)$    | -0.0004***<br>(0.0001)   | -0.0004***<br>(0.0001) | -0.0016***<br>(0.0004) | -0.0016***<br>(0.0004) |
| $\sum_i \tau_{it} \cdot (\bar{\theta}_{ik}\theta_{ijt} - \theta_{ijt}^2/2)$ |  | -0.0005<br>(0.0003)    |                        | -0.0168***<br>(0.0027) |
| $\kappa_{jk}$   | Yes  | Yes                    | Yes                    | Yes                    |
| $\kappa_{kt}$   | 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 plant-industry level.

+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$

reservation of inputs for small scale firms therefore lower industry adoption, and their estimated effect is equivalent to a 10.5 percentage tariff on inputs.

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 as:

$$CA_{jkt} \equiv \hat{\kappa}_0 \sum_i \bar{\theta}_{ik}\theta_{ijt} + \hat{\kappa}_1 \sum_i (\hat{\alpha}_B B_{it} + \hat{\alpha}_\tau \tau_{it}) \bar{\theta}_{ik}\theta_{ijt},$$

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

Table 8 shows summary statistics of  $CA_{jkt}$  for firms that produce in industry  $k$ . On average across firms and industries,  $CA_{jkt}$  increases the production probability by 4.3 percent, and for more than 13 percent in the top tenth percentile. On average,  $CA_{jkt}$  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,  $CA_{jkt}$  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 8. Comparative advantage  $CA_{jkt}$ , by industry 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  $CA_{jkt}$  for firms that do not produce in industry  $k$ . The expression then has the interpretation as the additional probability that firm  $j$  would produce in  $k$  by virtue of their input capabilities, holding fixed their capability choice. Naturally,  $CA_{jkt}$  is close to zero for the vast majority of triples  $(j, k, t)$  – after all, the space of inputs is large and many industries will not have inputs in common with the firm. But for certain firm-industry combinations, as suggested by Figure 1.1, the  $CA_{jkt}$  term is economically significant. Table 9 contrasts, for three different industries  $k$ , the average premium  $\overline{CA_{jkt}}$  among single-industry firms of two different industries that may be co-producing  $k$ . Single-industry firms in the Edible fruits and nuts/edible vegetables industry (code 121) on average enjoy a comparative advantage  $CA_{jkt}$  in the

Fruit and vegetable juices industry (135) of 8.5%, 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 0.6% 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  $CA_{jkt}$  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 6.8%, whereas the Plastic footwear industry’s premium is only 0.4%. Table 20 in Appendix E states the average  $CA_{jkt}$  for the industry  $k$  with the highest premium for 25 industries. Hence, 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 9. Average firm-level comparative advantage: Some examples

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

Note: The table shows the average firm-level comparative advantage  $\overline{CA}_{kk'}$  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 10 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



TABLE 10. Core Competency Sales Premium (%) from 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 |

contributes 5.2 percent to the production probability. Firms that produce in two industries experience a 6 percent premium on their core industry and about half of that, 2.9 per cent, 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 10 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 11 contain the size-weighted comparative advantage 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 5.5 per cent, compared to the typical single-industry firm. Firms in multiple industries now show large premia even when we move along

TABLE 11. Core Competency Sales Premium (Size) from Comparative Advantage

| 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 |

the rows of core industries for firms that operate in more and more industries. For example, a firm operating in nine industries has a 46.8% higher (size weighted) premium in its core industry compared to a 7.2% core premium for a two-industry firm. Moving down the columns, firms see larger premia on their core products, compared to their peripheral products. The lowest ranked industries of a firm show small premia, of under 1 per cent (compared to 5.5% for single-industry plants).

Tables 10 and 11 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. CONCLUSION

Even though multi-product and multi-industry firms account for a disproportionately large share of economic activity, the economics literature is thin regarding formal theories predicting the determinants of co-production within firms, often arguing that firms perform similar activities. In this paper we provide a theory of similarity in the product space through common use of firm-specific input capabilities that can be shared across product lines. 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 are an important determinant of firm-level comparative advantage and help explain the content of a firm's 'core competencies' through comparative advantage arising from input capability.

A key theoretical insight of our framework is that economies of scope within multi-product 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 its extent of input similarity with each industry. The theory allows us to derive an instrumental variable strategy that, when implemented, shows that input capabilities are quantitatively important in determining the production patterns of firms.

In a wider view, 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.

