THE COMPARATIVE ADVANTAGE OF FIRMS

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

April 17, 2019

Abstract. Multiproduct rms dominate production, and their product turnover contributes substantially to aggregate growth. Theories propose that multiproduct rms grow by diversifying into products which need the same know-how or capabili-

1. Introduction

Theories of the rm, dating back to

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).

(a) Industry Co-production matrix (b) Intermediate input similarity matrix

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 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.[4] 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

[4]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 rms better access to input supplies, enables an examination of how policy interacts with input similarity to a ect 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 ndings of Schott (2004), which shows that countries' within-product specialization re ects 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 rms. Starting from the primitive of industry-speci c production functions, di erences across rms 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 rms that have higher use for those inputs. These rms step up production, but much more so in industries which use these inputs more. In sum, policy-induced improv27(v)le0(the-276(b)-r)-1(7gr2)-403(s)-408(h)]ndu

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-

Conley and Dupor (2020

variation to identify input-based comparative advantage.[9] 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.[10] Recent macroeconomic studies stress the importance of input linkages in amplifying micro shocks and policy effects.[11] The development literature emphasizes their role in aggregate productivity and volatility (Koren and Tenreyro 2013), and in motivating policies such as domestic content requirements that have interested governments across the developing world (Harrison and Rodriguez-Clare 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.

[9] 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.

[10] 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.

[11] 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.[12]

The three-digit industries are in 60 two-digit sectors. To give a sense of the level of detail in this classification, consider the sector Cotton, Cotton yarn, and Fabrics sector (ASIC 63) which has various 3-digit industries, such as Cotton fabrics including cotton hosiery fabrics (ASIC 633), Made up articles of cotton including apparel (ASIC 634) and Processing or services of cotton, cotton yarn and fabrics (ASIC 638). To take another example, the 3-digit industry Stainless steel in primary and finished form (ASIC 714) is an industry in the sector Iron & Steel (incl. stainless steel), and articles thereof (ASIC 71).

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.

[12]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, rms that span multiple industries account for a dispro-

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:

$$\overline{IS}_{mn} = \sum_i \ {}_{mi} \ {}_{ni}$$

where ${}_{mi}$ is the sum of expenditure of single-industry rms that only produce m on intermediate inputs from i, divided by total expenditure of these rms 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 bres and yarns. Many base chemicals are applicable in di erent industrial processes. This correlation motivates an examination of rms' 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 rms' revealedofadv w and in explaining revealed comparative advantage. We nd that rms' 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 rm 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 aclose resemblance to the reduced-form regressions from this Section, but provides a structural interpretation of the estimated coe cients.

3.1. Input Similarity. A natural way to bring the industry-level input similarity from above to the rm level is to consider the inner product of the *rm's* vector of intermediate input expenditure shares, ${}_{j}$, with the vector of intermediate input expenditure shares of an industry k:

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

where i indexes the expenditure shares of spending on three-digit inputs and t denotes time. We construct the aggregate intermediate input shares ${}_{ki}$ by aggregating up the micro-data of single-industry plants that only produce in industry k. The input similarity measure ranges from zero, when rm j and sector k have no three-digit inputs in common, to one, when the input expenditure shares of rm j and sector k are identical. The crucial di erence between this rm-level input similarity and the aggregate input similarity constructed above in Section 2.2.2 is that this one incorporates idiosyncratic rm-speci c variation in input mixes. The rm's input mixes may deviate from the one observed in input-output tables because of the rm producing outputs belonging to multiple industries, or because of other sources of variation. This rm-speci c variation is quantitatively important: a set of input-output dummies explains only 61% of the overall variation in rm's cost shares ${}_{ij}$. As an inner product of a vector of

of revenue. These effects control for all shocks that might make all firms in industry k^0 more or less likely to start producing in industry k. Finally, ε^t_{jk} 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

Table 2. Industry Addition: Input Similarity and Vertical Relatedness

	Dependent variable: Add_{jkt}		
	(1)	(2)	(3)
InputSimilarity .69tdd			

of capital and skills (Government of India, 2009 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 sys-

when input i gets de-reserved, rms 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 tari -change-weighted input similarity measure, analogous to the derservation-weighted input similarity.[16] When input i gets de-reserved or gets tari reductions, rms that have been using i intensively are more likely to add products that rely heavily on i. Later, the structural estimation provides a tari equivalent for de-reservation.

