Prices, Markups and Trade Reform*

Jan De Loecker† Pinelopi K. Goldberg ‡ Amit K. Khandelwal§ Nina Pavcnik¶

First Draft: February 2012
Final Draft: October 2015

Forthcoming, *Econometrica*

Abstract

This paper examines how prices, markups and marginal costs respond to trade liberalization. We develop a framework to estimate markups from production data with multi-product firms. This approach does not require assumptions on the market structure or demand curves faced by firms, nor assumptions on how firms allocate their inputs across products. We exploit quantity and price information to disentangle markups from quantity-based productivity, and then compute marginal costs by dividing observed prices by the estimated markups. We use India's trade liberalization episode to examine how firms adjust these performance measures. Not surprisingly, we find that trade liberalization lowers factory-gate prices and that output tariff declines have the expected pro-competitive effects. However, the price declines are small relative to the declines in marginal costs, which fall predominantly because of the input tariff liberalization. The reason for this incomplete cost pass-through to prices is that firms offset their reductions in marginal costs by raising markups. Our results demonstrate substantial heterogeneity and variability in markups across firms and time and suggest that producers benefited relative to consumers, at least immediately after the reforms.

Keywords: Variable Markups, Production Function Estimation, Pass-through, Input Tariffs, Trade Liberalization

*The main work for this project was carried out while Goldberg was a Fellow of the Guggenheim Foundation, De Loecker was a visitor of the Cowles Foundation at Yale University and a visiting Professor at Stanford University, and Khandelwal was a Kenen Fellow at the International Economics Section at Princeton University. The authors thank the respective institutions for their support. We are grateful to Steve Berry, Elhanan Helpman, Ariel Pakes, Andres Rodriguez-Clare and Frank Wolak for useful discussions at early stages of this project and seminar participants at several institutions and conferences. We also thank the Editor and three anonymous referees for insightful comments and suggestions.

†Princeton University, Fisher Hall, Prospect Ave, Princeton, NJ 08540, *email*: jdeloeck@princeton.edu

‡Yale University, 37 Hillhouse, New Haven, CT 06520, *email*: penny.goldberg@yale.edu

§Columbia Business School, Uris Hall, 3022 Broadway, New York, NY 10027 *email*: ak2796@columbia.edu

¶Dartmouth College, 6106 Rockefeller Hall, Hanover, NH 03755, *email*: nina.pavcnik@dartmouth.edu

1 Introduction

Trade reforms have the potential to deliver substantial benefits to economies by forcing a more efficient allocation of resources. A large body of theoretical and empirical literature has analyzed the mechanisms behind this process. When trade barriers fall, aggregate productivity rises as less productive firms exit and the remaining firms expand (e.g., Melitz (2003) and Pavcnik (2002)) and take advantage of cheaper or previously unavailable imported inputs (e.g., Goldberg et al. (2010a), Amiti and Konings (2007), Halpern et al. (2011)). Trade reforms have also been shown to reduce markups (e.g., Levinsohn (1993) and Harrison (1994)). Based on this evidence, we should expect trade reforms to exert downward pressure on firm prices. However, we have little direct evidence on how prices respond to liberalization because they are rarely observed during trade reforms. We fill this gap by examining how prices, and their underlying markup and cost components, adjust during India's comprehensive trade liberalization. To obtain the markup and cost components we develop a unified framework to estimate jointly markups and marginal costs from production data.

Our paper makes three main contributions. First, we develop a unified framework to estimate markups and marginal costs of multi-product firms across a broad set of manufacturing industries. Since these measures are unobserved, we must impose some structure on the data. However, our approach does not require parametric assumptions on consumer demand, market structure or the nature of competition common in industrial organization studies. This flexibility is particularly appealing in settings when one wants to infer the full distribution of markups across firms and products over time in different manufacturing sectors. Since prices are observed, we can directly recover marginal costs from the markup estimates. Data containing this level of detail are becoming increasingly available, so this methodology is useful to researchers studying other countries and industries. The drawback of this approach is that we are unable to perform counterfactual simulations since we do not explicitly model consumer demand and firm pricing behavior.

The second and key contribution of our study is towards the methodology to estimate production functions. In order to infer markups, the proposed approach requires estimates of production functions. Typically, these estimates have well-known biases if researchers use revenue rather than quantity data. Estimates of "true" productivity (or marginal costs) are confounded by demand shocks and markups, and these biases may be severe (see Foster et al. (2008) and De Loecker (2011)). We address the output price-bias by estimating a quantity-based production function using data that contain the prices and quantities of firms' products over time. The focus on a quantity-based production function highlights the need for the estimation to address two additional biases that have not received much attention in the literature: the bias stemming from the unobserved allocation of inputs across products within multi-product firms and the bias stemming from unobserved input prices (due to the use of quality-differentiated inputs) by firms - the so-called input price bias. Our study contributes an approach to address these biases. This is important as future waves of plant- and firm-level data may start providing information on physical quantities of output forcing researchers to confront the challenges associated with multi-product production function estimation. Moreover, researchers may want to start combining data from firm-level production

surveys with fine-grained product-level information from consumer scanner data, which will also require an explicit treatment of multi-product firms in the production function estimation.

Third, existing studies that have analyzed the impact of trade reforms on markups have focused exclusively on the competitive effects from declines in output tariffs (e.g., Levinsohn (1993) and Harrison (1994)). Comprehensive reforms also lower tariffs on imported inputs and previous work, particularly on India, has emphasized this aspect of trade reforms (e.g., Goldberg et al. (2009)). These two tariff reductions represent distinct shocks to domestic firms. Lower output tariffs increase competition by changing the residual demand that firms face. Conversely, firms benefit from lower costs of production when input tariffs decline. It is important to account for both channels of liberalization to understand the overall impact of trade reforms on prices and markups. In particular, changes in markups depend on the extent to which firms pass these cost savings to consumers, the pass-through being influenced by both the market structure and nature of demand. For example, in models with monopolistic competition and CES demand, markups are constant and so by assumption, pass-through of tariffs on prices is complete. Arkolakis et al. (2012) demonstrate that several of the influential trade models assume constant markups and by doing so, abstract away from the markup channel as a potential source of gains from trade. This is the case in Ricardian models that assume perfect competition, such as Eaton and Kortum (2002), and models with monopolistic competition such as Krugman (1980) and its heterogeneous firm extensions like Melitz (2003). There are models that can account for variable markups by imposing some structure on demand and market structure.[1] While these studies allow for richer patterns of markup adjustment, the empirical results on markups and pass-through ultimately depend on the underlying parametric assumptions imposed on consumer demand and nature of competition. Ideally, we want to understand how trade reforms affect markups without having to rely on explicit parametric assumptions of the demand systems and/or market structures, which themselves may change with trade liberalization.

The structure of our analysis is as follows. We use production data to infer markups by exploiting the optimality of firms' variable input choices. Our approach is based on Hall (1988) and De Loecker and Warzynski (2012), but we extend their methodology to account for multi-product firms and to take advantage of observable price data and physical quantity of products. In order to infer markups we assume that firms minimize cost; then, markups are the deviation between the elasticity of output with respect to a variable input and that input's share of total revenue. We obtain this output elasticity from estimates of production functions across many industries. The use of physical quantity data alleviates the concern that the production function estimation is contaminated by prices, yet presents different challenges that we discuss in detail in Section 3. Most importantly, using physical quantity data forces us to conduct the analysis at the product level since without a demand system to aggregate across products, prices and physical quantities are only defined at the product level.

The approach we propose calls for an explicit treatment of multi-product firms. We show

[1] See Goldberg (1995), Bernard et al. (2003), Goldberg and Verboven (2005), Atkeson and Burstein (2008), Melitz and Ottaviano (2008), Feenstra and Weinstein (2010), Nakamura and Zerom (2010), Edmonds et al. (2011), Goldberg and Hellerstein (2013), Arkolakis et al. (2012), Mayer et al. (2014) and Atkin and Donaldson (2014).

how to exploit data on single-product firms along with a sample selection correction to obtain consistent estimates of the production functions. The benefit of using single-product firms at the production function estimation stage is that it does not require assumptions on how firms allocate inputs across products, something we do not observe in our data.[2] This approach assumes that the *physical* relationship between inputs and outputs is the same for single- and multi-product firms that manufacture the same product. That is, a single-product firm uses the same technology to produce rickshaws as a multi-product firm that produces rickshaws and cars. While this assumption may appear strong, it is already implicitly employed in all previous work that pools data across single- and multi-product firms (e.g., Olley and Pakes (1996) or Levinsohn and Petrin (2003)). Once we estimate the production functions from the single-product firms, we show how to back out allocation of inputs across products within a multi-product firm. We obtain the markups for each product manufactured by firms by dividing the output elasticity of materials by the materials share of total revenue.[3] Finally, we divide prices by the markups to obtain marginal costs.

The estimation of the production function provides plausible results and highlights the importance of addressing the input price bias. We also observe that firms have lower markups and higher marginal costs on products that are farther from their core competency, a finding consistent with recent heterogeneous models of multi-product firms. Foreshadowing the impact of the trade liberalizations, we find that changes in marginal costs are not perfectly reflected in changes in prices because of variable markups (i.e., incomplete pass-through).

Our main results focus on how prices, marginal costs, and markups adjust during India's trade liberalization. As has been discussed extensively in earlier work, the nature of India's reform provides an identification strategy that alleviates the standard endogeneity concerns associated with trade liberalization. Perhaps not surprisingly, we observe price declines during the reform period, but these declines appear modest relative to the size of the reform. On average, prices fall 18 percent despite average output tariff declines of 62 percentage points. Marginal costs, however, decline on average by 31 percent due primarily to input tariff liberalization; this finding is consistent with earlier work demonstrating the importance of imported inputs in India's trade reform. The predominant force driving down marginal costs are lower input tariffs reducing the costs of imported inputs, rather than output tariffs reducing X-inefficiencies. The importance of input tariffs is consistent with earlier results by Amiti and Konings (2007) on Indonesia and Topalova and Khandelwal (2011) on India who find that firm-level productivity changes were predominantly driven by input tariff declines. Since our prices decompose exactly into their underlying cost and markup components, we can show that the reason the relatively large decline in marginal costs did not translate to equally large price

[2]Suppose a firm manufactures three products using raw materials, labor and capital. To our knowledge, no dataset covering manufacturing firms reports information on how much of each input is used for each product. One way around this problem is to assume input proportionality. For example, Foster et al. (2008) allocate inputs based on products' revenue shares. Their approach is valid under perfect competition or the assumption of constant markups across all products produced by a firm. While these assumptions may be appropriate for the particular homogeneous good industries they study, we study a broad class of differentiated products where these assumptions may not apply. Moreover, our study aims to estimate markups without imposing such implicit assumptions.

[3]For multi-product firms, we use the estimated input allocations in the markup calculation.

declines was because markups increased: on average, the trade reform raised relative markups by 13 percent. The increases in markups are due to the fact that prices do not respond fully to cost, a finding that has been studied extensively in the exchange rate literature and is consistent with any model with variable markups. Finally, we observe that firms' ability to raise markups even further is mitigated by the pro-competitive impact of output tariff declines, particularly for those firms with very high initial markups. Our analysis is based on data representative of larger firms, so our results are representative of these larger firms.

Our results suggest that the most likely beneficiaries of the trade liberalization in the short-run are domestic Indian firms who benefit from lower production costs while simultaneously raising markups. The short-run gains to consumers appear small, especially considering that we observe factory-gate prices rather than retail prices. However, the additional short-run profits accrued to firms may have spurred innovation in Indian manufacturing, particularly in the introduction of many new products, that benefit consumers in the long run. These new products accounted for about a quarter of overall manufacturing growth (see Goldberg et al. (2010b)). In earlier work, we showed that the new product introductions were concentrated in sectors with disproportionally large input tariff declines that allowed firms access to new, previously unavailable imported materials (see Goldberg et al. (2010a)). In the present paper, we find that firms with larger increases in average markups were more likely to introduce new products, which suggests that higher profits may have financed the development of new products that contributed to long run gains to consumers. In addition, our empirical findings are consistent with an increase in the quality of existing products, which would have further benefited consumers. A more detailed investigation of these channels is beyond the scope of the present paper.

In addition to the papers discussed earlier, our work is related to a wave of recent papers that focus on productivity in developing countries, such as Bloom and Van Reenen (2007) and Hsieh and Klenow (2009). The low productivity in the developing world is often attributed to lack of competition (see Bloom and Van Reenen (2007) and Bloom and Van Reenen (2010)) or the presence of policy distortions that result in a misallocation of resources across firms (Hsieh and Klenow (2009)). Against this background, it is natural to ask whether there is any evidence that an increase in competition or a removal of distortions reduces production costs. India's reforms are an excellent context to study these questions because of the nature of the reforms and the availability of detailed data. Trade protection is a policy distortion that distorts resource allocation. Limited competition benefits some firms relative to others, and the high input tariffs are akin to the capital distortions examined by Hsieh and Klenow (2009). Our results suggest that the removal of barriers on inputs lowered production costs, so the reforms did indeed deliver gains in the form of lower production costs. However, the overall picture is more nuanced as firms do not appear to pass the entirety of the cost savings to consumers in the form of lower prices. Our findings highlight the importance of jointly studying changes in prices, markups and costs to understand the full distributional consequences of trade liberalization.

The remainder of the paper is organized as follows. In the next section, we provide a brief

overview of India's trade reform and the data used in the analysis. In Section 3 we lay out the general empirical framework that allows us to estimate markups and marginal costs. Section 3.1 presents the theoretical framework, Section 3.2 presents the empirical methodology to estimate the production function and discusses identification, and Section 3.3 explains the process to recover the allocation of inputs across products for multi-product firms. Section 4 presents the results and Section 5 concludes.

2 Data and Trade Policy Background

We first describe the Indian data since it dictates our empirical methodology. We also describe key elements of India's trade liberalization that are important for our identification strategy. Given that the Indian trade liberalization has been described in a number of papers (including several by a subset of the present authors), we keep the discussion of the reforms brief.