- 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.
- 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 (2019): “The origins of firm heterogeneity: A production network approach,” Tech. rep., National Bureau of Economic Research.
- 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.
- 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.
- 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.
- 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.
- COMBES, P.-P. AND L. GOBILLON (2014): “The Empirics of Agglomeration Economies,” *Working Paper*.
- CONLEY, T. G. AND B. DUPOR (2003): “A spatial analysis of sectoral complementarity,” *Journal of political Economy*, 111, 311–352.
- COSTINOT, A. (2009): “An elementary theory of comparative advantage,” *Econometrica*, 77, 1165–1192.
- DHINGRA, S. (2013): “Trading away wide brands for cheap brands,” *The American Economic Review*, 103, 2554–2584.
- DI GIOVANNI, J., A. A. LEVCHENKO, AND I. MÉJEAN (2014): “Firms, destinations, and aggregate fluctuations,” *Econometrica*, 82, 1303–1340.
- 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., L. IACOVONE, B. S. JAVORCIK, AND P. NEARY (2009): "Multi-product Firms at Home and Away," Tech. rep., Mimeo.
- ECKEL, C. AND P. NEARY (2010): "Multi-product Firms and Flexible Manufacturing in the Global Economy," *Review of Economic Studies*, 77, 188–217.
- 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*.
- 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*.
- 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. E. AND A. RODRIGUEZ-CLARE (2009): "Trade, Foreign Investment, and Industrial Policy for Developing Countries," *NBER Working Papers*.

- 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.
- 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.
- JOVANOVIĆ, B. (1987): “Micro shocks and aggregate risk,” *The Quarterly Journal of Economics*, 395–409.
- 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.

- 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*.
- MARRIS, R. (1964): *The economic theory of managerial capitalism*, vol. 258, Macmillan London.
- 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 (2009): "Market size, Competition, and the Product Mix of Exporters," *Working Paper*.
- (2014): "Market size, competition, and the product mix of exporters," *American Economic Review*, 104, 495–536.
- 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.
- ROBINS, J. AND M. F. WIERSEMA (1995): "A Resource-based Approach to the Multi-business 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. (1982a): “Inter-Industry Technology Flows and Productivity Growth,” *The Review of Economics and Statistics*, 64, pp. 627–634.
- (1982b): “Inter-industry technology flows and productivity growth,” *The review of economics and statistics*, 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.
- SUTTON, J. (2012): *Competing in capabilities: the globalization process*, Oxford University Press.
- SUTTON, J. AND D. TREFLER (2016): “Capabilities, wealth, and trade,” *Journal of Political Economy*, 124, 826–878.
- TEECE, D. J. (1980): “Economies of scope and the scope of the enterprise,” *Journal of economic behavior & organization*, 1, 223–247.
- (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.
- VANDEBUSSCHE, H. AND C. VIEGELAHN (2014): “Trade protection and input switching: Firm-level evidence from Indian importers,” *Working Paper*.

## APPENDIX A. DESCRIPTIVE STATISTICS

### APPENDIX B. ROBUSTNESS OF ESTIMATES AND FURTHER RESULTS

**B.1. Robustness of Industry Add Regressions.** Table 15 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

TABLE 12. Summary Statistics

|  | count      | mean    | sd   | min   | max |
|--|------------|---------|------|-------|-----|
| Industry Add Dummy   | 77,745,382 | 0.0007  | 0.03 | 0.00  | 1   |
| InputSimilarity <sup>0</sup> <sub>jk</sub>                 | 77,745,382 | 0.0128  | 0.06 | 0.00  | 1   |
| InputSimilarity-Dereservation <sup>0</sup> <sub>jkt</sub>  | 77,745,382 | 0.0006  | 0.01 | 0.00  | 1   |
| InputSimilarity-Tariff <sup>0</sup> <sub>jkt</sub>         | 77,745,382 | -0.0001 | 0.00 | -0.33 | 0   |
| OutputSimilarity <sup>0</sup> <sub>jk</sub>                | 77,745,382 | 0.0044  | 0.05 | 0.00  | 1   |
| OutputSimilarity-Dereservation <sup>0</sup> <sub>jkt</sub> | 77,745,382 | 0.0007  | 0.02 | 0.00  | 1   |
| Upstream <sup>0</sup> <sub>jk</sub>                        | 77,745,382 | 0.0040  | 0.04 | 0.00  | 1   |
| Downstream <sup>0</sup> <sub>jk</sub>                      | 77,745,382 | 0.0067  | 0.04 | 0.00  | 1   |