Table 5. Product Addition: The Impact of Dereservation

	Dependent variable: Add_{jkt}			
	(1)	(2)	(3)	(4)
InputSimilarity $^{0}_{jk}$	0.0220	0.0216	0.0157	0.0153
	(0.00021)	(0.00021)	(0.00035)	(0.00035)
InputSimilarity-Dereservation $^{0}_{jkt}$	0.0227	0.0228	0.0151	0.0143
	(0.0013)	(0.0013)	(0.0013)	(0.0013)
InputSimilarity-Tari $^{0}_{jkt}$				-0.0582
				(0.0054)
Firm Year FE $_{jt}$	Yes	Yes	Yes	Yes
Industry Year FE $_{kt}$		Yes		
k k^0 t FE $_{kk^0t}$			Yes	Yes
R^2	0.00840	0.00979	0.0417	0.0417
Observations	77745382	77745382	77726154	77726154

Standard errors in parentheses, clustered at the rm-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 rms who produce one, or a certain set of industries, are able to obtain

[16] This is constructed by replacing the de-reservation indicator $_{jit}$ with the change in India's import tari s $_{jit}$. For the precis02 0 Tdb e]TJ/ andencx..

relatively higher prices on products from another industry.[17] To capture such complementarities, we construct a measure of output similarity analogously to our input similarity index as an inner product between rm j's sales shares and the aggregate industry k's sales shares:

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

where i runs over the set of three-digit industries. The vector $_{j}$ denotes the sales of rm j belonging to industry i at time t, divided by the total of j's sales at time t. The vector $\bar{}_{k}$ denotes the (size-weighted) average $_{j^0 i}$ among rms j^0 that derive their highest fraction of revenue from sales in k. Again, this measure captures the degree of overlap between rm j's portfolio of sales (across industries), and the average portfolio of rms that sell most in is the expected expenditure share of industry k on rms that feature the same product mix as j.

Table 6. Product Addition: Robustness

	Dependent variable: Add_{jkt}				
	(1)	(2)	(3)	(4)	(5)
InputSimilarity$^0_{jk}$	0.0220	0.0146	0.0142	0.0111	0.0107
	(0.00021)	(0.00027)	(0.00027)	(0.00035)	(0.00035)
InputSimilarity-Dereservation$^0_{jkt}$	0.0227	0.0212	0.0212	0.0128	0.0121
	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)
OutputSimilarity$^0_{jk}$		0.00860	0.00852	0.0599	0.0599
		(0.00039)	(0.00039)	(0.0011)	(0.0011)
OutputSimilarity-Dereservation$^0_{jkt}$		0.0160	0.0156	0.00622	0.00623
		(0.00086)	(0.00086)	(0.0012)	(0.0012)
Upstream$^0_{jk}$		0.0197	0.0186	0.0160	0.0160
		(0.00055)	(0.00055)	(0.0017)	(0.0017)
Downstream$^0_{jk}$		-0.00526	-0.00479	-0.00238	-0.00244
		(0.00033)	(0.00033)	(0.00083)	(0.00083)
InputSimilarity-Tari $^0_{jkt}$					-0.0549
					(0.0054)
Firm Year FE $_{jt}$	Yes	Yes	Yes	Yes	Yes
Industry Year FE $_{kt}$			Yes		
k k^0 t FE $_{kk^0t}$				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 rm-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 coe cient output similarity is positive and significant, in particular in the speci cations with k k^0 t xed e ects. 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.

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 dereserved 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 ve sellers of spice oleoresins in the world is a good example of how the product mix of rms changed with the dereservation of spices.