2.1 Production and Price data

We use the Prowess data that is collected by the Centre for Monitoring the Indian Economy (CMIE). Prowess includes the usual set of variables typically found in firm-level production data, but has important advantages over the Annual Survey of Industries (ASI), India's manufacturing census over the 1989-2003 period that spans India's trade liberalization. First, unlike the repeated cross section in the older versions of the Annual Survey of Industries (ASI), Prowess is a panel that tracks firm performance over time. Second, the data span India's 1991 trade liberalization. Third, Prowess records detailed product-level information for each firm. This enables us to distinguish between single-product and multi-product firms, and track changes in firm scope over the sample period. Fourth, Prowess collects information on quantity and sales for each reported product, so we can construct the prices of each product a firm manufactures. These advantages make Prowess particularly well-suited for understanding the mechanisms of firm-level adjustments in response to trade liberalizations that are typically hidden in other data sources, and deal with measurement issues that arise in most studies that estimate production functions.[4]

Prowess enables us to track firms' product mix over time because Indian firms are required by the 1956 Companies Act to disclose product-level information on capacities, production and sales in their annual reports. As discussed extensively in Goldberg et al. (2010b), several features of the database give us confidence in its quality. Product-level information is available for 85 percent of the manufacturing firms, which collectively account for more than 90 percent of Prowess' manufacturing output and exports. Since product-level information and overall output are reported in separate modules, we can cross check the consistency of the data. Product-level sales comprise 99 percent of

[4]The ASI has recently released panel data that contain similar product-level information. These data have the advantage of being a representative survey of Indian manufacturing activity and contain both the wage bill and number of employees, but because these recent waves do not span the Indian trade liberalization period, we are unable to use them for our analysis.

the (independently) reported manufacturing sales. We refer the reader to Appendix C and Goldberg et al. (2010a,b) for a more detailed discussion of the data.

The definition of a product is based on the CMIE's internal product classification, which is based on India's national industrial classification (NIC). There are 1,400 products in the sample for estimation.[5] Table 1 reports basic summary statistics by two-digit NIC (India's industrial classification system) sector. As a comparison, the U.S. data used by Bernard et al. (2010), contain approximately 1,500 products, defined as five-digit SIC codes across 455 four-digit SIC industries. Thus, our definition of a product is similar to earlier work that has focused on the U.S. Table 2 provides a few examples of products available in our data set. In our terminology, we will distinguish between "sectors" (which correspond to two-digit NIC aggregates), "industries" (which correspond to four-digit NIC aggregates) and "products" (the finest disaggregation we observe); we emphasize that since the "product" definition is available at a highly disaggregated level, unit values are plausibly interpreted as "prices" in our application.

The data also have some disadvantages. Unlike Census data, the CMIE database is not well suited for understanding firm entry and exit. However, Prowess contains mainly medium large Indian firms, so entry and exit is not necessarily an important margin for understanding the process of adjustment to increased openness within this subset of the manufacturing sector.[6]

We complement the production data with tariff rates from 1987 to 2001. The tariff data are reported at the six-digit Harmonized System (HS) level and were compiled by Topalova (2010). We pass the tariff data through India's input-output matrix for 1993-94 to construct input tariffs. We concord the tariffs to India's national industrial classification (NIC) schedule developed by Debroy and Santhanam (1993). Formally, input tariffs are defined as $\tau_{it}^{\text{input}} = \sum_k a_{ki} \tau_{kt}^{\text{output}}$, where $\tau_{kt}^{\text{output}}$ is the tariff on industry k at time t, and a_{ki} is the share of industry k in the value of industry i.

2.2 India's Trade Liberalization

A key advantage of our approach is that we examine the impact of openness by relying on changes in trade costs induced by a large-scale trade liberalization. India's post-independence development strategy was one of national self-sufficiency and heavy government regulation of the economy. India's trade regime was amongst the most restrictive in Asia, with high nominal tariffs and non-tariff barriers. In response to a balance-of-payments crisis, India launched a dramatic liberalization of the economy as part of an IMF structural adjustment program in August 1991. An important part of this reform was to abandon the extremely restrictive trade policies it had pursued since independence.

Several features of the trade reform are crucial to our study. First, the external crisis of 1991, which came as a surprise, opened the way for market oriented reforms (Hasan et al. (2007)).[7] The

[5] We have fewer products than in Goldberg et al. (2010b) because we require non-missing values for quantities and revenues rather than just a count of products, and drop small sectors that do not have enough observations to implement the methodology.

[6] Firms in Prowess account for 60 to 70 percent of the economic activity in the organized industrial sector and comprise 75 percent of corporate taxes and 95 percent of excise duty collected by the Government of India (CMIE).

[7] Some commentators (e.g., Panagariya (2008)) noted that once the balance of payments crisis ensued, market-

liberalization of the trade policy was therefore unanticipated by firms in India and not foreseen in their decisions prior to the reform. Moreover, reforms were passed quickly as sort of a "shock therapy" with little debate or analysis to avoid the inevitable political opposition (see Goyal (1996)). Industries with the highest tariffs received the largest tariff cuts implying that both the average and standard deviation of tariffs across industries fell. While there was significant variation in the tariff changes across industries, Topalova and Khandelwal (2011) show that tariff changes through 1997 were uncorrelated with pre-reform firm and industry characteristics such as productivity, size, output growth during the 1980s and capital intensity. The tariff liberalization does not appear to have been targeted towards specific industries and appears relatively free of usual political economy pressures until 1997 (which coincides with an election that changed political power). We estimate the production function and markups on the full sample, but restrict our analysis of the trade reform to the 1989-1997 period when trade policy did not respond to pre-existing industry- or firm-level trends. We again refer the reader to previous publications that have used this trade reform for a detailed discussion (Topalova and Khandelwal (2011); Topalova (2010); Sivadasan (2009); Goldberg et al. (2010a,b)).

3 Methodology: Recovering Markups and Marginal Costs

This section describes the framework to estimate markups and marginal costs using product- and firm-level production data. Section 3.1 presents the theoretical framework and explicitly states the assumptions required to implement the approach. The computation of markups and marginal costs requires estimates of production function coefficients and information about the allocation of inputs across products. Section 3.2 describes the methodology to estimate the production function and identification. Once the production function parameters are estimated, Section 3.3 explains how we recover the allocation of inputs across products for multi-product firms. In section 3.4 we discuss how we compute markups and marginal costs. Section 3.5 comments on the assumptions required to implement our methodology.

3.1 Theoretical Framework

Consider a production function for a firm f producing a product j at time t:

$$Q_{fjt} = F_{jt}(\boldsymbol{V}_{fjt}, \boldsymbol{K}_{fjt})\Omega_{ft} \tag{1}$$

where Q is physical output, $\boldsymbol{V}$ is a vector of variable inputs that the firm can freely adjust and $\boldsymbol{K}$ is a vector of fixed inputs that face adjustment costs. The firm's productivity is denoted Ω_{ft}. A firm produces a discrete number of products J_{ft}. Collect the inputs into a vector $\boldsymbol{X} = \{\boldsymbol{V}, \boldsymbol{K}\}$. Let

based reforms were inevitable. While the general direction of the reforms may have been anticipated, the precise changes in tariffs were not. Our empirical strategy accounts for this shift in broad anticipation of the reforms, but exploits variation in the sizes of the tariff cuts across industries.

W^v_{fjt} denote the price of a variable input v and W^k_{fjt} denote the price of a dynamic input k, with $v = \{1, ..., V\}$ and $k = \{1, ..., K\}$.

We begin by characterizing conceptual assumptions necessary to estimate markups and marginal costs for multi-product firms. We refer to these assumptions as conceptual because they are independent of the particular data and setting. Implementing the approach requires additional assumptions dictated by particular features of our data and our focus on India's trade reforms (e.g., functional form and identification assumptions), and we describe these in the next section. The approach requires the following assumptions:

Assumption 1: The production technology is product-specific. Our notation reflects this assumption. The production function $F(.)$ is indexed by product j. This assumption implies that a single-product firm and a multi-product firm that produce the same product have the same production technology, although their productivities Ω_{ft} might differ.

Assumption 2: $F_{jt}(.)$ is continuous and twice differentiable w.r.t. at least one element of $\boldsymbol{V}_{fjt}$, and this element of $\boldsymbol{V}_{fjt}$ is a static (i.e., freely adjustable or variable) input in the production of product j. This assumption restricts the technology so that the firm can adjust its output quantity by changing a particular variable input.[8] Furthermore, this assumption implies that firm cost minimization involves at least one static first order condition with respect to a variable input of production.

Assumption 3: Hicks-neutral productivity Ω_{ft} is log-additive and firm-specific. This assumption implies that a multi-product firm has the same productivity Ω_{ft} in the production of all its products.[9] This assumption follows the tradition of modeling productivity in the multi-product firm literature in this manner (e.g., Bernard et al. (2011)). For single-product firms, this assumption is of course redundant.

Assumption 4: Expenditures on all variable and fixed inputs are attributable to products. This assumption implies that we can always write the expenditure on input X attributable to product j as $W^X_{fjt}X_{fjt} = \tilde{\rho}_{fjt}\sum_j \left(W^X_{fjt}X_{fjt}\right)$ where W^X_{fjt} is the price for input X with $X \in \boldsymbol{X}$, and $\tilde{\rho}_{fjt}$ is the share of input expenditures attributable to product j with the restriction that $\sum_j \tilde{\rho}_{fjt} = 1$. Note that $\tilde{\rho}_{fjt}$ is not observed in the data. Assumption 4 allows for economies (or diseconomies) of scope in costs of production; we discuss this issue below in Section 3.5.

Assumption 5: The state variables of the firm are

$$\boldsymbol{s}_{ft} = \{J_{ft}, \boldsymbol{K}_{f,j=1,t}, ..., \boldsymbol{K}_{f,J_{ft},t}, \Omega_{ft}, \mathbf{G}_f, \boldsymbol{r}_{fjt}\}$$

[8] Assumption 2 rules out a fixed proportion technology (e.g., Leontief) in *all* variable inputs. The assumption seems reasonable at the level of aggregation of our data. We observe total labor, capital and intermediate inputs at the firm level, and so there is ample room for firms to substitute, say, workers for capital while keeping output constant.

[9] In principle, we can allow for $F_{jt}(\boldsymbol{V}_{fjt}, \boldsymbol{K}_{fjt}, \Omega_{fjt})$ to derive a theoretical expression for markups. However, assumption 3 is required to estimate markups for multi-product firms.

The state variables include the number of products produced (J_{ft}), the dynamic inputs for all products ($\boldsymbol{K}_{fjt}$), productivity (Ω_{ft}), exogenous factors (e.g., location of the firm) ($\mathbf{G}_f$),[10] and all payoff relevant serially correlated variables, such as tariffs and the firm's export status (EXP_{ft}), which we collect in $\boldsymbol{r}_{fjt}$.

Assumption 6: Firms minimize short-run costs taking output quantity and input prices $\mathbf{W}_{fjt}$ at time t as given. Firms face a vector of variable input prices $W^v_{fjt} = W^v_t(\nu_{fjt}, \mathbf{G}_f, \boldsymbol{a}_{fjt-1})$, which depends on the quality ν_{fjt} of product j, exogenous factors $\mathbf{G}_f$, and firm/product-level actions $\boldsymbol{a}_{fjt-1}$ taken prior to time t. The latter can capture pre-negotiated input prices through contracts, for example, as long as the contracts do not specify input prices as a function of current input purchase quantities (i.e., quantity discounts). The important assumption is that a firm's variable input price does not depend on input quantity. This assumption rules out static sources of market power in input markets. We discuss this assumption in more detail at the end of this subsection.

We consider the firm's cost minimization problem conditioning on state variables. From assumptions 2 and 6, firms minimize costs with respect to variable inputs. Assumptions 4 and 6 imply that costs are separable across products since a firm's product mix is a dynamic choice and pre-determined at time t when variable inputs are chosen. Hence, we can minimize costs product-by-product for multi-product firms.

The associated Lagrangian function for any product j at time t is:

$$\begin{aligned}\mathcal{L}(\mathbf{V}_{fjt}, \mathbf{K}_{fjt}, \lambda_{fjt}) &= \sum_{v=1}^{V} W^v_{fjt} V^v_{fjt} + \sum_{k=1}^{K} W^k_{fjt} K^k_{fjt} \\ &\quad + \lambda_{fjt} \left[Q_{fjt} - Q_{fjt}(\mathbf{V}_{fjt}, \mathbf{K}_{fjt}, \Omega_{ft}) \right] \end{aligned} \tag{2}$$

The first order condition for any variable input V^v used on product j, is

$$\frac{\partial \mathcal{L}_{fjt}}{\partial V^v_{fjt}} = W^v_{fjt} - \lambda_{fjt} \frac{\partial Q_{fjt}(.)}{\partial V^v_{fjt}} = 0, \tag{3}$$

where the marginal cost of production at a given level of output is λ_{fjt} since $\frac{\partial \mathcal{L}_{fjt}}{\partial Q_{fjt}} = \lambda_{fjt}$. Rearranging terms and multiplying both sides by $\frac{V_{fjt}}{Q_{fjt}}$, provides the following expression:

$$\frac{\partial Q_{fjt}(.)}{\partial V^v_{fjt}} \frac{V^v_{fjt}}{Q_{fjt}} = \frac{1}{\lambda_{fjt}} \frac{W^v_{fjt} V^v_{fjt}}{Q_{fjt}}. \tag{4}$$

The left-hand side of the above equation represents the elasticity of output with respect to variable input V^v_{fjt} (the "output elasticity"): $\theta = \frac{\partial Q_{fjt}(.)}{\partial V^v_{fjt}} \frac{V^v_{fjt}}{Q_{fjt}}$. Define the markup μ_{fjt} as $\mu_{fjt} \equiv \frac{P_{fjt}}{\lambda_{fjt}}$.

[10] In our data we only observe the location of the firms' headquarters, and not the site of production, so in practice we exclude this from the analysis. But the general framework can nevertheless account for differences in locations of firms (which may affect, for instance, exogenous spatial differences in factor prices).

The cost-minimization condition can be rearranged to express the markup for each product j as:

$$\mu_{fjt} = \theta^v_{fjt} \left(\frac{P_{fjt} Q_{fjt}}{W^v_{fjt} V^v_{fjt}} \right) = \theta^v_{fjt} (\alpha^v_{fjt})^{-1} \tag{5}$$

where α^v_{fjt} is the share of expenditure on input V^v allocated to product j in the total sales of product j. This expression forms the basis for our approach to compute markups. To compute the markup, we need the output elasticity on V^v for product j, and the share of the input's expenditure allocated to product j in the total sales of product j, α^v_{fjt}.

The expression for the markup in (5) looks similar to the one derived in De Loecker and Warzynski (2012) with one *crucial* difference: all variables are indexed by j. This seemingly small distinction has significant ramifications for the analysis and precludes us from using the existing approach in De Loecker and Warzynski (2012) to obtain the subcomponents of (5). De Loecker and Warzynski (2012) focus on firm-level markups and implement the conventional production function methodology using revenue data. Because of their focus and data, they do not need to confront the challenges posed by multi-product firms. Specifically, the firm-specific expenditures shares are directly observed in their data and the output elasticity is obtained by estimating a firm-level production function using deflated revenues. In contrast, our framework utilizes product-specific information on quantities and prices. This forces us to conduct the analysis at the product-level because aggregation to the firm-level is not possible without an explicit model of market demand.