TABLE 13. Correlation of Similarity Indexes

|                                  | IS <sup>0</sup> <sub>jk</sub> | OS <sup>0</sup> <sub>jk</sub> | IS-DR <sup>0</sup> <sub>jk</sub> | OS-DR <sup>0</sup> <sub>jk</sub> | Up <sup>0</sup> <sub>jk</sub> | Down <sup>0</sup> <sub>jk</sub> |
|----------------------------------|-------------------------------|-------------------------------|----------------------------------|----------------------------------|-------------------------------|---------------------------------|
| IS <sup>0</sup> <sub>jk</sub>    | 1.00                          |                               |                                  |                                  |                               |                                 |
| OS <sup>0</sup> <sub>jk</sub>    | 0.38                          | 1.00                          |                                  |                                  |                               |                                 |
| IS-DR <sup>0</sup> <sub>jk</sub> | 0.17                          | 0.06                          | 1.00                             |                                  |                               |                                 |
| OS-DR <sup>0</sup> <sub>jk</sub> | 0.09                          | 0.40                          | 0.09                             | 1.00                             |                               |                                 |
| Up <sup>0</sup> <sub>jk</sub>    | 0.46                          | 0.54                          | 0.05                             | 0.10                             | 1.00                          |                                 |
| Down <sup>0</sup> <sub>jk</sub>  | 0.55                          | 0.45                          | 0.06                             | 0.08                             | 0.50                          | 1.00                            |

TABLE 14. Structural Summary Statistics

|   | count      | mean      | sd       | min   | max  |
|---|------------|-----------|----------|-------|------|
| Industry Production Dummy   | 77,745,382 | .0059156  | .076685  | 0     | 1    |
| $\sum_i (\bar{\theta}_{ik}\theta_{ijt} - \theta_{ijt}^2/2)$                     | 77,745,382 | -.3585109 | .1483231 | -.5   | .5   |
| $\sum_i B_{it} \cdot (\bar{\theta}_{ik}\theta_{ijt} - \theta_{ijt}^2/2)$        | 77,745,382 | -.0141903 | .0682496 | -.5   | .485 |
| $\sum_i \tau_{it} \cdot (\bar{\theta}_{ik}\theta_{ijt} - \theta_{ijt}^2/2)$     | 77,745,382 | -.0066979 | .02447   | -1.05 | .480 |
| $\sum_i (\bar{\theta}_{ik}\theta_{ijt-1} - \theta_{ijt-1}^2/2)$                 | 46,185,150 | -.3374241 | .1689226 | -1.46 | .5   |
| $\sum_i B_{it} \cdot (\bar{\theta}_{ik}\theta_{ijt-1} - \theta_{ijt-1}^2/2)$    | 46,185,150 | -.0129705 | .0658083 | -1    | .483 |
| $\sum_i \tau_{it} \cdot (\bar{\theta}_{ik}\theta_{ijt-1} - \theta_{ijt-1}^2/2)$ | 46,185,150 | -.006114  | .0233025 | -.91  | .480 |
| $\sum_i \chi_{jkt-1} (\bar{\theta}_{ik} - \theta_{ijt-1})^2$                    | 46,185,150 | .0000208  | .0014629 | 0     | .719 |
| $\sum_i B_{it} \cdot \chi_{jkt-1} (\bar{\theta}_{ik} - \theta_{ijt-1})^2$       | 46,185,150 | .0011376  | .0240802 | 0     | 2.52 |
| $\sum_i \tau_{it} \cdot \chi_{jkt-1} (\bar{\theta}_{ik} - \theta_{ijt-1})^2$    | 46,185,150 | .0000592  | .0051217 | 0     | 1.05 |

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).