Headquartered in Cochin, Kerala, the Akay Group is a large Indian rm 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 avoured oil. Following the dereservation, Akay expanded its product o erings 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

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, rm j combines inputs from industry i, M_{ijkt}, using a constant return to scale Cobb-Douglas technology with industry input expenditure shares ¯$_{ik}$ and idiosyncratic industry productivity labeled ′$_{jk}$.[18] At input prices S_{ijt} $_{it}$

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 $\underline{c}_{jt} \quad c_{jkt}$ in each industry, where

$$\underline{c}_{jt} \quad \exp \left(\sum_i \quad \ln \underline{c}_{i0} \quad \ln \underline{c}_{ijt} \; {}^2 = 2 \right) :$$

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 $\underline{c}_{i0} = 1$.[20]

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

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

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

$$_{jt} = \sum_k \; _{jkt} = \sum_k p_{jkt}\, q_{jkt} \quad \sum_k \sum_i \quad \underline{c}_{jt} \; S_{it}\, M_{ijkt} = \sum_k \; D_{kt}\, q_{jkt} \quad \underline{c}_{jt} \; c_{jkt}\, q_{jkt} :$$

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* $_{ijt}$*, the optimal capability choice is*

$$\ln \underline{c}_{ijt} = \quad _{it} \quad _{ijt}$$

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

where $_{it}$ $1 +$ $_{it} = ($ $1)$ *is the elasticity of input price w.r.t. capability and rm-industry revenues are given by*

$$\ln R_{jkt} = \underbrace{\ln \frac{1}{\quad} \frac{}{1} D_{kt}^{\frac{1}{1}}}_{\text{Demand }(kt)} \underbrace{\frac{}{1} \sum_{i} \bar{}_{ik} \ln {}_{it} \quad 1 \quad {}_{it}{}^{1} \frac{1}{1} \frac{s_m^{it} \quad 1}{\bar{}_{ik}}}$$

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 $'_{jk}$).

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

$$(4.2)\quad \ln R_{jkt} = {}_{kt} + {}_{jk} + \underbrace{\frac{}{1} \sum_{i} {}^{2}_{i0} + \frac{2\ {}_{i0}}{(\quad 1)}\ {}_{P}\,(P_{it} \quad P_{i0}) \quad \bar{}_{ik}\ {}_{ijt} \quad \frac{{}^{2}_{ijt}}{2}}_{\text{Firm Capability Change } (\ {}_{jkt}\ )} :$$

The theory above signs $_{it}$ as the same sign as $\ 1$, so estimating $_{P}\ 2\ {}_{i0}=(\quad 1)$ gives the same sign as $_{P}$ and allows for testing hypotheses about $_{P}$.

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) $_{B}$ at the three digit level (with B_{it} equal to 1 if a product is reserved and zero otherwise) and a linear effect of tariffs on entry for three digit tariffs $_{it}$ (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 $_{i0} = \bar{\ }$, so that

$$_{it} = \bar{\ } + {}_B B_{it} + \quad (1 + {}_{it}):$$

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

$$(4.3) \qquad \ln R_{jkt} = {}_0 \sum_i \bar{\ }_{ik} \; {}_{ijt} \; \frac{{}^2_{ijt}}{2} + {}_1 \sum_i ({}_B B_{it} + \quad {}_{it}) \; \bar{\ }_{ik} \; {}_{ijt} \; \frac{{}^2_{ijt}}{2}$$
$$+ \; {}_{kt} + {}_{jk}:$$

with $_0 = \; {}^2_{i0} = (1 \quad)$, $_1 = 2 \; {}_{i0} = (1 \quad)(\quad 1)$. The tari equivalent of dereser-vation can then be computed from $_B \; {}_1 = \; {}_1 = \; {}_B = \;$. 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 $R_{jkt} > (1 \quad) f_{kt}$, so we estimate Equation (4.3) as a linear probability model for the outcome that observed revenues of the rm-industry are positive each period.[22] As we are estimating probabilities, we can think of how comparative advantage shifts the *production probability frontier* of rms.