The focus on products rather than firms calls for an explicit treatment of multi-product firms. In a multi-product setting, *both* components in equation (5) are unobserved. In contrast to a single-product firm setting, we must estimate the output elasticity separately for each product manufactured by each firm. Furthermore, the product-specific input expenditure shares α^v_{fjt} cannot be calculated from the data because firms do not report the input expenditure allocations $\tilde{\rho}_{fjt}$.[11] Our framework, presented below, confronts these two challenges by proposing a methodology for estimating production functions that explicitly deals with multi-product firms and allows one to impute the input expenditure allocations across the products of a multi-product firm.

An additional advantage of focusing on products rather than firms is that once we derive estimates of product-level markups, we can calculate marginal costs using information on product-level prices, which are observed directly in the data:

$$mc_{fjt} = \frac{P_{fjt}}{\mu_{fjt}}. \tag{6}$$

A brief discussion of the assumptions underlying the analysis is in order. Assumptions 1-5 have been explicitly or implicitly assumed throughout the literature estimating production functions.[12] For example, Assumption 1 is made implicitly whenever researchers pool single- and multi-product firm data to estimate production functions, which is almost always the case. The only difference is

[11] We are unaware of any data set that provides this information for all inputs.

[12] See Ackerberg et al. (2006) for an overview of this literature.

that the standard approach uses firm-level deflated sales and expenditure data; this practice does not force the researcher to confront multi-product firms in the data since the analysis is conducted at the firm level. Our framework strictly nests this approach, but since we use price data, and because prices are only defined at the product level (unless one is willing to make additional assumptions on demand that will allow aggregation to the firm level), we must specify physical production functions at the *product* level. We therefore explicitly state the assumptions that underlie the treatment of multi-product firms (Assumptions 1, 3 and 4).

Variants of Assumption 4 have been invoked in the few studies that have addressed the price bias in production function estimation (e.g., Foster et al. (2008) and De Loecker (2011)). Foster et al. (2008) allocate input expenditures according to revenue shares, while De Loecker (2011) allocates them based on the number of products. These variants are considerably stronger than, and are strictly nested within, Assumption 4. Relaxing these input allocation assumptions is one of the methodological contributions of this paper.

The product-by-product short-run cost minimization with respect to variable inputs in (2) follows from Assumptions 2, 4 and 6. Assumption 2 assures the existence of a variable input and is essential for our approach. If all inputs are dynamic, we can still estimate the production function, but we cannot derive markups using the approach we described above. However, the assumption that there is at least one factor of production that the firm can freely adjust over the period of a year (we have annual production data) is both plausible and standard in empirical work.

Our framework allows for economies (or diseconomies) of scope. While physical synergies in production are ruled out by Assumption 1, other forms of economies (or diseconomies) of scope are consistent with Assumptions 1 and 4. Economies of scope can operate through the Hicks-neutral productivity shocks Ω_{ft}, through pre-negotiated firm-level contracts for input prices W^v_{fjt} (as long as these input prices do not depend on quantity of inputs), and also through the spreading of fixed costs (unrelated to physical synergies in production) across multiple products in multi-product firms.[13]

Finally, an important assumption we maintain throughout the analysis is that input prices do not depend on input quantities (Assumption 6). While restrictive, this assumption is more general than the one employed in almost all production function studies, in which it is assumed that all firms face the same input prices (in contrast, we allow for input prices to differ across firms because of locational differences and/or quality differentiation). If firms have monopsony power in input markets, Assumption 6 will be violated. In this case, one can show that our approach will tend to understate the *level* of markups. However, the approach can still be used to trace and explain *changes* in markups, as long as there are no contemporaneous changes in firms' monopsony power, or, even if there are such changes, as long as changes in firms' monopsony power are uncorrelated with trade policy changes. Appendix D provides a detailed discussion of the conditions under which our approach is valid in the case of monopsony power.[14]

[13] We discuss economies of scope in more detail in Section 3.5.

[14] In principle, one could make the argument that trade policy might lead to exit of smaller, less productive firms, which might give monopsony power to the remaining firms in the market. In practice, we do not observe firm exit in our sample, so we do not consider such a scenario as a likely explanation for our empirical results. We have explored

In sum, our approach to recover estimates of markups and marginal costs requires estimates of the parameters of the production function $F_{jt}(.)$ at the product level and the input allocations $\tilde{\rho}_{fjt}$ across products within each multi-product firm. Section 3.2 discusses the production function estimation method and the identification strategy we employ in order to obtain the output elasticities for both single- and multi-product firms.

3.2 Estimation

We take logs of equation (1) and allow for log-additive measurement error and/or unanticipated shocks to output (ϵ_{fjt}). To simplify notation, and since we do not have enough data to estimate different production functions for different time periods, we assume that the production function coefficients remain constant over the sample period and drop the subscript t in the writing of the production function $f(.)$ Log output is given by: $q_{fjt} = \ln\left(Q_{fjt}\exp\left(\epsilon_{fjt}\right)\right)$. Letting $\boldsymbol{x}_{fjt}$ be the vector of (log) physical inputs, $\boldsymbol{x}_{fjt} = \{\boldsymbol{v}_{fjt}, \boldsymbol{k}_{fjt}\}$, and ω_{ft} be $\ln(\Omega_{ft})$, we obtain:

$$q_{fjt} = f_j(\boldsymbol{x}_{fjt}; \boldsymbol{\beta}) + \omega_{ft} + \epsilon_{fjt}. \tag{7}$$

By writing the production function in terms of physical output rather than revenue, we exploit separate information on quantities and prices that is available in the data. The use of physical output in equation (7) eliminates concerns of a price bias that arises if output is constructed by deflating firm revenues by an industry-level price index.[15]

Unobserved productivity ω_{ft} potentially leads to well known simultaneity and selection biases. These two biases have been the predominant focus of the production function estimating literature and we follow the insights of Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2006) in addressing them. Note that if we theoretically had data on the physical inputs $(\boldsymbol{v}_{fjt}, \boldsymbol{k}_{fjt})$ for all products, these existing approaches to estimating production functions would in principle suffice to obtain consistent estimates of the production function coefficients $\boldsymbol{\beta}$.

In reality, no dataset records product-specific inputs, so estimating equation (7) requires dealing with two additional issues: (a) we do not observe input allocations across products in multi-product firms; and (b) we observe industry-wide deflated firm-level input expenditures rather than firm-level input quantities. The latter is not merely a measurement problem because firms typically rely on differentiated inputs to manufacture differentiated products, so physical input and output are not readily comparable across firms.

To understand the implications of these two issues for estimation, let $\tilde{\boldsymbol{x}}_{ft}$ denote the (observed) vector of deflated input expenditures, deflated by a sector-specific price index. From Assumption

heterogeneity in our results by identifying business groups in our sample who may have some degree of monopsony power, but we do not find differential effects with respect to the impacts of tariffs on their prices, markups and marginal costs (results available upon request).

[15] For a detailed discussion, see De Loecker (2011) and Foster et al. (2008).

4, product-level input quantities, x_{fjt}, for each input x relate to firm-level expenditures as follows:

$$x_{fjt} = \rho_{fjt} + \tilde{x}_{ft} - w^x_{fjt} \tag{8}$$

where $\rho_{fjt} = \ln \tilde{\rho}_{fjt}$ is the (log) share of firm input expenditures allocated to product j and w^x_{fjt} denotes the deviation of the unobserved (log) firm-product-specific input price from the (log) industry-wide input price index.[16] By substituting this expression for physical inputs into equation (7) and defining $\boldsymbol{w}_{fjt}$ as the vector of log firm-product-specific input prices, we obtain:[17]

$$q_{fjt} = f_j(\tilde{\boldsymbol{x}}_{ft}; \boldsymbol{\beta}) + A(\rho_{fjt}, \tilde{\boldsymbol{x}}_{ft}, \boldsymbol{\beta}) + B(\boldsymbol{w}_{fjt}, \rho_{fjt}, \tilde{\boldsymbol{x}}_{ft}, \boldsymbol{\beta}) + \omega_{ft} + \epsilon_{fjt} \tag{9}$$

Compared to equation (7), there are two additional unobserved terms in (9). First, the term $A(.)$ that arises from the unobserved product-level input allocations ρ_{fjt} and second, the term $B(.)$ that captures unobserved firm-product-specific input prices $\boldsymbol{w}_{fjt}$. The exact form of terms $A(.)$ and $B(.)$ depends on the functional form of $f(.)$. Both terms depend on the vector of coefficients $\boldsymbol{\beta}$, the input expenditures $\tilde{\boldsymbol{x}}_{ft}$, and the unobserved product-level input allocation shares ρ_{fjt}. It is evident from (9) that even after controlling for the unobserved productivity ω_{ft} using standard estimation techniques, the presence of the terms $A(.)$ and $B(.)$ leads to biased production function coefficients since both terms are correlated with the deflated input expenditures $\tilde{\boldsymbol{x}}_{ft}$. We refer to the bias arising from the term $A(.)$ as the "input allocation" bias and the bias arising from $B(.)$ as the "input price" bias. The methodology we develop in this subsection addresses these biases.

Neither the "input allocation" nor the "input price" bias have received much attention in the literature on production function estimation to date because the standard practice regresses deflated sales on deflated expenditures at the firm level.[18] De Loecker and Goldberg (2014) discuss the conditions under which these biases interact so as to produce reasonable estimates. But although such estimates may look plausible, this does not imply that the coefficients are consistent estimates of the production function. Failing to correct these biases traces the elasticity of sales with respect to input expenditures, but that elasticity is not useful in our approach because equation (5) requires the elasticity of output quantities with respect to input quantities.

To deal with these biases, we proceed in four steps. Subsection 3.2.1 explains how the estimation addresses the unobserved input allocation bias. Subsection 3.2.2 explains how to address the bias arising from unobserved input prices. Subsection 3.2.3 explains our treatment of the unobserved productivity shock and selection correction. Subsection 3.2.4 explains the moment conditions and further elaborates on identification and estimation. The first two steps are new to the literature on

[16] We allow for multi-product firms to face different input prices in the production of their various products. Accordingly, the input prices w are indexed by both f and j. This would be the case if a multi-product firm manufactured products of different qualities that relied on inputs of different qualities; see subsection 3.2.2 for a discussion of the relationship between output and input quality.

[17] To simplify notation, we will always use $\boldsymbol{w}_{fjt}$ to denote the deviations of firm-product-specific input prices from industry input price indexes. Similarly, from now on, we will use the term "firm input prices" to denote firm-specific deviations from industry averages.

[18] Katayama et al. (2009) is the only study to our knowledge that acknowledges the existence of the input price bias.

production function estimation; the last two steps build on existing work.

3.2.1 Unobserved Input Allocations: The Use of Single-Product Firms

Assumptions 1 and 4 imply that a firm f's technology used to produce product j is independent of the other products manufactured by the firm. This also implies that a multi-product firm uses the same technology as a single-product firm producing the same product.[19] We can therefore rely on single-product firms to estimate the product-level production function in (9), without having to address the unobserved input allocations in multi-product firms. For single-product firms, $A(\cdot) = 0$ because by definition, $\tilde{\rho}_{fjt} = 1$. Since estimation is based on the single-product sample, we omit the product subscript j for the remainder of the exposition of the estimation algorithm.

Equation (9) simplifies to:

$$q_{ft} = f(\tilde{\boldsymbol{x}}_{ft}; \boldsymbol{\beta}) + B(\boldsymbol{w}_{ft}, \tilde{\boldsymbol{x}}_{ft}, \boldsymbol{\beta}) + \omega_{ft} + \epsilon_{ft}. \tag{10}$$

The approach of using the single-product firm estimates to infer the production function coefficients for all firms raises the concern that the estimates may suffer from a selection bias since we rely only on single-product firms in the estimation. The selection bias arises if firms' choice to add a second product and become multi-product depends on the unobserved firm productivity ω_{ft} and/or firms' input use. Our estimation procedure utilizes the selection correction insights from Olley and Pakes (1996) to address this potential selection bias in two ways. First, we use an unbalanced panel that consists of firms that are single-product *at a given point in time*. At time t, the unbalanced panel includes both firms who always remain single-product firms and those that manufacture a single product at t but add additional products at a later date. This feature of the sample is important since many firms start off as single-product firms and add products during our sample. The use of the unbalanced panel is helpful in addressing the selection concern arising from the non-random event that a firm becomes a multi-product producer based on unobserved productivity ω_{ft}.[20] Second, to account for the possibility that the productivity threshold determining the transition of a firm from single- to multi-product status is correlated with production inputs (in particular, capital), we additionally apply a sample selection correction procedure. We describe the details of the sample selection correction procedure in subsection 3.2.3.[21]

We consider three inputs in the (deflated) input expenditure vector $\tilde{\boldsymbol{x}}_{ft}$: labor ($\tilde{l}$), intermediate

[19] For example, imagine a single-product firm produces a t-shirt using a particular technology, and another single-product firm produces carpets using a different combination of inputs. We assume that a multi-product firm that manufactures both products will use each technology on its respective product, rather than some third technology.

[20] This non-random event of adding a second product results in a sample selection issue analogous to the non-random exit of firms discussed in Olley and Pakes (1996). In their context, Olley and Pakes (1996) are concerned about the left tail of the productivity distribution; here, a balanced panel of single-product firms would censor the right tail of the productivity distribution. The use of the unbalanced panel of single-product firms improves upon this selection problem.

[21] Firms in our sample very rarely drop products, so we do not observe the reverse transition from multi- to single-product status. We refer the reader to Goldberg et al. (2010b) for a detailed analysis of product adding and dropping in our data. Unlike Olley and Pakes (1996), we are also not concerned with firm exit. Firm exit is rare in our data because Prowess covers the medium and large firms in India.

inputs ($\tilde{m}$) and capital ($\tilde{k}$). It is clear from equation (10) that we still need to correct for the term related to unobserved firm-specific input price variation, $B(\boldsymbol{w}_{ft}, \tilde{\boldsymbol{x}}_{ft}, \boldsymbol{\beta})$ and the unobserved firm-level productivity (ω_{ft}) in order to obtain consistent estimates of the production function parameters $\boldsymbol{\beta}$, and hence the output elasticities that are used to compute markups and marginal costs. We turn to these issues next.

3.2.2 Unobserved Input Prices

The treatment of unobserved input prices is important for two reasons. First, we need to control for them in $B(\boldsymbol{w}_{ft}, \tilde{\boldsymbol{x}}_{ft}, \boldsymbol{\beta})$ in equation (10) to recover consistent estimates of the production function parameters $\boldsymbol{\beta}$.[22] Second, the input demand equation that is used to control for productivity ω_{ft} naturally depends on input prices (see next subsection 3.2.3).