TABLE 15. Revealed comparative advantage – Robustness

|   | Dependent variable: $Add_{jkt}$ |                         |                                   |                          |
|---|---------------------------------|-------------------------|-----------------------------------|--------------------------|
|   | (1)                             | (2)                     | (3)                               | (4)                      |
| InputSimilarity $^0_{jk}$                 | 0.0111**<br>(0.00035)           | 0.0108**<br>(0.00037)   | 0.0188**<br>(0.00067)             | 0.0204**<br>(0.00068)    |
| InputSimilarity-Dereservation $^0_{jkt}$  | 0.0128**<br>(0.0013)            | 0.0114**<br>(0.0014)    | 0.00959**<br>(0.0019)             | 0.0170**<br>(0.0024)     |
| OutputSimilarity $^0_{jk}$                | 0.0599**<br>(0.0011)            | 0.0550**<br>(0.0012)    | 0.0873**<br>(0.0019)              | 0.0519**<br>(0.0011)     |
| OutputSimilarity-Dereservation $^0_{jkt}$ | 0.00622**<br>(0.0012)           | 0.00630**<br>(0.0013)   | 0.00844**<br>(0.0021)             | 0.00480**<br>(0.0011)    |
| Upstream $^0_{jk}$                        | 0.0160**<br>(0.0017)            | 0.0112**<br>(0.0018)    | 0.0263**<br>(0.0030)              | 0.00995**<br>(0.0020)    |
| Downstream $^0_{jk}$                      | -0.00238**<br>(0.00083)         | -0.00304**<br>(0.00088) | -0.00254 <sup>+</sup><br>(0.0014) | -0.00916**<br>(0.0017)   |
| Sample                                    | Full                            | Single-plant firms      | Census plants                     | Co-production industries |
| Firm $\times$ Year FE                     | Yes                             | Yes                     | Yes                               | Yes                      |
| $k \times k' \times t$ FE                 | Yes                             | Yes                     | Yes                               | Yes                      |
| $R^2$                                     | 0.0459                          | 0.0454                  | 0.0740                            | 0.0827                   |
| Observations                              | 77726154                        | 65110309                | 33544764                          | 8677381                  |

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

<sup>+</sup>  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$

**B.2. Robustness of Industry Add Regressions to Logit.** Table 16 shows the results of the logit estimation of the industry addition regressions, corresponding to the baseline specifications of Table 5.

**B.3. Robustness of Unit Value Regressions.** The unit values in the ASI are very noisy. 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. We try to correct for this problem by appropriately

TABLE 16. Revealed Comparative Advantage – Robustness

|                           | Dependent variable: $Add_{jkt}$ |                         |                        |                          |
|---------------------------|---------------------------------|-------------------------|------------------------|--------------------------|
|                           | (1)                             | (2)                     | (3)                    | (4)                      |
| $IS_{jkt}^0$              | 6.60053***<br>(0.03198)         | 8.21798***<br>(0.04748) | 5.1412***<br>(0.07137) | 5.10027***<br>(0.07209)  |
| $ISDR_{jkt}^0$            | 2.14728***<br>(0.17232)         | 2.44737***<br>(0.19811) | 1.4510***<br>(0.24678) | 1.38364***<br>(0.24712)  |
| $ISDTariff_{jkt}^0$       |                                 |                         |                        | -3.72533***<br>(0.89856) |
| Firm $\times$ Year FE     | Yes                             | Yes                     | Yes                    | Yes                      |
| $k \times t$ FE           |                                 | Yes                     |                        |                          |
| $k \times k' \times t$ FE |                                 |                         | Yes                    | Yes                      |
| Estimator                 | Logit (ML)                      | Logit (ML)              | Logit (ML)             | Logit (ML)               |
| Observations              | 77111718                        | 77111718                | 77111718               | 77111718                 |

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

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

scaling unit values if they fall outside a particular interval (in log terms) from the median. The reported unit values are those after this correction. In columns 3 and 4 of Table 17 we also 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).

**B.4. Industry Drop and Intensive Margin (Sales) Regressions.** Tables 18 shows how the probability to drop an industry from the industry mix is shaped by input similarity. Table 19 shows how log sales are correlated with input similarity.

**B.5. Estimated Technology Changes from Dereservation.**

## APPENDIX C. DATA APPENDIX

### C.1. Data sources.

TABLE 17. Domestic input unit values after dereservation – Robustness

|                                   | Dependent variable: $\log p_{jit}$ |                      |                      |                      |
|-----------------------------------|------------------------------------|----------------------|----------------------|----------------------|
|                                   | (1)                                | (2)                  | (3)                  | (4)                  |
| $t \geq$ year $i$ was de-reserved | -0.128**<br>(0.014)                | -0.0864**<br>(0.015) | -0.0477**<br>(0.012) | -0.0635**<br>(0.014) |
| Sample                            | 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.