4.3. Structural Instrumentation. In Equation (4.2), rm expenditure shares $_{ijt}$ are a function of xed technology $\bar{\ }_{ik}$, time varying input prices $_{it}$, demand shocks D_{kt} and idiosyncratic productivities $'_{jk}$. Input price and demand shocks are estimated through industry-time xed e ects. Idiosyncratic productivities are estimated through rm-industry xed e ects, expressed as Revealed Comparative Advantage. Technology

---

[22] 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 di culties with IV-Logit due to the high dimensional parameter space and well known sensitivity of that estimatorn

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

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

There might be omitted variables from our structural equation that cause $_{ijt}$ 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 within rm distribution of activity. [24] In fact, examining the estimating Equation (4.2), from the Proposition is then:

$$(4.4)\quad \sum_i \bar{\ }_{ik}\ {}_{ijt}\ \frac{{}^{2}_{ijt}}{2} = \sum_i \bar{\ }_{ik}\ {}_{ijt\ 1}\ \frac{{}^{2}_{ijt\ 1}}{2} + {}_{kt}\sum_i {}_{jkt\ 1}\ \bar{\ }_{ik}\ {}_{ijt\ 1}{}^{2}$$
$$+ {}_{kt} + {}_{jk}.$$

Equation (4.4) is composed of three parts: the xed e ects found in the main structural equation for revenues, a lagged term for the endogenous sum $\sum_i \bar{\ }_{ik}\ {}_{ijt\ 1}\ {}^{2}_{ijt\ 1}/2$, and linear adjustment based on predicted input share changes from lagged revenue shares ${}_{jkt\ 1}$ and contemporaneous industry level demand shocks ${}_{kt}$. This last term is essentially a (lagged) sales weighted `technological distance' measure of the rm away from an industry k times the magnitude of the demand innovation which predicts the change in $\sum_i \bar{\ }_{ik}\ {}_{ijt}\ {}^{2}_{ijt}/2$ 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 rst stage prediction equation.[25] Accordingly, de ne both $e_{ijkt}\ \bar{\ }_{ik}\ {}_{ijt}\ {}^{2}_{ijt}/2$ and $e_{jkt}\ {}_{jkt}\ \bar{\ }_{ik}\ {}_{ijt}{}^{2}$ and the following sums for and the KxT vector :

$$I_{jkt}(\ ;\ )\quad \sum_i e_{ijkt\ 1} + {}_{kt}\sum_i e_{jkt\ 1};$$
$$I^{B}_{jkt}(\ ;\ )\quad \sum_i B_{it}\, e_{ijkt\ 1} + {}_{kt}\sum_i B_{it}\, e_{jkt\ 1};$$
$$I_{jkt}(\ ;\ )\quad \sum_i {}_{it}\, e_{ijkt\ 1} + {}_{kt}\sum_i {}_{it}\, e_{jkt\ 1}.$$

[25] 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[26]:

$$\frac{\sum_i e_{ijkt}}{\sum_i B_{it}} = {}_{kt} + {}_{jk} + I_{jkt}\ {}^{11}, {}^{11} + I^{B}_{jkt}\ {}^{12}, {}^{12} + I_{jkt}\ {}^{13}, {}^{13} + {}_{jkt} \tag{4.5}$$

Fruit and vegetable juices industry (135) of 8.5%, whereas the single-industry rms 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 $\overline{CA}_{jkt}$ 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 $\overline{CA}_{jkt}$ for the industry *k* with the highest premium for 25 industries. Hence, the examples below are not outliers: in many industries input capabilities shape rm-level comparative advantage to an extent that is economically relevant to rms.

Table 9. Average rm-level comparative advantage: Some examples

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

Note: The table shows the average rm-level comparative advantage $\overline{CA}_{kk^0}$ 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 rms operate in and the rows contain the rm sales ranking of each industry. For rms 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	# of Industries With Positive Sales									
rank	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 rms 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 rms that o er 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 diversi ed multiproduct rms 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 diversi ed rms, which are much bigger than other rms. To highlight this selection e ect, entries in Table 11 contain the size-weighted comparative advantage of rms. We normalize sales weights by the average sales of a single-product rm in that industry, so that the interpretation is premia weighted by the equivalent number of typical single-product rms. The single-industry premium from acquiring capabilities is hardly changed at 5.5 per cent, compared to the typical single-industry rm. 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) prerelevance 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 multiproduct firms imply production choices and input capabilities are jointly determined. Production choices are interdependent on the relative demands a firm faces and the portfolio of industries a firm enters depends on 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

Amirapu, A., M. Gechter, and G. Smagghue (2018): Dynamic E ects of Product Market Regulation: Evidence from India, *Working Paper*.