In our framework (see Assumption 6), firm-specific input price variation can arise through exogenous variation in input prices across local input markets (G_f) and/or variation in input quality (ν_{ft}).[23] This implies that two firms in the same industry that produce in the same location only face the exact same input prices if they buy the exact same input quality. We propose an approach to control for unobserved input price variation across firms using information on observables, particularly (but not exclusively) output prices. The intuition is that output prices contain information about input prices. For example, using data from Colombia that uniquely record price information for both inputs and outputs, Kugler and Verhoogen (2011) document that producers of more expensive products also use more expensive inputs.

We provide a formal model that rationalizes our approach to control for input prices in Appendix A. We show that in a large class of models of consumer demand and imperfect competition used in the Industrial Organization and International Trade literatures, we can proxy for unobserved input prices using a function of the firm's output price, market share, and product dummies. Here, we sketch the main argument and provide the economic intuition underlying our empirical strategy.

We define product quality as the mean utility associated with consuming a product net of price. Product quality can be modeled as a function of observable and unobservable product characteristics. Intuitively, our quality concept encompasses all attributes that increase the utility consumers receive from consuming the product, conditional on its price. The main premise of our correction procedure is that manufacturing high-quality products requires high-quality inputs, and that high-quality inputs are expensive. We further assume *complementarity* in input quality: manufacturing high-quality products requires combining high-quality materials with high-quality labor and capital. This is a common assumption in the literature and underlies 'O-Ring'-type theories of production (e.g., Kremer (1993), Verhoogen (2008) and Kugler and Verhoogen (2011)). This complementarity implies that the prices of *all* inputs facing a firm can be expressed as functions of a single index of product

[22] This subsection considers single-product firms since we use only these firms to estimate the production functions, but all relationships described below also apply to multi-product firms (in which case all relevant variables should be indexed by j).

[23] We abstract from lagged action variables $\boldsymbol{a}_{ft-1}$, since we do not have rich enough data to measure these (e.g., past contracts specifying input prices independent of quantities).

quality. We assume that all firms producing the same product category (e.g., apparel) face the same production function for quality, but allow the production function for quality to differ across product categories (e.g., between apparel and food products). Appendix A shows that input prices are an increasing function of product quality in this setting. Accordingly, we can control for input price variation across firms using differences in output quality across firms.

Given that input prices are an increasing function of input quality, which is an increasing function of output quality, we can use the variables proxying for output quality (i.e., output price, market share and product dummies) to proxy for input prices. Formally, we write input prices w_{ft}^x as a function of output quality ν_{ft} and firm location $\boldsymbol{G}_f$:[24]

$$w_{ft}^x = w_t(\nu_{ft}, \boldsymbol{G}_f). \tag{11}$$

This expression for input prices generalizes Assumption 6 to all inputs. Appendix A shows that the input price control function w_t will generally be input-specific (so it should be indexed by x). As we discuss in Appendix A and elaborate in section 3.5.2, allowing for input-specific input price control functions always allows one to identify the coefficients of the production function $\boldsymbol{\beta}$. However, in this general case, one will not be able to identify the coefficients of the input price control function, which are needed in our application to compute the input allocations ρ_{fjt} (and markups) for multi-product firms in sections 3.3 and 3.4. Therefore, we impose the same control function w_t across all inputs.

Using the results from Appendix A we get:

$$w_{ft}^x = w_t(p_{ft}, \boldsymbol{ms}_{ft}, \boldsymbol{D}_f, \boldsymbol{G}_f, EXP_{ft}), \tag{12}$$

where p_{ft} is the output price of the firm, $\boldsymbol{ms}_{ft}$ is a vector of market shares, $\boldsymbol{D}_f$ captures the vector of product dummies, and EXP_{ft} denotes the export status of a firm.[25] It is important to note that our approach to control for unobserved input quality does not assume that products are only vertically differentiated. It allows for horizontal differentiation, but horizontal differentiation is costless. In contrast, differentiation along the vertical dimension requires higher quality inputs that have higher input prices. This assumption is common in trade models (e.g., Verhoogen (2008) and Khandelwal (2010)). Moreover, because we model output quality as a flexible function of output prices, market share, and product dummies, the approach does not require us to commit to a particular demand function since it encompasses a large class of demand models used in the literature. For example, in a purely vertical differentiation model, there is a one-to-one mapping between product quality and product prices, so output prices perfectly proxy for quality; in this case, one would not require

[24] We remind the reader that we have defined the input price w_{ft}^x for input x as the deviation of the actual input price from the relevant input price index (i.e., the weighted industry mean), and therefore $w_{ft}^x = 0$ for the producer paying exactly the (weighted) average $\bar{w}_t^x$. Formally $w_{fjt}^x = w_{fjt}^{x*} - \bar{w}_{jt}^x$, where * denotes the actual input price faced by firm f for its product j at time t.

[25] We include the export status of a firm to allow for market demand conditions in export destinations to differ from the domestic market. In our data we do not observe the product-destination trade flows for each firm. Otherwise this information could be included here.

controls for market share or product characteristics. In the simple logit model, quality is a function of output prices and market shares (see Khandelwal (2010) for a detailed exposition). In more general models, such as the nested logit or random coefficients models, quality is a function of additional variables, such as product characteristics, conditional market shares, etc. While product characteristics are typically not observed in manufacturing surveys, product dummies can proxy for the unobserved product characteristics (as long as these do not change over time) and accommodate these more general demand specifications as in Berry (1994). Finally, using output prices as a proxy for quality does not imply that we assume complete pass-through of input to output prices; the degree of pass-through is dictated by the (unspecified) underlying demand and market structure and by the firm behavioral assumptions. Accordingly, the approach is consistent with *any* degree of pass-through between input and output prices.

The final step is to substitute the input price control function from (12) into the expression for w_{ft} in $B(\boldsymbol{w}_{ft}, \tilde{\boldsymbol{x}}_{ft}, \boldsymbol{\beta})$ in equation (10), we get:

$$B(\boldsymbol{w}_{ft}, \tilde{\boldsymbol{x}}_{ft}, \boldsymbol{\beta}) = B((p_{ft}, \boldsymbol{ms}_{ft}, \boldsymbol{D}_f, \boldsymbol{G}_f, EXP_{ft}) \times \tilde{\boldsymbol{x}}^c_{ft}; \boldsymbol{\beta}, \boldsymbol{\delta}) \tag{13}$$

A few words on notation are in order. The function $B(.)$ is different from the input price function $w(.)$ as described in equation (12). The function $B(.)$ depends on the input prices w_{ft} and will therefore take as arguments the elements of $w(.)$. However, it also contains interactions of the input prices (w_{ft}) with the vector of deflated input expenditures $\tilde{\boldsymbol{x}}_{ft}$. We use the notation $\tilde{\boldsymbol{x}}^c_{ft}$ to highlight the fact that the input price term $w(.)$ enters also by itself, without being interacted with the input expenditures $\tilde{\boldsymbol{x}}_{ft}$, and thus we include a constant term: $\tilde{\boldsymbol{x}}^c_{ft} = \{1, \tilde{\boldsymbol{x}}_{ft}\}$. The notation highlights that the use of the input price control function requires us to estimate an additional parameter vector $\boldsymbol{\delta}$ alongside the production function parameters $\boldsymbol{\beta}$.

3.2.3 Unobserved Productivity and Selection Correction

The only remaining source of potential bias in (10) is the unobserved firm-level productivity ω_{ft}. Firms' choices of inputs and number of products are in part affected by this (to the econometrician) unobserved productivity, potentially leading to simultaneity and selection bias in estimation. We control for unobserved productivity ω_{ft} in (10) using a control function based on a static input demand equation. In addition, we implement a selection correction for the potential selection bias stemming from the use of single-product firms in the estimation procedure, discussed in subsection 3.2.1. We describe both procedures here.

We follow the literature on production function estimation, as initiated by Olley and Pakes (1996) and extended by Levinsohn and Petrin (2003), and control for unobserved productivity ω_{ft} in (10) using a static input demand equation. The materials demand function in our setting will take as arguments all state variables of the firm noted in Assumption 5, including productivity, and all additional variables that affect a firm's demand for materials. These include firm location ($\boldsymbol{G}_f$), output prices (p_{ft}), product dummies ($\boldsymbol{D}_f$), market shares ($\boldsymbol{ms}_{ft}$), input prices ($w_t(.)$), the export

status of a firm (EXP_{ft}) and the input (τ_{it}^{input}) and output tariffs ($\tau_{it}^{\text{output}}$) that the firm faces on the product it produces. From (12) input prices are themselves a function of output price, market share and product dummies[26], so materials demand is given by:

$$\tilde{m}_{ft} = m_t(\omega_{ft}, \tilde{k}_{ft}, \tilde{l}_{ft}, \boldsymbol{G}_f, p_{ft}, \boldsymbol{D}_f, \boldsymbol{ms}_{ft}, EXP_{ft}, \tau_{it}^{\text{input}}, \tau_{it}^{\text{output}}). \tag{14}$$

We collect all the variables determining intermediate input demand, except for the input expenditures and unobserved productivity, in $\boldsymbol{z}_{ft} = \{\boldsymbol{G}_f, p_{ft}, \boldsymbol{D}_f, \boldsymbol{ms}_{ft}, EXP_{ft}, \tau_{it}^{\text{input}}, \tau_{it}^{\text{output}}\}$. The number of products (J_{ft}) is omitted from the set of state variables since the sample we use for estimation contains only single-product firms. The subscript i on the tariff variables denotes an industry to indicate that tariffs vary at a higher level of aggregation than products. Inverting (14) gives our control function for productivity:[27]

$$\omega_{ft} = h_t(\tilde{\boldsymbol{x}}_{ft}, \boldsymbol{z}_{ft}). \tag{15}$$

Our approach also encompasses a selection correction to address the potential selection bias stemming from the use of only single-product firms in the estimation discussed in subsection 3.2.1. The selection bias arises if a firm's choice to add a second product and become a multi-product firm depends on unobserved firm productivity ω_{ft} in equation (10) and/or the firm's input use. Following Olley and Pakes (1996), who address the selection bias due to plant exit in their setting, we model the probability that a firm continues to produce one product non-parametrically as a function of the firm's productivity forecast and all state variables $\boldsymbol{s}_{ft}$.

The underlying model behind our sample selection correction is one where the number of products manufactured by firms increases with productivity. Several multi-product firm models generate this correlation, with Mayer et al. (2014) matching our setup most closely. In that model, the number of products a firm produces is an increasing step function of the firms' productivity. Firms have a productivity draw which determines their core product. Conditional on entry, the firm produces this core product and incurs an increasingly higher marginal cost of production for each additional product it manufactures. This structure generates a competence ladder that is characterized by a set of cutoff points, each associated with the introduction of an additional product.[28]

[26] Note that we consider (log) intermediate input expenditure, defined as the sum (in logs) of the intermediate input demand and the input price. This implies that the materials expenditure function $\tilde{m}_t(.)$ takes as arguments the same variables as the physical materials demand function $m_t(.)$: $m_{ft} = m_t(w_{ft}^m, .)$ and $\tilde{m}_{ft} = m_t(.) + w_{ft}^m = \tilde{m}_t(w_{ft}^m, .)$, where w_{ft}^m is the input price.

[27] As discussed in Olley and Pakes (1996), the proxy approach does not require knowledge of the market structure for the input markets; it simply states that input demand depends on the firm's state variables and variables affecting input demand. By using a static control to proxy for productivity, we do not have to revisit the underlying dynamic model and prove invertibility when modifying Olley and Pakes (1996) for our setting to include additional state variables (e.g., tariffs). See De Loecker (2011) and Ackerberg et al. (2006) for an extensive discussion. A recent literature has discussed alternative estimation procedures that do not rely on this inversion. In the absence of shocks to output ϵ_{ft}, these procedures can be implemented without additional assumptions. However, the ϵ_{ft} shocks end up being important, especially when estimating physical output production functions, where the $\epsilon_{ft}'s$ absorb unit fixed effects.

[28] Alternative models such as Bernard et al. (2010) introduce firm-product-specific demand shocks that generate product switching (e.g., product addition and dropping) in each period. We avoid this additional complexity since

The cutoff point relevant to our sample selection procedure is the one associated with the introduction of a *second* product. We denote this cutoff by $\bar{\omega}_{ft}$. Firms with productivity that exceeds $\bar{\omega}_{ft}$ are multi-product firms that produce two (or more) products while firms below $\bar{\omega}_{ft}$ remain single-product producers and are included in the estimation sample.

If the threshold $\bar{\omega}_{ft}$ is independent of the right-hand side variables in the production function in equation (10), there is no selection bias and we obtain consistent estimates of production function coefficients (as long as we use the unbalanced panel of single product firms, i.e., the sample of firms that are single-product at any point in time, but may become multi-product in the future). A bias arises when the threshold is a function of capital and/or labor. For example, it is possible that even conditional on productivity, a firm with more capital finds it easier to finance the introduction of an additional product; or, a firm that employs more workers may have an easier time expanding into new product lines. In these cases, firms with more capital and/or labor are less likely to be single-product firms, even conditional on productivity, and this generates a negative bias in the capital and labor coefficients.

To address the selection bias, we allow the threshold $\bar{\omega}_{ft}$ to be a function of the state variables $\boldsymbol{s}_{ft}$ and the firm's information set at time $\mathcal{I}_{t-1}$ (we assume the decision to add a product is made in the previous period). The selection rule requires that the firm make its decision to add a product based on a forecast of these variables in the future. Define an indicator function χ_{ft} to be equal to 1 if the firm remains single-product (SP) and 0 otherwise. The selection rule can be written as:

$$\begin{aligned}
\Pr(\chi_{ft}=1) &= \Pr\left[\omega_{ft} \leq \bar{\omega}_{ft}(\boldsymbol{s}_{ft}) | \bar{\omega}_{ft}(\boldsymbol{s}_{ft}), \omega_{ft-1}\right] && (16)\\
&= \kappa_{t-1}(\bar{\omega}_{ft}(\boldsymbol{s}_{ft}), \omega_{ft-1}) \\
&= \kappa_{t-1}(\tilde{\boldsymbol{x}}_{ft-1}, i_{ft-1}, \boldsymbol{z}_{ft-1}) && (17)\\
&\equiv SP_{ft}
\end{aligned}$$

Note that the variables included in $\boldsymbol{z}$ are a subset of the state variables that appear in $\boldsymbol{s}$ (the latter include the dynamic inputs that are part of $\tilde{\boldsymbol{x}}$). We use the fact that the threshold at t is predicted using the firm's state variables at $t-1$, the accumulation equation for capital, and $\omega_{ft} = h_t(\tilde{\boldsymbol{x}}_{ft}, \boldsymbol{z}_{ft})$ from equation (15) to arrive at the last equation.[29] As in Olley and Pakes (1996), we have two different indexes of firm heterogeneity, the productivity and the productivity cutoff point. Note that $SP_{ft} = \kappa_{t-1}(\omega_{ft-1}, \bar{\omega}_{ft})$ and therefore $\bar{\omega}_{ft} = \kappa_{t-1}^{-1}(\omega_{ft-1}, SP_{ft})$.

product dropping is not a prominent feature of our data (Goldberg et al. (2010b)). Moreover, in Section 4 we find strong support that firms' marginal costs are lower on their core competent products (products that have higher sales shares).