+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$

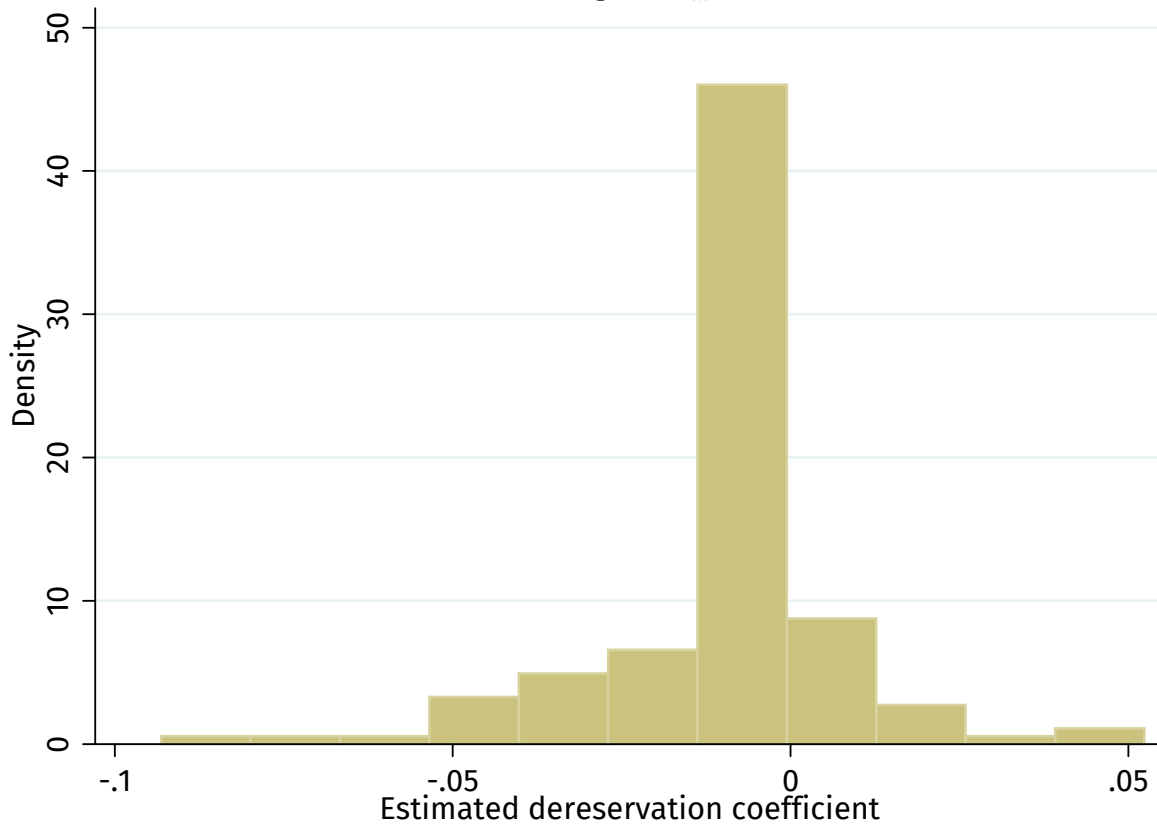
FIGURE B.1. Estimated Changes in  $\bar{\theta}_{ik}$  from Dereservation

TABLE 18. Industry Drop Regressions:

|                                    | Dependent variable: Drop <sub>ikt</sub> |                                 |                                 |
|------------------------------------|---|---------------------------------|---------------------------------|
|                                    | (1)                                     | (2)                             | (3)                             |
| IS <sub>ikt</sub> <sup>0</sup>     | -0.00940 <sup>+</sup><br>(0.0050)       | -0.112**<br>(0.0068)            | -0.0839**<br>(0.0076)           |
| ISDR <sub>ikt</sub> <sup>0</sup>   | -0.185**<br>(0.028)                     | -0.0541 <sup>+</sup><br>(0.029) | -0.0842*<br>(0.034)             |
| OS <sub>ikt</sub> <sup>0</sup>     | -0.185**<br>(0.0034)                    | -0.170**<br>(0.0040)            | -0.136**<br>(0.0058)            |
| OSDR <sub>ikt</sub> <sup>0</sup>   | -0.0534**<br>(0.0060)                   | -0.0462**<br>(0.0068)           | -0.0459**<br>(0.0080)           |
| UP0                                | -0.0273**<br>(0.0061)                   | -0.0556**<br>(0.010)            | -0.0360 <sup>+</sup><br>(0.020) |
| DOWN0                              | 0.0993**<br>(0.0096)                    | 0.0246*<br>(0.011)              | 0.0391 <sup>+</sup><br>(0.021)  |
| Firm × Year FE α <sub>jt</sub>     | Yes                                     | Yes                             | Yes                             |
| Industry × Year FE α <sub>kt</sub> |   | Yes                             |                                 |
| k × k' × t FE α <sub>kk't</sub>    |   |                                 | Yes                             |
| R <sup>2</sup>                     | 0.573                                   | 0.601                           | 0.669                           |
| Observations                       | 251028                                  | 250963                          | 220611                          |