Amiti, M. and J. Konings (2007): Trade liberalization, intermediate inputs, and

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 Diversi cation Strategy, *Strategic Management Journal*, 26, 1153 1171.

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 rms 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 Diversi cation, *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

Liu, R. (2010): Import competition and rm 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 Multibusiness Firm: Empirical Analysis of Portfolio Interrelationships and Corporate Financial Performance, *Strategic Management Journal*, 16, 277 299.

Rondi, L. and D. Vannoni (2005): Are EU leading rms returning to core business? Evidence on refocusing and relatedness in a period of market integration, *Review of Industrial Organization*, 27, 125 145.

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	0.0108		

Table 17. Domestic input unit values after dereservation Robustness

	Dependent variable: log p_{jit}			
	(1)	(2)	(3)	(4)
t year i was de-reserved	-0.128	-0.0864	-0.0477	-0.0635
	(0.014)	(0.015)	(0.012)	(0.014)
Sample	All	All	Safe	Safe
Year FE	Yes	Yes	Yes	Yes
Input Product FE	Yes		Yes	
Firm 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 rm-year level.

$^{+}$ $p < 0{:}10$, $p < 0{:}05$, $p < 0{:}01$

Figure B.1. Estimated Changes in $^{-}{}_{ik}$ from Dereservation

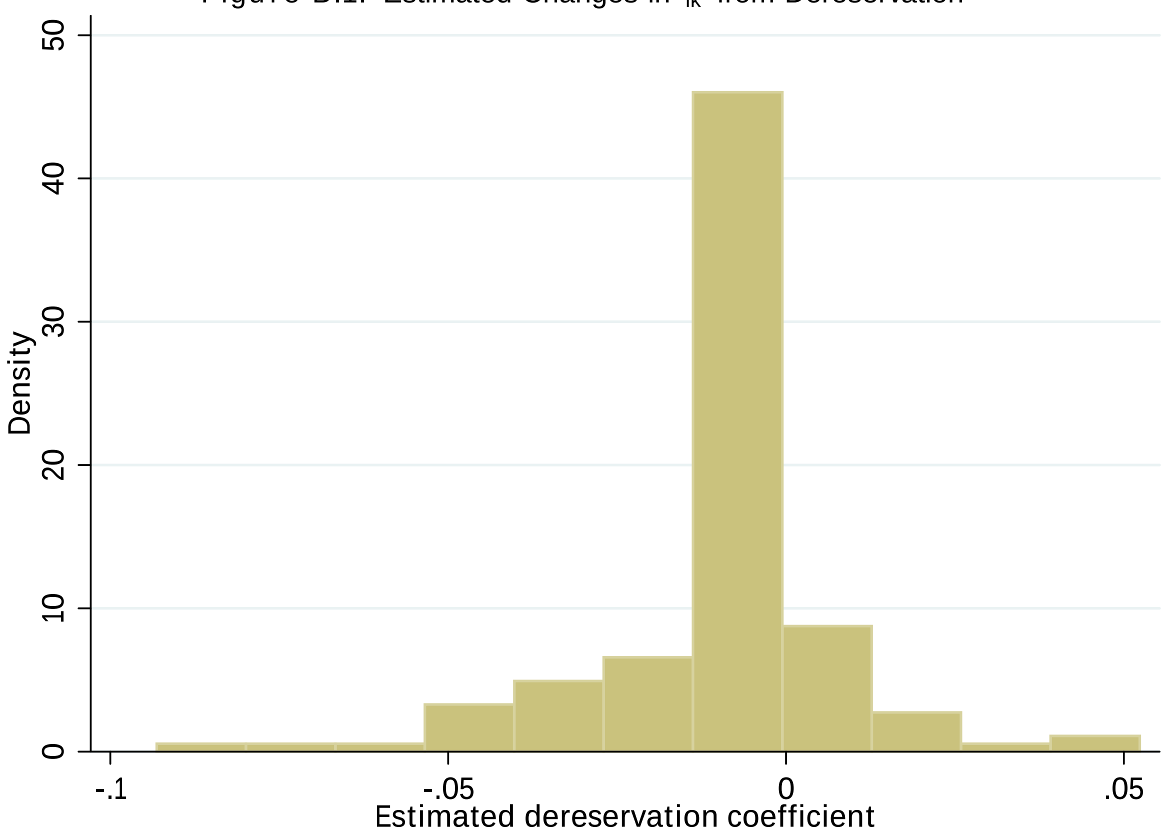

Table 18. Industry Drop Regressions:

	Dependent variable: $Drop_{jkt}$		
	(1)	(2)	(3)
IS^0_{jkt}	-0.00940†	-0.112	-0.0839
	(0.0050)	(0.0068)	(0.0076)
$ISDR^0_{jkt}$	-0.185	-0.0541†	-0.0842
	(0.028)	(0.029)	(0.034)
OS^0_{jkt}	-0.185	-0.170	-0.136
	(0.0034)	(0.0040)	(0.0058)
$OSDR^0_{jkt}$	-0.0534	-0.0462	-0.0459
	(0.0060)	(0.0068)	(0.0080)
UP0	-0.0273	-0.0556	-0.0360†
	(0.0061)	(0.010)	(0.020)
DOWN0	0.0993	0.0246	0.0391†
	(0.0096)	(0.011)	(0.021)

Firm

Table 19. Intensive Margin Regressions:

	Dependent variable: $\log \text{Sales}_{jkt}$		
	(1)	(2)	(3)
IS^{0}_{jkt}	0.451	0.799	0.466
	(0.037)	(0.048)	(0.047)
$ISDR^{0}_{jkt}$	0.538	0.145	0.392
	(0.19)	(0.19)	(0.18)
OS^{0}_{jkt}	3.821	3.326	1.414
	(0.023)	(0.027)	(0.029)
$OSDR^{0}_{jkt}$	-0.341	-0.497	-0.239
	(0.039)	(0.044)	(0.041)
UP0	-1.279	0.0798	0.304
	(0.045)	(0.070)	(0.10)
DOWN0	1.876	0.526	0.134
	(0.075)	(0.081)	(0.11)
Firm Year FE $_{jt}$	Yes	Yes	Yes
Industry Year FE $_{kt}$		Yes	
k k^{0} t FE $_{kk^{0}t}$			Yes
R^2	0.804	0.833	0.911
Observations	251028	250963	220611

Standard errors in parentheses, clustered at the rm-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 de nitions.

Add dummies Add_{jkt} : one if and only if j does not produce any product in 3-digit industry k at time t and does produce a product in k at time $t + 1$. We exclude outputs with zero or missing sales from the set of produced products.

Drop dummies Drop_{jkt} : one if and only if j does produce a product in 3-digit industry k at time t and does not produce any product in k at time $t + 1$. We exclude outputs with zero or missing sales from the set of produced products.

Tari change $_{jit}$: Di erence between year t Indian import tari and year 2000 tari on 5-digit products in 3-digit category i, weighted by j's expenditure on 5-digit imports in i. We concord tari s from the 6

C.3. Sample de nition. 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* $_{it} > 1$ *which is necessary for non-degenerate variety choices. De ne 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 rm j at time t are*

$$S_{ijt} = \frac{_{it}}{_{it} + (\quad 1)}^{1=(1\quad)} \underline{c}_{ijt}^{1\quad_{it}=(1\quad)} s_m^{\ _{it}=(1\quad)}:$$

(2) *Since* $d \ln S_{ijt} = d \ln \underline{c}_{ijt} = 1 + \ _{it} = (\quad 1)$*, it follows that when inputs are*

(a) *substitutes (* > 1*), increasing varieties lowers costs (Love for Variety),*

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

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

$$c_{jkt} = \frac{1}{\quad}$$

Cost minimization conditional on $\underline{c}_{ijt}$ implies a rst order condition of[28]

$$m_{ijkt}^{(\ \ 1)=} = M_{ijkt}^{(\ \ 1)=} \frac{}{1} \frac{s_{it}}{}^{1} \quad \text{where} \quad _{it} = \left(\int_{1}^{\underline{c}_{ijt}} \frac{}{1} s^{1} \, dG_{it}(s) \right)^{1=(1\ \ )} :$$