[29] The accumulation equation for capital is: $K_{ft} = (1-\delta)K_{ft-1} + I_{ft-1}$, where δ is the depreciation rate of capital. The specification of the selection rule takes into account that firms hire and/or fire workers based on their labor force at time $t-1$ and their forecast of future demand and costs captured by $\boldsymbol{z}$ and ω. So all variables entering the nonparametric function $\kappa_{t-1}(.)$ help predict the firm's employment at time t.

3.2.4 Productivity Process, Moment Conditions, and Identification

To estimate the parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$, we follow Ackerberg et al. (2006) and form moments based on the innovation in the productivity shock ξ_{ft}. We consider the following law of motion for productivity:

$$\omega_{ft} = g(\omega_{ft-1}, \tau_{it-1}^{\text{output}}, \tau_{it-1}^{\text{input}}, EXP_{ft-1}, SP_{ft}) + \xi_{ft}. \tag{18}$$

The tariff variables and export dummy are included in the law of motion to account for the fact that trade policy and exporting *may* affect productivity. As De Loecker (2013) shows, if one expects these variables to have an effect on productivity, then the theoretically consistent treatment is to include them directly in the law of motion. Otherwise, their omission may lead to biased production function coefficients. Of course, the fact that these variables are allowed to have an impact on productivity does not mean that they will in fact have an effect. It is entirely possible that the empirical estimates indicate that the trade variables have no effect on productivity. Hence, including trade variables in the law of motion does not assume a particular result regarding the effects of tariffs or exporting on productivity.

Trade related variables are expected to affect productivity both through exporting and importing channels. For example, a large literature suggests "learning by exporting" effects. Likewise, trade economists have postulated that a reduction in output tariffs that exposes firms to intensified import competition may lead to reduction in X-inefficiencies and adoption of better management practices. In this case, output tariff reductions may lead to productivity improvements. On the input side, input tariff reductions may lead to the import of new, previously unavailable intermediate products, which will lead to increases in productivity (see Halpern et al. (2011) for a formalization of this argument). We emphasize that the specification we adopt for the law of motion for productivity in equation (18) *allows* for these mechanisms to generate productivity improvements, but by no means assumes the result. The inclusion of the probability that a firm remains single-product in the next period SP_{ft} in the law of motion addresses the selection correction from equation (16). In principle, there could be additional variables that affect firm productivity (e.g., a firm's R&D), but we do not include those in the law of motion as we have no information on them in our data.

To form moments based on the innovation in the productivity shock in (18), one needs to express the productivity ω_{ft} as a function of data and parameters. Plugging the expressions for the input price correction from (13) and for unobserved productivity from (15) into the production function equation (10), we get:

$$q_{ft} = \phi_t(\tilde{\boldsymbol{x}}_{ft}, \boldsymbol{z}_{ft}) + \epsilon_{ft}, \tag{19}$$

where we remind the reader that the vector $\boldsymbol{z}_{ft}$ includes all variables that affect intermediate input demand, except for the input expenditures and unobserved productivity:

$$\boldsymbol{z}_{ft} = \{\boldsymbol{G}_f, p_{ft}, \boldsymbol{D}_f, \boldsymbol{ms}_{ft}, EXP_{ft}, \tau_{it}^{\text{input}}, \tau_{it}^{\text{output}}\},$$

an the term $\phi_t(.)$ is equal to $f(\tilde{\boldsymbol{x}}_{ft}; \boldsymbol{\beta}) + B(\boldsymbol{w}_{ft}, \tilde{\boldsymbol{x}}_{ft}, \boldsymbol{\beta}) + \omega_{ft}$ and captures output net of noise ϵ_{ft}.

Estimation of (19) enables one to get rid of unanticipated shocks and/or measurement error ϵ_{ft}. We note that although the variables proxying for input prices (see equation (12)) also enter the input demand equation in equation (15), this has no implications for the identification of the production function parameters. The only purpose of the first stage estimation is to purge the output quantity data from unanticipated shocks and/or measurement error (i.e., purge ϵ_{ft} in equation (10)).[30] For example, output prices (p_{ft}) enter this first stage both to control for unobserved productivity and input price differences, but we do not need to distinguish between them when forecasting output. Note that even if we observed (quality-corrected) input prices, we would still include output prices and the function $\phi_t(.)$ would reflect this.

The first stage of the estimation in (19) yields an estimate of predicted output $\hat{\phi}_{ft}$.[31] One can then express productivity ω_{ft} as a function of data and parameters. In particular, using equations (10), (13) and (19) we have:

$$\omega_{ft}(\boldsymbol{\beta}, \boldsymbol{\delta}) = \hat{\phi}_{ft} - f(\tilde{\boldsymbol{x}}_{ft}; \boldsymbol{\beta}) - B((p_{ft}, \boldsymbol{ms}_{ft}, \boldsymbol{D}_f, \boldsymbol{G}_f, EXP_{ft}) \times \tilde{\boldsymbol{x}^c}_{ft}; \boldsymbol{\delta}), \tag{20}$$

where the last term, the function $B(.)$, represents the input price control function.[32]

It is important to note that even though the input expenditures $\tilde{\boldsymbol{x}}_{ft}$ enter both the production function $f(.)$ and the input price control function $B(.)$, the coefficients of the production function $\boldsymbol{\beta}$ are identified because $\tilde{\boldsymbol{x}}_{ft}$ enter the input price control function in (13) only interacted with input prices, or put differently, the input expenditures do not enter the input price function $w(.)$ in (12). This identification insight does not rest on any functional form assumptions; it results from the fact that the control function for quality, and hence input prices, rests on the demand side alone and hence does not include input expenditures.

The main parameters of interest to compute markups are the vector of production function coefficients $\boldsymbol{\beta}$. However, from (13), note that the parameter vector $\boldsymbol{\delta}$ allows us to identify the input prices: after we have estimated $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$, we can recover the input prices from equation (12).[33]

To estimate the parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$, we form moments based on the innovation in the productivity shock ξ_{ft} in law of motion in equation (18). We use (20) to project $\omega_{ft}(.)$ on the

[30] We could set $\epsilon_{ft} = 0$; in this case, we no longer need to invert the input demand function to control for unobserved productivity. However, we feel that the input demand specification addresses first-order empirical issues with the data: measurement error in output and differences in units across products within sectors, which are absorbed by unit fixed effects in the first stage.

[31] In practice we approximate the function $\phi_t(.)$ with a third-order polynomial in all its elements, with the exception of product dummies. We add the product dummies linearly to avoid having to estimate all cross terms. This seems innocuous since the first stage R^2 is very close to one.

[32] We approximate $B(.)$ with a flexible third-order polynomial. At this point the reader might find it useful to consider a special case of a Cobb-Douglas production function and a vertical differentiation model of consumer demand. In this special case equation (20) reduces to: $\omega_{ft}(\boldsymbol{\beta}, \boldsymbol{\delta}) = \phi_{ft} - \tilde{\boldsymbol{x}}'_{ft}\boldsymbol{\beta} - \Gamma w_t(p_{ft}; \boldsymbol{\delta})$, where Γ denotes the returns to scale parameter. Please see Appendix B for details.

[33] In other words, we specify the function $w(.)$ and therefore the $\boldsymbol{\delta}$ parameters are a function of both the production function coefficients $\boldsymbol{\beta}$, and the parameters in $w(.)$. It is at this stage where we need the assumption that the function $w(.)$ does not vary across inputs. If we allowed for input-specific $w(.)$ functions, we would still be able to consistently estimate the parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$, but we would not be able to identify the input-specific coefficients of the $w(.)$ functions from $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$. See Appendix B for a more detailed discussion based on a Cobb-Douglas production function.

elements of $g(.)$ to obtain the innovation ξ_{ft} as a function of the parameters $\xi_{ft}(\boldsymbol{\beta}, \boldsymbol{\delta})$:

$$\xi_{ft}(\boldsymbol{\beta}, \boldsymbol{\delta}) = \omega_{ft}(\boldsymbol{\beta}, \boldsymbol{\delta}) - E\left(\omega_{ft}(\boldsymbol{\beta}, \boldsymbol{\delta}) | \omega_{ft-1}(\boldsymbol{\beta}, \boldsymbol{\delta}), \tau_{it-1}^{\text{output}}, \tau_{it-1}^{\text{input}}, EXP_{ft-1}, SP_{ft}\right) \tag{21}$$

The moments that identify the parameters are:

$$E\left(\xi_{ft}(\boldsymbol{\beta}, \boldsymbol{\delta}) \mathbf{Y}_{ft}\right) = 0, \tag{22}$$

where $\boldsymbol{Y}_{ft}$ contains lagged materials, current capital and labor, and their higher order and interaction terms, as well as lagged output prices, lagged market shares, lagged tariffs, and their appropriate interactions with the inputs.

This method identifies the production function coefficients by exploiting the fact that current shocks to productivity will immediately affect a firm's materials choice while labor and capital do not immediately respond to these shocks; moreover, the degree of adjustment can vary across firms and time. These moments that rely on adjustment costs in inputs are by now standard in this literature. In our context, we assume that firms freely adjust materials and treat capital and labor as dynamic inputs that face adjustment costs. In other settings, one may choose to treat labor as a flexible input. Since materials are the flexible input, we use lagged materials when we construct moments.[34]

We use lagged output prices, market shares, and tariffs and their interactions with appropriately lagged inputs to form additional moment conditions to identify jointly the production function coefficients $\boldsymbol{\beta}$ and the coefficients $\boldsymbol{\delta}$ capturing the input price variation. For example, the parameter related to the output price is identified off the moment $E(\xi_t p_{t-1}) = 0$; this moment condition is based on the insight that current prices do react to productivity shocks, so we need to use lagged output prices which exploit the serial correlation of prices.

We estimate the model using a GMM procedure on a sample of firms that manufacture a single product for at least three consecutive years.[35] We choose three years since the moment conditions require at least two years of data because of the lagged values; we add an additional (third) year to allow for potential measurement error in the precise timing of a new product introduction. We discuss the timing assumptions further in subsection 3.5.2. In principle, one could run the estimation separately for each product. In practice, our sample size is too small to allow estimation at the product level, so we estimate (10) at the two-digit sector level.[36]

Estimation of equation (10) requires choosing a functional form for f. We adopt a translog specification because of its flexibility.[37] Specifically, the translog offers the advantage that it generates output elasticities that are not constant over time and across firms (though the production coeffi-

[34] In our setting, input tariffs are serially correlated and since they affect input prices, input prices are serially correlated over time, creating a link between current and lagged intermediate input usage.

[35] We follow the procedure suggested by Wooldridge (2009) that forms moments on the joint error term $(\xi_{ft} + \epsilon_{ft})$.

[36] This follows the standard practice in the literature where production functions are estimated at the industry level. For example, see Levinsohn and Petrin (2003).

[37] The translog production function is $q_{ft} = \beta_l l_{ft} + \beta_{ll} l_{ft}^2 + \beta_k k_{ft} + \beta_{kk} k_{ft}^2 + \beta_m m_{ft} + \beta_{mm} m_{ft}^2 + \beta_{lk} l_{ft} k_{ft} + \beta_{lm} l_{ft} m_{ft} + \beta_{mk} m_{ft} k_{ft} + \beta_{lmk} l_{ft} m_{ft} k_{ft} + \omega_{ft}$.

cients are constrained to be the same across years and firms); hence, large firms can have different elasticities than small firms. The exact functional form for $f(.)$ does not generate any identification results. The crucial assumption is that productivity enters in a log-additive fashion (Assumption 3 in Section 3.1).

Finally, the standard errors on the coefficients are obtained using block-bootstrapping, where we draw an entire firm time series. Since our ultimate objective is to estimate the impact of the trade reforms on markups and marginal costs, we correct the standard errors of the regressions in Section 4 by block-bootstrapping over our entire empirical procedure.

3.3 Recovering Input Allocations

As shown in equations (5) and (6), computing markups and marginal costs requires the product-specific output elasticity and product-specific revenue shares on a variable input (in our case, materials). We obtain the output elasticity from the estimation outlined in Section 3.2 based on single-product firms, but we do not know the product-specific revenue shares of inputs for multi-product firms. Here, we show how to compute the input allocations across products of a multi-product firm in order to construct α^M_{fjt}.

From Assumption 6, recall that $\rho_{fjt} = \ln\left(\frac{W^X_{fjt}X_{fjt}}{\widetilde{X}_{ft}}\right)\ \forall X \in \{V, K\}$, is product j's input cost share. We solve for ρ_{fjt} in multi-product firms as follows. We first eliminate unanticipated shocks and measurement error from the product-level output data by following the same procedure as in the first stage of our estimation routine for the single-product firms in (19). We project q_{fjt} on the exact same variables used in the first stage of the estimation procedure, $\widehat{q}_{fjt} \equiv E\left(q_{fjt}|\phi_t\left(\tilde{\boldsymbol{x}}_{ft}, \boldsymbol{z}_{ft}\right)\right)$, which allows us to eliminate any measurement error and unanticipated shocks to output from the recorded output data.

Given the aforementioned assumptions that productivity is firm-specific and log-additive and that inputs are divisible across products, we can rewrite the production function as:
$\widehat{q}_{fjt} = f(\tilde{\boldsymbol{x}}_{ft}, \hat{\boldsymbol{\beta}}, \widehat{w}_{fjt}, \rho_{fjt}) + \omega_{ft}$, and recover $\left\{\{\rho_{fjt}\}_{j=1}^{J}, \omega_{ft}\right\}$ using:

$$\widehat{q}_{fjt} - f_1(\tilde{\boldsymbol{x}}_{ft}, \hat{\boldsymbol{\beta}}, \widehat{w}_{fjt}) = f_2(\tilde{\boldsymbol{x}}_{ft}, \widehat{w}_{fjt}, \rho_{fjt}) + \omega_{ft} \tag{23}$$

$$\sum_j \exp\left(\rho_{fjt}\right) = 1, \tag{24}$$

where f_1 and f_2 depend on the functional form of the production function and the input prices $\widehat{w}_{fjt}$ for each product j are computed based on the input price function (12). In other words, to recover the input allocations ρ_{fjt}, we separate the production function into a component f_1 that captures all terms that do not depend on ρ_{fjt} and a component f_2 that collects all terms that involve ρ_{fjt}. Because the input allocation shares have to sum up to 1 across all products in a multi-product firm, this yields a system of $J_{ft} + 1$ equations (where J_{ft} is the number of products produced by firm f at time t) in $J_{ft} + 1$ unknowns (the J_{ft} input allocations ρ_{fjt} and ω_{ft}) for each firm-year pair.