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

<sup>+</sup>  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$

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](mailto: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:

TABLE 19. Intensive Margin Regressions:

|   | Dependent variable: log Sales <sub>jkt</sub> |                     |                     |
|---|--|---------------------|---------------------|
|   | (1)  | (2)                 | (3)                 |
| IS <sub>jkt</sub> <sup>0</sup>            | 0.451**<br>(0.037)                           | 0.799**<br>(0.048)  | 0.466**<br>(0.047)  |
| ISDR <sub>jkt</sub> <sup>0</sup>          | 0.538**<br>(0.19)                            | 0.145<br>(0.19)     | 0.392*<br>(0.18)    |
| OS <sub>jkt</sub> <sup>0</sup>            | 3.821**<br>(0.023)                           | 3.326**<br>(0.027)  | 1.414**<br>(0.029)  |
| OSDR <sub>jkt</sub> <sup>0</sup>          | -0.341**<br>(0.039)                          | -0.497**<br>(0.044) | -0.239**<br>(0.041) |
| UP0                                       | -1.279**<br>(0.045)                          | 0.0798<br>(0.070)   | 0.304**<br>(0.10)   |
| DOWN0                                     | 1.876**<br>(0.075)                           | 0.526**<br>(0.081)  | 0.134<br>(0.11)     |
| Firm × Year FE $\alpha_{jt}$              | Yes  | Yes                 | Yes                 |
| Industry × Year FE $\alpha_{kt}$          |  | Yes                 |                     |
| $k \times k' \times t$ FE $\alpha_{kk't}$ |  |                     | Yes                 |
| $R^2$                                     | 0.804  | 0.833               | 0.911               |
| Observations                              | 251028                                       | 250963              | 220611              |

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

+  $p < 0.10$ , \*  $p < 0.05$ , \*\*  $p < 0.01$

<http://www.dcmsme.gov.in/publications/reserveditems/resvex.htm> (accessed December 2014). 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<sub>jkt</sub> : 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.

- *Drop dummies*  $\text{Drop}_{jkt}$  : 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*  $\text{Sales}_{jkt}$  :  $j$ 's total sales of products in 3-digit industry  $k$  at time  $t$ .
- *Plant expenditure shares*  $\theta_{ijt}$  : 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 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.
- *Plant sales shares*  $\chi_{jkt}$  : 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  $t$ .
- *Aggregate sales shares*  $\bar{\chi}_{ik}$  : 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.
- *Dereservation dummy*  $\delta_{ijt}$  : 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  $B_{it}$  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*  $\Delta\tau_{jit}$  : Difference between year  $t$  Indian import tariff and year 2000 tariff on 5-digit products in 3-digit category  $i$ , weighted by  $j$ 's expenditure 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 MOSPI, 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,  $\tau_{it}$  is defined analogously as the level of that tariff.

- *Input Similarity*  $IS_{jk}^t$  :

$$IS_{jk}^t \equiv \sum_{i \in \Omega} \theta_{ijt} \bar{\theta}_{ik}$$

- *Output Similarity*  $OS_{jk}^t$  :

$$OS_{jk}^t \equiv \sum_{i \in \Omega} \sigma_{ijt} \bar{\sigma}_{ik}$$

- *Input Similarity weighted by a policy change:*

$$ISDR_{jk}^t \equiv \sum_{i \in \Omega} \theta_{ijt} \bar{\theta}_{ik} \delta_{it}, \quad IST_{jk}^t \equiv \sum_{i \in \Omega} \theta_{ijt} \bar{\theta}_{ik} \Delta\tau_{it}$$

- *Output Similarity weighted by a policy change:*

$$OSDR_{jk}^t \equiv \sum_{i \in \Omega} \sigma_{ijt} \bar{\sigma}_{ik} \delta_{it}, \quad OST_{jk}^t \equiv \sum_{i \in \Omega} \sigma_{ijt} \bar{\sigma}_{ik} \Delta\tau_{it}$$