Under these distributional assumptions, we have

$$_{it} = \frac{}{1} \left(\frac{_{it}}{_{it} + (\ \ 1)} s_m^{_{it}} \underline{c}_{ijt}^{1\ \ _{it}} \right)^{1=(1\ \ )}$$

under the condition $_{it} > 1$, $_{it}$ is nite and the input choice is non-degenerate.[29] De ning the cost index of input i as S_{ijt} we have minimum costs of $S_{ijt} M_{ijkt}$ where

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

and therefore

$$d \ln S_{ijt} = d \ln \underline{c}_{ijt} = 1 + \ _{it} = (\ \ 1) :$$

Now the restriction $_{it} > 1$ is especially informative as if > 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 < 1 consistent with hate for variety. Unit input costs c_{jkt} conditional on capabilities are then as above.

Proposition. For rm-input expenditure shares $_{ijt}$, the optimal capability choice is

$$\ln \underline{c}_{ijt} = \ \ _{it}\ _{ijt}$$

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

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

where $_{it}$ $1 +$ $_{it} = ($ $1)$ is the elasticity of input price w.r.t. capability and rm-industry revenues are given by

$$\ln R_{jkt} = \ln \underbrace{\frac{1}{\quad} \; \frac{\quad}{1} D_{kt}^{\frac{1}{1}}}_{\text{Demand }(kt)} \; \underbrace{\frac{\quad}{1} \sum_{i} \bar{}_{ik} \ln {}_{it} \; 1 \; {}_{it}^{1} \; \frac{1}{1} \; \frac{s_m^{it} \quad 1}{\bar{} \; ik}}$$

Supn..105 4.948 Td [( )]TJ/F27 7.9701 Tf 6.586 0 Td [(1)]TJ/F30 7.97

Holding $\underline{c}_{ijt}$ xed, for $_{jkt}$ $C_{jk} = C_j$ the cost share of industry k for rm j (equal to revenue shares), it is the case that

$$\frac{d \;_{ijt}}{dD_{kt}} = \frac{1}{C_j^2}\left[\frac{\bar{}_{ik}}{1} \frac{C_{jk}}{D_{kt}} C_j \quad \frac{1}{1} \frac{C_{jk}}{D_{kt}} \sum_k \bar{}_{ik} C_{jk} \right] = \frac{_{jkt}}{1} \frac{\bar{}_{ik} \quad _{ijt}}{D_{kt}}$$

it follows from the mean value theorem that for some $f \;_{jk}\, g$ with each $_{jk}$ 2 $[D_{kt \; 1}; D_{kt}]$ and cost shares $_{jk}$ and expenditure shares $_{ij}$ evaluated at $f \;_{jk}\, g$ that

$$\sum_i \bar{}_{ik} \;_{ijt} \quad {}^2_{ijt} = 2 \quad \bar{}_{ik} \;_{ijt \; 1} \quad {}^2_{ijt \; 1} = 2 = \sum_i \bar{}_{ik} \quad _{ij} \frac{_{jk}}{1} \bar{}_{ik} \quad _{ijt} \frac{D_{kt} \quad D_{kt \; 1}}{_{jk}}:$$

Rede ning $_{jk} = D_{kt \; 1}$ as common across rms, yields the (feasible) approximation

$$\sum_i \bar{}_{ik} \;_{ijt} \frac{{}^2_{ijt}}{2} \quad \sum_i \bar{}_{ik} \;_{ijt \; 1} \frac{{}^2_{ijt \; 1}}{2} + \sum_i \bar{}_{ik} \quad _{ijt \; 1} {}^2 \frac{_{jkt \; 1}}{1} \frac{D_{kt} \quad D_{kt \; 1}}{D_{kt \; 1}}:$$

Appendix E. Average Firm-level Comparative Advantage, by industry

Table 20 shows the average comparative advantage of single-industry rms in industry k^0, for the industry in which they enjoy the highest average CA_{jkt} .

Table 20. Comparative advantage of single-industry plants, by industry

Industry k^0	Highest average comparative advantage industry (except k^0)	Comp Adv
Dairy products	Live animals, chie y for food	15.8**
Other jute and natural bre 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 bre 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 bre, incl. cotton thread	10.2**
Cotton yarn and bre, incl. cotton thread	Ginned cotton, cotton, and raw cotton waste	10.0*