Let $\widehat{\omega}_{fjt} = \widehat{q}_{fjt} - f_1(\tilde{\boldsymbol{x}}_{ft}, \hat{\boldsymbol{\beta}}, w_{ft})$. Applying our translog functional form to (23), we obtain:

$$\widehat{\omega}_{fjt} = \omega_{ft} + \hat{a}_{fjt}\rho_{fjt} + \hat{b}_{fjt}\rho_{fjt}^2 + \hat{c}_{fjt}\rho_{fjt}^3 \tag{25}$$

The terms $\hat{a}_{ft}$, $\hat{b}_{ft}$, and $\hat{c}_{ft}$ are functions of the estimated parameter vector $\hat{\boldsymbol{\beta}}$ and the estimated input price correction $\widehat{w}_{fjt}$.[38]

For each year, we obtain the firm's productivity and input allocations, the $J+1$ unknowns $(\omega_{ft}, \rho_{f1t}, \ldots, \rho_{fJ_t})$, by solving a system of $J+1$ equations:

$$\begin{aligned}
\widehat{\omega}_{f1t} &= \omega_{ft} + \hat{a}_{f1t}\rho_{f1t} + \hat{b}_{f1t}\rho_{f1t}^2 + \hat{c}_{f1t}\rho_{f1t}^3 && (26)\\
&\ldots && (27)\\
\widehat{\omega}_{fJ_t t} &= \omega_{ft} + \hat{a}_{fJ_t t}\rho_{fJ_t t} + \hat{b}_{fJ_t t}\rho_{fJ_t t}^2 + \hat{c}_{fJ_t t}\rho_{fJ_t t}^3 && (28)\\
\sum_{j=1}^{J} \exp(\rho_{fjt}) &= 1, \qquad \exp(\rho_{fjt}) \leq 1 \qquad \forall fjt && (29)
\end{aligned}$$

This system imposes the economic restriction that each input share can never exceed one and they must together sum up to one across products in a firm. We numerically solve this system for each firm in each year.

3.4 Markups and Marginal Costs

We can now apply our framework to compute markups and marginal costs using the estimates of the production function coefficients ($\boldsymbol{\beta}$) and the input allocations ($\boldsymbol{\rho}$). We calculate the markup for each product-firm pair f, j in each time period t using:

$$\hat{\mu}_{fjt} = \hat{\theta}^M_{fjt} \frac{P_{fjt}Q_{fjt}}{\exp(\hat{\rho}_{fjt})\widetilde{X}^M_{ft}}, \tag{30}$$

where $\hat{\theta}^M_{fjt} = \theta(\hat{\boldsymbol{\beta}}, \tilde{\boldsymbol{x}}_{ft}, \widehat{w}_{fj}, \hat{\rho}_{fjt})$ and $\widetilde{X}^M_{ft}$ denotes the firm's expenditure on materials.

The product-specific output elasticity for materials $\hat{\theta}^M_{fjt}$ is a function of the production function coefficients and the materials allocated to product j. Hence, it can be easily computed once the allocation of inputs across products has been recovered.[39]Marginal costs mc_{fjt} are then recovered

[38]For the translog, these terms are

$$\begin{aligned}
\hat{a}_{ft} &= \hat{\beta}_k + \hat{\beta}_l + 3\hat{w}_{fjt}^2\hat{\beta}_{lmk} + \tilde{l}_{ft}\left(\hat{\beta}_{lk} + 2\hat{\beta}_{ll} + \hat{\beta}_{lm} + \tilde{k}_{ft}\hat{\beta}_{lmk} + \tilde{m}_{ft}\hat{\beta}_{lmk} - 2\hat{w}_{fjt}\hat{\beta}_{lmk}\right) + \hat{\beta}_m + \tilde{k}_{ft}\left(2\hat{\beta}_{kk} + \hat{\beta}_{lk} + \tilde{m}_{ft}\hat{\beta}_{lmk}\right)\\
&\quad + \tilde{k}_{ft}\left(-2\hat{w}_{fjt}\hat{\beta}_{lmk} + \hat{\beta}_{mk}\right) + \hat{w}_{fjt}\left(-2\hat{\beta}_{kk} - 2\hat{\beta}_{lk} - 2\hat{\beta}_{ll} - 2\hat{\beta}_{lm} - 2\hat{\beta}_{mk} - 2\hat{\beta}_{mm}\right) + \tilde{m}_{ft}\left(\hat{\beta}_{lm} - 2\hat{w}_{fjt}\hat{\beta}_{lmk} + \hat{\beta}_{mk} + 2\hat{\beta}_{mm}\right)\\
\hat{b}_{ft} &= \hat{\beta}_{kk} + \hat{\beta}_{lk} + \hat{\beta}_{ll} + \hat{\beta}_{lm} + \hat{\beta}_{lmk}\tilde{k}_{ft} + \hat{\beta}_{lmk}\tilde{l}_{ft} + \hat{\beta}_{lmk}\tilde{m}_{ft} - 3\hat{w}_{fjt}\hat{\beta}_{lmk} + \hat{\beta}_{mk} + \hat{\beta}_{mm}\\
\hat{c}_{ft} &= \hat{\beta}_{lmk}
\end{aligned}$$

[39]The expression for the materials output elasticity for product j at time t is: $\widehat{\theta}^M_{fjt} = \hat{\beta}_m + 2\hat{\beta}_{mm}m_{fjt} + \hat{\beta}_{lm}l_{fjt} + \hat{\beta}_{mk}k_{fjt} + \hat{\beta}_{lmk}l_{fjt}k_{fjt}$. As before, to obtain the physical inputs, we rely on our estimates of the input prices $\widehat{w}_{fjt}$

by dividing price by the relevant markup according to equation (6).

Note that both markups and marginal costs are estimates since they depend on the estimated production function coefficients and the input cost allocation parameters, which are estimates themselves since they depend on the production function coefficients. Hence, the only source of uncertainty in our markup (and marginal cost) estimates comes from using estimated coefficients (the production function coefficients $\hat{\boldsymbol{\beta}}$ and the input price correction coefficients $\hat{\boldsymbol{\delta}}$). We account for the measurement error in these variables when we estimate the reduced form regressions in Section 4 by bootstrapping over the entire procedure. We execute the following steps in sequence: 1) estimate the production function, 2) recover the input allocations, 3) calculate markups (marginal costs), and 4) project markups and costs on trade policy variables. We then repeat this procedure 500 times, using bootstrapped (with replacement) samples that keep the sample size equal to the original sample size. This allows us to compute the bootstrapped standard error on the trade policy coefficients in Section 4.

3.5 Discussion

In addition to the conceptual assumptions discussed in Section 3.1, the actual implementation of the approach requires a set of assumptions to accommodate limitations of the data. Some of these limitations are specific to our data set (for example, we do not have information on physical labor units and wages, but only the wage bill) and may be of little general relevance. But other limitations are present in every firm-level data set and will need to be addressed by any study using such data. To our knowledge, no dataset reports the allocation of input expenditures across products in multi-product firms or contains the complete information on the firm-specific input prices (including firm-specific price of capital). The additional assumptions we impose are needed in order to deal with these features of the data. Apart from measurement issues, the assumptions we employ also address challenges that arise from product differentiation.

In this section we discuss these additional assumptions and our identification strategy. We start by discussing the way we deal with the unobserved input allocations in multi-product firms.

3.5.1 The Use of Single-Product Firms: Economies of Scope and Relationship to Cost Function Estimation

This subsection expands on the discussion of economies of scope in our setting and relates it to discussion of economies of scope in the cost function literature. Our approach does not rule out economies (or diseconomies) of scope, which may be important for multi-product firms. Panzar (1989) defines economies of scope in terms of cost. Baumol et al. (1983) speak of economies of scope in production if the cost function is sub-additive: $c_{ft}\left([q_1, q_2], \mathbf{w}_{ft}, \omega^2_{ft}\right) \leq c_{ft}([q_1, 0], \mathbf{w}_{ft}, \omega^1_{ft}) + c_{ft}([0, q_2], \mathbf{w}_{ft}, \omega^1_{ft})$ where $c_{ft}(.)$ is a firm's cost function, ω_{ft} is (log) factor-neutral productivity, and $\mathbf{w}_{ft}$ denotes a vector of (log) input prices. The superscripts in the productivity denote the

and the input allocation shares $\hat{\rho}_{fjt}$.

number of products produced by a firm. Our framework allows for factor-neutral productivity to depend on the number of products produced by a firm.

The assumption we impose is that the function $c(.)$ is the same across single- and multi-product firms producing the same product. However, costs between the two types of firms can still differ if there are factor-neutral productivity differences between multi- and single-product firms. To see this, consider the thought experiment of splitting a firm that produces two products to two sub-firms, each of which produces only one product. Economies of scope will exist if $c_{ft}\left(q_1, q_2, \mathbf{w}_{ft}, \omega^2_{ft}\right) < c_{ft}(q_1, \mathbf{w}_{ft}, \omega^1_{ft}) + c_{ft}(q_2, \mathbf{w}_{ft}, \omega^1_{ft})$. Note that this condition is conceptually distinct from the equation implied by Assumption 4, which states that it is possible to allocate all input expenditures of a multi-product firm to individual products, i.e., $c_{ft}\left(q_1, q_2, \mathbf{w}_{ft}, \omega^2_{ft}\right) = c_{ft}(q_1, \mathbf{w}_{ft}, \omega^2_{ft}) + c_{ft}(q_2, \mathbf{w}_{ft}, \omega^2_{ft})$. The indexing of productivity by the number of products is important here. When we allocate expenditures of a multi-product firm to individual products, we hold the firm's productivity constant. In contrast, in the counterfactual of splitting a firm into two subdivisions, we allow for the productivity of each subdivision to be different than the productivity of the original multi-product firm. The dependence of productivity on the number of products a firm produces could arise for several reasons. For example, it is possible that there is learning associated with the production of multiple products or additional managerial experience that makes the firm more efficient; and vice versa, it is possible that the production of multiple lines overwhelms managers resulting in a decline in total factor-productivity.

A further possibility (not borne out in our notation) is that factor prices $\mathbf{w}$ differ across the two types of firms because of pre-negotiated contracts. Such differences are consistent with our assumptions regarding input prices as long as the contracts do not specify bulk discounts that would make current input prices a function of current input quantities. For example, it is possible in our framework for a firm such as Walmart to have lower input prices because it has negotiated good deals with its suppliers in the past; but we do not allow the price Walmart faces on each delivery of supplies to be a function of the size of the delivery. We do not have any data on pre-negotiated prices that would allow us to investigate this possibility, so we do not go down this road empirically. Finally, economies of scope can arise in the short run because of the amortization of fixed costs F across multiple products for multi-product firms.We emphasize that we allow for economies of scope rather than assume it. For example, our results could find no productivity differences between single- and multi-product firms, or find that multi-product firms are less productive implying diseconomies of scope. Likewise, finding economies of scope in the range of our data does not imply existence of economies of scope over any range of products produced by a firm; it is possible that economies of scope switch to diseconomies once a firm reaches a certain number of products. This paper does not attempt to provide a theory of multi-product firms. We simply point out that our approach does not *a priori* rule out economies or diseconomies of scope in the range of our data.

The discussion above raises the natural question of why we do not exploit the duality between production and cost function and estimate a multi-product cost function. The main reason for focusing on the production function is that we do not have information on firm costs (as we do not

observe the firm-specific user cost of capital) or wages. Furthermore, a multi-product cost function estimation would require additional identification assumptions in order to deal with the endogeneity of multiple product outputs on the right-hand side. Finally, even if one could come up with such identification assumptions, the product portfolios in our particular context are not stable. While Indian firms very rarely drop products, they often add products during this period (see Goldberg et al. (2010b)). These frequent additions require explicitly modeling a firm's decision to add a particular product (in contrast, our approach requires us to model only the change from single- to multi-product status). Given these challenges, the approach to estimate production functions from single-product firms while accounting for the potential selection bias is an appealing alternative.

3.5.2 Control Function for Input Prices and Timing Assumptions

This subsection explains how the control function for input prices, the law of motion for productivity and the timing assumptions allow us to identify the coefficients. Recall that the identification strategy involves two control functions for the two unobservables: input prices and productivity:

$$w_{ft} = w_t(p_{ft}, \boldsymbol{ms}_{ft}, \boldsymbol{D}_f, \boldsymbol{G}_f, EXP_{ft}) \tag{31}$$

$$\omega_{ft} = g(\omega_{ft-1}, \tau^{\text{output}}_{it-1}, \tau^{\text{input}}_{it-1}, EXP_{ft-1}, SP_{ft}) + \xi_{ft}. \tag{32}$$

While ω_{ft} enters the production function (10) linearly, the input prices enter non-linearly as part of the term $B(.)$. By substituting the input price control function into the expression for w, we get equation (13).

First, note that we make use of the input price control function in the first stage of the estimation, when we purge the data from the noise ϵ. At this stage, we use materials as a proxy for productivity. Given that materials demand depends on input prices, it is important to control for the input prices using the control function specified above. However, the first stage has no implications for the identification of the production function coefficients; its sole purpose is to net out ϵ.

Next, consider the identification of the production function coefficients $\boldsymbol{\beta}$ and the coefficients associated with the input price correction term $\boldsymbol{\delta}$. These are identified off our timing assumptions. To review these assumptions, we assume that materials are a freely adjustable input and hence they will be correlated with contemporaneous productivity. Similarly, output prices will be correlated with current productivity. In contrast, capital and labor are dynamic inputs. Therefore, they will be uncorrelated with the productivity innovation ξ_{ft}. We rely on these assumptions to form moment conditions.[40]

[40] These timing assumptions are standard in the production function estimation literature. For example, both Olley and Pakes (1996) and Levinsohn and Petrin (2003) assume that capital is a dynamic input and use this assumption to identify the capital coefficient. Our treatment of capital is identical to its treatment in those papers. Our treatment of labor differs as we treat labor as a dynamic input, while the aforementioned papers assume that labor is static. This difference is due to our effort to use assumptions that match the institutional setting in India, a country characterized by significant labor market rigidities. However, the assumption that labor is a dynamic input has no significant implications for our identification strategy; we can easily modify the assumptions to treat labor as a static input and adjust the moment conditions accordingly.