- *Upstream and Downstream:*

$$\text{upstream}_{jk}^t = \sum_{i=1}^N \sigma_{ji}^t \bar{\theta}_{ik}, \quad \text{downstream}_{jk}^t = \sum_{i=1}^N \sigma_{ji}^t \bar{\theta}_{ki}.$$

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

## APPENDIX D. THEORY APPENDIX

### D.1. Firm Input Choice.

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

(1) *The cost index for inputs from industry  $i$  for firm  $j$  at time  $t$  are*

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

(2) *Since  $d \ln S_{ijt} / d \ln \underline{c}_{ijt} = 1 + \Omega_{it} / (\sigma - 1)$ , 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*

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

*Proof.* Firms solve

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

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

Cost minimization conditional on  $\underline{c}_{ijt}$  implies a first order condition of<sup>28</sup>

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

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  $\Omega_{it} > 1 - \sigma$ ,  $\eta_{it}$  is finite and the input choice is non-degenerate.<sup>29</sup>

Defining the cost index of input  $i$  as  $S_{ijt}$  we have minimum costs of  $S_{ijt} M_{ijkt}$  where

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

and therefore

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

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

**Proposition.** *For firm-input expenditure shares  $\theta_{ijt}$ , the optimal capability choice is*

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

<sup>28</sup>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  $\sigma = 1$ ).

<sup>29</sup>Otherwise for  $\sigma < 1$  it is optimal to use all of the cheapest input and for  $\sigma > 1$ , input vectors of the type  $\kappa s^{1-\sigma}$  all satisfy the production constraint so as  $\kappa \rightarrow 0$ , costs go to zero.

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

$$(D.1) \quad \ln R_{jkt} = \underbrace{\ln \frac{1-\rho}{\rho} \left( \rho^{\frac{\rho}{1-\rho}} D_{kt}^{\frac{1}{1-\rho}} \right)}_{\text{Demand } (kt)} - \underbrace{\frac{\rho}{1-\rho} \sum_i \bar{\theta}_{ik} \ln \psi_{it} (1 - \Theta_{it}^{-1})^{\frac{1}{1-\sigma}} \frac{s_m^{\Theta_{it}-1}}{\bar{\theta}_{ik}}}_{\text{Supplier } (kt)} \\ + \underbrace{\frac{\rho}{1-\rho} \ln \varphi_{jk}}_{\text{RCA } (jk)} + \underbrace{\frac{\rho}{1-\rho} \sum_i \Theta_{it}^2 (\bar{\theta}_{ik} \theta_{ijt} - \theta_{ijt}^2/2)}_{\text{Firm Capability } (jkt)}$$

with the dimension of variation listed below each term.

*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

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

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

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

it follows that the first order condition for profit maximization

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

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

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

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

**D.2. Extensive Product Margin.** Equation (4.3) can be modified to consider the extensive product margin choice of firms. Assume firms face a fixed cost  $(1 - \rho) f_{kt}$  to produce in an industry  $k$  each period, so produce when profits  $\pi_{jkt} = (1 - \rho) R_{jkt} > (1 - \rho) f_{kt}$ . From Equation (4.2), with identical coefficients and fixed effects similar to Equation (4.3) and error terms with  $-\epsilon_{jkt}$  logistic, firms operate in industry  $k$  when either of the following equations is positive:

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

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.** *An input similarity approximation for an instrumental variable first stage regression, holding capabilities constant based on demand shocks is*

$$\sum_i (\bar{\theta}_{ik} \theta_{ijt} - \theta_{ijt}^2/2) \approx \lambda \sum_i (\bar{\theta}_{ik} \theta_{ijt-1} - \theta_{ijt-1}^2/2) + \gamma_{kt} \sum_i \chi_{jkt-1} (\bar{\theta}_{ik} - \theta_{ijt-1})^2$$

where  $\chi_{jkt}$  are firm revenue shares for a firm in year  $t$ . The coefficients are as follows:

- $\lambda$  should equal one,
- $\gamma_{kt}$  is a demand innovation term  $(D_{kt}/D_{kt-1} - 1) / (1 - \rho)$ .

*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

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

Holding  $c_{ijt}$  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

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

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

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

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

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

□

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

Table 20 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 20. Comparative advantage of single-industry plants, 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. \*\* $p < 0.05$ , \* $p < 0.10$ .