There are two remaining identification issues that need to be discussed. First, as we noted earlier, the term $B(.)$ will in general include input expenditures $\tilde{\boldsymbol{x}}_{ft}$. This raises the question of whether the production coefficients $\boldsymbol{\beta}$ are identified. They are identified because the input expenditures $\tilde{\boldsymbol{x}}_{ft}$ enter the input price term $B(.)$ *only through interaction* with the input prices. It is because of the complexity of the translog that $\tilde{\boldsymbol{x}}_{ft}$ appear in $B(.)$ through interactions with input prices. In a Cobb-Douglas specification, the input expenditures do not appear in $B(.)$. In fact, under a constant returns to scale Cobb-Douglas production function the input correction term $B(.)$ simplifies to $w(.)$.[41]

The second question is how the coefficients on variables that enter both the law of motion for productivity and the input price control function are identified. One example of such a variable is the export dummy. The law of motion for productivity includes a dummy for exporting in $t-1$, while it is also included in the input price control. The answer is that these coefficients are again identified off timing assumptions. We assume that productivity responds with a lag to changes in a firm's environment, since it plausibly takes time for a firm to take the actions required to increase its efficiency (e.g., hiring better managers, adopting better management practices, changing organizational structure, importing new intermediate inputs, etc.). Accordingly, variables that may influence a firm's productivity, such as tariffs or exporting, enter with a lag in the law of motion of productivity. In contrast, output and input prices respond immediately to changes in the economic environment. Accordingly, the variables included in the input price control function enter with their current values. As noted earlier, it is precisely because these variables enter with their current values that we face an identification problem; the current values will be correlated with ξ_{ft} since by assumption they respond to contemporaneous shocks. It is this potential correlation that leads us to form moment conditions based on the lags, and not the current values, of the corresponding variables (the vector $\mathbf{Y}_{ft}$ contains *lagged* output prices, *lagged* market shares, etc.).

As noted in section 3.2.2, we assume that there is a single input price control function across all inputs, $w_t(.)$. This assumption allows us to identify the coefficients of the input price control function once the parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ have been estimated. The coefficients of the $w_t(.)$ function are required to compute firm- and product-specific input prices that are then used to obtain input allocations ρ_{fjt} in multi-product firms in section 3.3. Without the assumption of a common control function for the prices of all inputs, we would still be able to estimate the production function coefficients consistently, but the parameter vector $\boldsymbol{\delta}$ would in this case be a function of all parameters of the input-specific input price control functions. Because our data does not report firm-specific input prices, it would be impossible to identify the parameters of each input price-control function in our case (see the particular example of a Cobb-Douglas production function in Appendix B). However, some data sets report input prices for a subset of (though never for all) firms' inputs. With this additional information, it would be possible to specify and estimate input-specific input price control functions.

[41] See Appendix B for details of the special case of Cobb-Douglas.

4 Empirical Results

4.1 Output Elasticities, Marginal Costs and Markups

In this subsection, we present the output elasticities recovered from the production function estimation procedure. We describe how failing to correct for input price variation or account for the selection bias affects the parameters. Finally, we present and discuss our markup and marginal cost estimates.

The output elasticities are reported in Table 3.[42] The nice feature of the translog is that unlike in a Cobb-Douglas production function, output elasticities can vary across firms (and across products within firms). We report both the average and standard deviation of the elasticities across sectors, and the final column reports the returns to scale. We note that a few sectors appear to have low returns to scale, but these are driven by outliers; Table 4 reports median output elasticities which are less influenced by outliers. Since the returns to scale vary across firms, it is possible for many firms in a sector to have increasing returns to scale, while the estimate of the industry-average returns to scale is close to one. At the firm level, 68 percent of the sample exhibits increasing returns to scale.

The left panel of Table 5 repeats the production function estimation without implementing the correction for the unobserved input price variation discussed in subsection 3.2.2. The uncorrected procedure yields nonsensical estimates of the production function. For example, the output elasticities and returns to scale are sometimes negative, very low or very high. These results are to be expected given that we estimate a quantity-based production function using deflated input expenditures, i.e., we relate physical output to input expenditures. It is clear that failing to account for input price variation yields distorted estimates. To understand the source of the distortion, consider the following concrete example from our data: in 1995, Ashnoor Textile Mills and Delight Handicrafts Palace sold 71,910 and 67,000,000 carpets, respectively. Ashnoor, however, had about three times higher input expenditures and three times higher revenues. It is easiest to understand the implications of this example for the estimates using a Cobb-Douglas specification. A quantity production function estimation that ignores input price variation would result in very large and negative output elasticities (more input expenditures result in lower quantity for Ashnoor). In the more general translog specification, it is impossible to sign this bias because there are three inputs which interact in complicated ways with each other and input prices, but it is clear that one needs to correct for input price variation across firms. By introducing the input price control, we are effectively comparing output quantities to input quantities, and the resulting output elasticities then look reasonable.

The importance of the input price correction is not apparent in the earlier literature, which traditionally estimates a Cobb-Douglas specification of the form: $q + p = \tilde{x}\beta + \tilde{\omega}$. This specification relates deflated sales to deflated expenditures and implies that $\tilde{\omega} = \omega + p - w(.)$. That is, the

[42] The output elasticities for capital and labor are defined analogous to the materials elasticity reported in Footnote 39.

unobserved productivity measure includes both (unobserved) output price p and (unobserved) input prices w. If one does not control for either output or input price variation (the typical practice in this literature until recently), there is no apparent problem as the two price biases tend to work in opposite directions. To obtain some intuition for the combined impact of these biases on the estimation, suppose that higher input prices were completely passed through to higher output prices, so that $p = w(.)$. In this case, $\tilde{\omega} = \omega$, and a regression of revenues $(q + p)$ on input expenditures $\tilde{x}$ would deliver unbiased estimates of the coefficients β. De Loecker and Goldberg (2014) discuss the conditions under which this happens, which turn out to be highly restrictive.[43] In the general case, the output and input biases will not completely offset each other, but they will still partially neutralize each other as higher input prices will generally be partially passed through to higher output prices. This will lead to output elasticities that appear plausible without immediately calling for a correction. In fact, when we estimate a firm-level revenue-based production function using the standard approach, we obtain production function coefficients that look similar to the previous literature (see Appendix E). Of course, this does not mean that the two biases exactly cancel each other, so the final estimates will generally still be biased. Moreover, estimation of the production function using the revenue-based approach implies that one can only conduct the analysis at the firm-level. Such firm-level analysis would not allow one to obtain marginal costs and markups at the product level and exploit product-specific variation in tariffs in order to identify the effects of the trade reforms.

The right panel of Table 5 presents the median output elasticities from an estimation of the production function that does not include the sample selection correction described in Section 3.2.3. The coefficients change slightly when the selection correction is not implemented. The stability of the coefficient estimates with and without selection correction for the unbalanced panel suggests that the use of the unbalanced panel of single-product firms (which includes firms that are always single-product and firms that ultimately transition to a multi-product status) likely alleviates most of the concerns about the selection bias. This is consistent with the findings in Olley and Pakes (1996).

The markups are reported in Table 6. The mean and median markups are 2.70 and 1.34, respectively, but there is considerable variation across sectors and across products and firms within sectors. Some firms report markups below one for individual products, but multi-product firms maximize profits across products, so they may lose money on some products while being profitable on others. To get a better sense of the plausibility of our estimates, we aggregate the product-level markups to the firm level using the share of sales as weights. The firm-level markups are below one for only about 8 percent of the sample and the median firm-level markup is 1.60. In fact, we find a strong positive (and statistically significant) relationship between firm markups and reported accounting profits, measured as operating profits divided by total sales (results available upon request). Importantly, for our main results below, we rely on *changes* in markups over time by exploiting variation within firm-product pairs rather than variation in levels across firms.

[43] See Sections 2.2.2 and 2.2.4 in De Loecker and Goldberg (2014) for a discussion of this issue.

The methodology provides measures of markups and marginal costs without *a priori* assumptions on the returns to scale. The estimates show that many firms are characterized by increasing returns to scale, so we expect to observe an inverse relationship between a product's marginal cost and quantity produced. Accordingly, another way to assess the plausibility of the measures is to plot marginal costs against production quantities in Figure 1 (we de-mean each variable by product-year fixed effects in order to facilitate comparisons across firms). The figure shows indeed that marginal costs vary inversely with production quantities. The left panel of the figure shows that quantities and markups are positively related indicating that firms producing more output also enjoy higher markups (due to their lower marginal costs).

We also examine how markups and marginal costs vary across products within a firm. Our analysis here is guided by the recent literature on multi-product firms. Our correlations are remarkably consistent with the predictions of this literature, especially with those of Eckel and Neary (2010) and the multi-product firm extension of Melitz and Ottaviano (2008) developed by Mayer et al. (2014). A key assumption in these models is that multi-product firms each have a "core competency". The core product has the lowest (within a firm) marginal cost. For the other products, marginal costs rise with a product's distance from the core competency. Mayer et al. (2014) assume a linear demand system which implies that firms have non-constant markups across products. Furthermore, firms have their highest markups on their "core" products with markups declining as they move away from their main product. Figure 2 provides evidence supporting these implications. They plot the de-meaned markups and marginal costs against the sales share of the product within each firm (markups and marginal costs are de-meaned by product-year and firm-year fixed effects in order to make these variables comparable across products within firms). Marginal costs rise as a firm moves away from its core competency while the markups fall. In other words, the firm's most profitable product (excluding any product-specific fixed costs) is its core product. Despite not imposing any assumptions on the market structure and demand system in our estimation, these correlations are remarkably consistent with the predictions from the multi-product firm literature.

4.2 Pass-Through

Foreshadowing the results in the next subsection, we also find evidence of imperfect pass-through of costs on prices because of variable markups. This subsection explains how we estimate pass-through.

Consider the identity that decomposes the (log) price of a firm f producing product j into its two subcomponents: (log) marginal cost, $\ln mc_{fjt}$, and (log) markup, $\ln \mu_{fjt}$:

$$\ln P_{fjt} = \ln mc_{fjt} + \ln \mu_{fjt} \tag{33}$$

This identity can also be written as:

$$\ln P_{fjt} = \ln \mu_{fj} + \ln mc_{fjt} + (\ln \mu_{fjt} - \ln \mu_{fj}) \tag{34}$$

where $\ln \mu_{fj}$ is the (time-invariant) average (log) markup for this particular firm-product pair and

$(\ln \mu_{fjt} - \ln \mu_{fj})$ is the deviation of the markup from its average. If markups are constant, then the last term becomes zero. This is the case of complete pass-through: a proportional change in marginal cost is passed entirely to prices. If markups are variable, then marginal costs are correlated with the term in parenthesis and pass-through is incomplete. For example, if the price elasticity of demand is increasing in price, then an increase in marginal cost (which will tend to raise the price) will lead to an increase in the price elasticity of demand and a decrease in the markup. In this case, the marginal cost is negatively correlated with the (variable) markup and the pass-through of a marginal cost change onto price is below one. This correlation between marginal costs and markups is *not* an econometric issue since the equation above is an identity. Rather, it is a correlation dictated by economic theory: any model that implies variable markups will also imply a correlation between marginal cost and markup and result in incomplete pass-through.

To understand the implications of variable markups and incomplete pass-through in our setting, first consider the hypothetical case where marginal cost can be measured exactly. Suppose we run the following pass-through regression:

$$\ln P_{fjt} = a_{fj} + \zeta \ln mc_{fjt} + \varepsilon_{fjt} \tag{35}$$

where a_{fj} is a firm-product fixed effect. In this setup, the error term ε_{fjt} has a structural interpretation. It reflects the deviation of the actual markup in period t from the average (i.e., it corresponds to $(\ln \mu_{fjt} - \ln \mu_{fj})$) .

If markups are constant, then we would expect to find that $\zeta = 1$ and $\varepsilon_{fjt} = 0$ (i.e., an exact fit). The firm-product fixed effect a_{fj} would accurately measure the constant markup and the coefficient ζ would measure the pass-through of marginal cost to price which would be complete ($\zeta = 1$). The deviation of the actual markup from the average, ε_{fjt}, would be zero if markups were constant. Of course, in reality we would never get an exact fit of the regression line. But as long as ε_{fjt} captures random variation in price (due for example to recording errors) that is orthogonal to the marginal cost, we would estimate complete pass-through.

If markups are variable, then the error term ε_{fjt} will be correlated with the marginal cost $\ln mc_{fjt}$.[44] We again emphasize that this correlation is dictated by theory and not by econometrics. If the price elasticity facing the firm is increasing in price, then a marginal cost increase will lead to a price increase, which will raise the price elasticity and lower the markup. Hence, ε_{fjt} and $\ln mc_{fjt}$ will be negatively correlated and the pass-through coefficient ζ will be below one. This is the case of incomplete pass-through.

When observing marginal cost, the coefficient ζ reflects markup variability and pass-through. There would be no need to instrument for marginal costs. In fact, instrumenting marginal costs is conceptually incorrect because the correlation between marginal costs and the structural error of the regression (i.e., the markup) is precisely what the coefficient ζ is supposed to capture. However, in our application (and almost every other empirical study), we only observe an estimate of marginal

[44] Variable markups can be generated in many different ways through various combinations of market structure, firm behavior and demand function. See Goldberg and Hellerstein (2013) for a discussion.

cost, $\ln \widehat{mc}_{fjt} = \ln mc_{fjt} + \sigma_{fjt}$. The pass-through regression becomes

$$\ln P_{fjt} = a_{fj} + \zeta \ln \widehat{mc}_{fjt} + (\varepsilon_{fjt} - \zeta\sigma_{fjt}) = a_{fj} + \zeta \ln \widehat{mc}_{fjt} + u_{fjt} \tag{36}$$

Measurement error results in a downward bias in the pass-through coefficient ζ leading us to conclude, potentially erroneously, that pass-through is incomplete. We therefore require instruments to address measurement error in marginal costs. It is important to note that in this setting, instruments must be uncorrelated with the measurement error, σ_{fjt}. However, we do not require that they are uncorrelated with the part of the error term that reflects the deviation in markup, ε_{fjt}. Indeed, such a condition would be inconsistent with the exercise which is precisely to measure the correlation between marginal cost and markup, that is the correlation between $\widehat{mc}_{fjt}$ and ε_{fjt}.

We instrument for marginal cost in equation (36) with input tariffs and lagged marginal cost. Both variables are certainly correlated with marginal cost. The former should be uncorrelated with the measurement error in our marginal cost estimate, but input tariffs do not vary at the firm level. The advantage of lagged marginal cost is that it varies at the firm-product-year level. Although lagged marginal costs contain measurement error, we have no reason to expect this measurement error to be serially correlated.

Table 7 presents the pass-through results from estimating (36).[45] OLS results are reported in column 1, and the coefficient is 0.337. The second column instruments marginal costs with both lagged marginal cost and input tariffs. The coefficient becomes 0.305, but is not statistically different from the OLS estimate. In case one is concerned about first-order serial correlation in measurement error, the third column uses input tariffs and two-period lagged marginal cost as the instruments, and the IV estimate is now 0.405 and significant at the 10.1 percent level. Thus, the results seem robust to the use of alternative instruments and consistently point to low pass-through. This imperfect pass-through means that shocks to marginal costs, for example shocks from trade liberalization, do not lead to proportional changes in factory-gate prices because of changes in markups. We examine this markup adjustment in detail in the subsequent section.

4.3 Prices, Markups and Trade Liberalization

We now examine how prices, markups and marginal costs adjusted as India liberalized its economy. As discussed in Section 2, we restrict the analysis to 1989-1997 since tariff movements after this period appear correlated with industry characteristics.

We begin by plotting the distribution of raw prices in 1989 and 1997 in Figure 3. Here, we include only firm-product pairs that are present in both years, and we compare the prices over time by regressing them on firm-product pair fixed effects plotting the residuals. As before, we remove outliers in the bottom and top 3rd percentiles. This comparison of the same firm-product pairs over time exploits the same variation as our regression analysis below. The figure shows that the distribution of (real) prices did not change much between 1989 and 1997. This might at first be

[45] As noted in Section 3.4, we report bootstrap standard errors.

a surprising result given nature of India's economic reforms during this period that were designed to reduce entry barriers and increase competition in the manufacturing sector. As a first pass, the figure suggests that prices did not move much despite the reforms.

Of course, the figure includes only firm-product pairs that are present at the beginning and end of the sample, and summarizes aggregate trends, thereby not controlling for sector-specific factors that could influence prices beyond the trade reforms. We use the entire sample and control for macroeconomic trends in the following specification:[46]

$$p_{fjt} = \lambda_{fj} + \lambda_{st} + \lambda_1 \tau_{it}^{\text{output}} + \eta_{fjt}. \tag{37}$$

We exploit variation in prices and output tariffs within a firm-product over time through the firm-product fixed effects (λ_{fj}) and control for macroeconomic fluctuations through sector-year fixed effects λ_{st}. Since the trade policy measure varies at the industry level, we cluster our standard errors at this level.[47] We report the price regression with just year fixed effects in column 1 of Table 8. The coefficient on the output tariff is positive implying that a 10 percentage point decline is associated with a small–1.36 percent–decline in prices.[48] Between 1989 and 1997, output tariffs fall on average by 62 percentage points; this results in a precisely estimated average price decline of 8.4 percent (=62*0.136). This is a small effect of the trade reform on prices and it is consistent with the raw distributions plotted in Figure 3. The basic message remains the same if we control more flexibly for trends with sector-year fixed effects in column 2. The results imply that the average decline in output tariffs led to a 10.4 (=62*.167) percent relative drop in prices.

These results show that although the trade liberalization led to lower factory-gate prices, the decline is more modest than we would have expected given the magnitude of the tariff declines. Since earlier studies (Goldberg et al. (2010a), Topalova and Khandelwal (2011)) have emphasized the importance of declines in input tariffs in shaping firm performance, we separate the effects of output tariffs and input tariffs on prices. Output tariff liberalization reflects primarily an increase in competition, while the input tariff liberalization should provide access to lower cost (and more variety of) inputs. We run the analog of the regression in (37), but separately include input and output tariffs:

[46] One could try to capture the net impact of tariff reforms using the effective rate of protection (ERP) measure proposed by Corden (1966). However, this measure is derived in a setting with perfect competition and infinite export-demand and import-supply elasticities which imply perfect pass-through. As we show below, these assumptions are not satisfied in our setting, so that the concept of the "effective rate of protection" is not well defined in our case. The ERP has two further limitations in our context. The first is that the ERP combines the decline in output and input tariffs which blurs the two thought experiments of reducing the marginal cost and changing the residual demand facing firms. The second limitation is that a specification with ERP on the right-hand side, by construction, restricts the marginal effect of a unit decline in output tariff on the outcome of interest to be the same as the marginal effect of a unit increase in an input tariff. The specifications we employ below are more flexible. We nevertheless report results using the ERP in Appendix Table A2. The results suggest that prices decline with a decline in ERP, but we do not find statistically significant effects on marginal costs and markups. As noted above, it is not clear how to interpret these results given that the ERP is conceptually wrong in the our context.

[47] Recall from Section 2 that tariffs vary at a 4-digit level, while sector is defined as a 2-digit industry.

[48] Our result is consistent with Topalova (2010) who finds that a 10 percentage point decline in output tariffs results in a 0.96 percent decline in wholesale prices in India during this period.

$$p_{fjt} = \lambda_{fj} + \lambda_{st} + \lambda_1 \tau_{it}^{\text{output}} + \lambda_2 \tau_{it}^{\text{input}} + \eta_{fjt}. \tag{38}$$

The results are shown in column 1 of Table 9.[49] There are two interesting findings that are important for understanding how trade affects prices in this liberalization episode. First, there is a positive and statistically significant coefficient on output tariffs. This result is consistent with the common intuition that increases in competitive pressures through lower output tariffs will lead to price declines. The effect is traditionally attributed to reductions in markups and/or reductions in X-inefficiencies within the firm. The point estimates imply that a 10 percentage point decline in output tariffs results in a 1.56 percent decline in prices. On the other hand, the coefficient on input tariffs is noisy. Holding input tariffs fixed and reducing output tariffs, we would observe a precisely estimated decline in prices. Overall, average output tariffs and input tariffs fall by 62 and 24 percentage points, respectively, and using the point estimates in column 1, this implies that prices fall on average by 18.1 percent (a decline that is statistically significant).

We use the estimates of markups and costs to examine the mechanisms behind these moderate changes in factory-gate prices. We begin by plotting the distribution of markups and costs in Figure 4. Like Figure 3, this figure considers only firm-product pairs that appear in both 1989 and 1997. The figure indicates that between 1989 and 1997, the marginal cost distribution shifted left indicating an efficiency gain. However, this marginal cost decline is offset by a corresponding rightward shift in the markup distribution. Since (log) marginal costs and (log) markups exactly sum to (log) prices, the net effect results in little changes to prices. Hence, the raw data point towards imperfect pass-through of cost declines to prices. As before, these patterns are only suggestive and presented only for illustrative purposes, given that the figures do not condition on the policy and other changes that took place over this period.

We re-run specification (38) using marginal costs and markups as the dependent variables to formally analyze these relationships. Since prices decompose exactly to the sum of marginal costs and markups, the coefficients in columns 2 and 3 sum to their respective coefficients in column 1 in Table 9. We first focus on the marginal cost regressions reported in column 2. The coefficient on output tariffs is statistically insignificant, suggesting that marginal costs are insensitive to output tariff liberalization. However, the coefficient on input tariffs is both positive and large in magnitude. This is strong evidence that improved access to cheaper and more variety of imported inputs results in large cost declines. The final row of Table 9 reports the average effect on marginal costs using the average declines in input and output tariffs. On average, marginal costs fell 30.7 percent.[50]

[49] The regressions exclude outliers in the top and bottom 3rd percent of the markup distribution. We trim to ensure that the results are not driven by outliers. Nevertheless, the results are robust (e.g., magnitudes change slightly but statistical significance is unaffected) to alternative trims (e..g, the top and bottom 1st) and to not trimming at all (results are available upon request).

[50] This decline is sizable, but consistent with earlier work documenting the effects of input tariffs on input prices and input varieties, with the latter further lowering the exact price index for intermediate inputs in India. Specifically, calculations from Goldberg et al. (2010a) suggest that prices of imported intermediaries fell by 21 percent as a result of the tariff reductions, while new varieties of intermediate inputs increased by 8.9 percent. These estimates cannot be converted to estimates of marginal cost declines without further structure, but they suggest large effects of tariff reductions on firms' costs.

This magnitude of the marginal costs decline is sizable and would translate to larger prices declines if markups were constant. However, Figure 4 suggests that markups rose during this period, and in column 3 of Table 9, we directly examine how input and output tariffs affected markups. The coefficient on input tariffs is large and negative implying that input tariff liberalization resulted in higher markups. The results indicate that firms offset the beneficial cost reductions from improved access to imported inputs by raising markups. The overall effect, taking into account the average declines in input and output tariffs between 1989 and 1997, is that markups, on average, increased by 12.6 percent. This increase offsets almost half of the average decline in marginal costs, and as a result, the overall effect of the trade reform on prices is moderated.[51]

Although tempting, it is misleading to draw conclusions about the pro-competitive effects of the trade reform from the markup regressions in column 3 of Table 9. The reason is that one needs to control for the impacts of the output tariff liberalization on marginal costs in order to isolate the pro-competitive effects. For example, if output tariffs affect costs through changes in X-inefficiencies, firms may adjust markups in response to these cost changes. The simultaneous effects that tariffs have on both costs and markups make it difficult to identify pro-competitive effects of the reform based on the specification in column 3.

To isolate the pro-competitive effects, we need to control for simultaneous shocks to marginal costs. We do this by re-running the markup regression while controlling flexibly for marginal costs. Conditioning on marginal costs, the output tariff coefficient isolates the direct pro-competitive effect of the trade liberalization on markups. We report the results in Table 10.[52] The coefficient on output tariffs in column 1 is positive and significant; this provides direct evidence that output tariff liberalization exerted pro-competitive pressure on markups. The way to interpret the results in column 1 is to consider the markups of two products in different industries. Conditional on any (potentially differential) impact of the trade reforms on their respective costs, the product in the industry that experiences a 10 percentage point larger decline in output tariffs will have a 1.43 percent relative decline in markups.[53] Column 2 instruments marginal costs to account for measurement error (see discussion in Section 4.2) with input tariffs and a second-order polynomial in lagged marginal costs, and the coefficient increases slightly and remains statistically significant. In sum, our analysis demonstrates that although India's trade reform led to large cost reductions, firms responded by raising markups. Once we control for these cost effects, output tariff reductions do have pro-competitive effects by putting downward pressure on markups.

The pro-competitive effects might differ across products. For example, output tariffs may exert more pressure on products with high markups prior to the reform. We explore this heterogeneity by creating a time-invariant indicator for firm-product pairs in the top decile of their industry's

[51] These results are robust to controlling India's delicensing policy reform; see Appendix Table A1.

[52] To control for marginal costs as flexibly as possible, we use a second-order polynomial for marginal costs and suppress these coefficients in Table 10. We find very similar results if we simply include marginal costs as the only control (results are available upon request).

[53] In unreported results, we include input tariffs in the regression. As discussed earlier, input tariffs should affect markups *only through* the imperfect transmission of their impact on costs through improved access to imported inputs. Once we control for marginal costs, input tariffs should have *no* effect on markups and that is what we find.

markup distribution in the first year that a product-pair is observed in the data. We interact output tariffs with this indicator to allow for differential effects of output tariffs on markups for these high markup products. The results are reported in column 3 of Table 10. The table shows a very strong effect of output tariffs on these high markup products: a 10 percentage point decline in output tariffs leads to a 1.29 percent fall in markups for products initially below the 90th percentile in the markup distribution. For high markup products, the same policy reform results in an additional 3.14 percent decline in markups. In short, once we control for the incomplete pass-through of costs, output tariffs reduce markups and these reductions are substantially more pronounced on products with initially high markups. If we instrument marginal costs, the coefficient on output tariffs increases even further, while the coefficient on the interaction remains positive, but is not statistically significant.

4.4 Interpretation of Results: Variable Markups and Incomplete Pass-through

Our results call for a nuanced evaluation of the effects of the Indian trade liberalization on markups. While we do find evidence that the tariff reductions have pro-competitive effects, especially at the right tail of the markup distribution, our results suggest that the most significant effect of the reforms is to reduce costs to producers. Due to variable markups, cost reductions are not passed through completely to consumers.

This last finding raises the question of why prices do not fully respond to cost reductions. Our results here relate to a voluminous literature on price rigidities and incomplete pass-through in macroeconomics and international macroeconomics. While this literature has focused primarily on exchange rate pass-through, its findings are equally relevant to tariff reductions given that exchange rate and tariff changes have similar effects on firm profits. Structural approaches within this literature explain incomplete pass-through through a combination of demand side and market structure assumptions. As discussed in Section 4.2, there is a large class of potential models (i.e., combinations of demand side and market structure assumptions) that can generate this phenomenon. Incomplete pass-through requires the demand elasticity perceived by the firm to be rising in price, so any model that delivers a demand elasticity increasing in price will also deliver incomplete pass-through. For example, this pattern can be generated in a setting with a linear consumer demand and monopolistic competition as in Melitz and Ottaviano (2008). Alternatively, one could assume CES preferences and Cournot (e.g., Atkeson and Burstein (2008)), or nested logit and Bertrand (e.g., Goldberg (1995) or Goldberg and Verboven (2005)); or random coefficients and Bertrand (e.g., Goldberg and Hellerstein (2013) or Nakamura and Zerom (2010)). Which assumptions are appropriate depends on the industry under investigation. Against this background, the advantage of our approach is precisely the fact that it establishes the existence of incomplete pass-through and explores its implications for trade policy without committing to a particular structure. Such structure may be defensible in the context of Industrial Organization case studies which rely on a careful study of the industry under consideration and its institutional setting to inform their assumptions. But it is less defensible in the context of an analysis of the entire Indian manufacturing

sector that includes many heterogeneous industries, each likely characterized by different demand and market conditions. Our study demonstrates that variable markups generate incomplete cost pass-through in many different sectors, but it cannot answer the question of which fundamentals in each case generate variable markups. To answer this last question, one would need to impose more structure along the lines of the aforementioned studies, yet doing so would undermine the fundamental rationale and advantage of our approach.

Our results suggest that the trade reforms benefited producers relatively more than consumers, at least in the short run. However, this does not necessarily imply that the reform lowered consumer welfare. There are two channels through which consumers may have benefited from the trade reforms, despite the fact that prices did not decrease significantly. First, it is possible that the quality of existing products improved. The finding that prices did not decline in full proportion to the decline in trade barriers is consistent with this possibility. Note however that quality upgrading is costly. In the absence of changes in input prices and productivity due to the trade liberalization, we would expect quality upgrading to be associated with an increase in marginal costs, while our study documents a decrease in marginal costs. However, it is possible that in the absence of quality upgrading, marginal costs would have fallen even further. Our results in Table 9 capture the composite effect of all these factors (lower input prices, productivity increases and potential quality changes) on marginal costs. Moreover, the estimates are net of trends, captured by sector-year fixed effects, so we cannot rule out absolute increases in quality. Similarly, the increase in markups is consistent with, but cannot be attributed exclusively to quality upgrading. Without variable markups, a marginal cost change caused by quality changes would have been reflected in proportional changes to prices. However, Table 7 shows that the pass-through of marginal cost changes on prices is incomplete; this is direct evidence that markups changed conditional on marginal cost changes. A model with only quality-upgrading (and no incomplete pass-through) would not generate such a finding. In general, our results are consistent with quality upgrading in response to the trade reform, but cannot be explained by quality upgrading alone.

The second channel through which trade liberalization may have benefited consumers is through long-term dynamic gains. Though such gains are difficult to pin down empirically, they are potentially important. There is an active literature studying the relationship between competition, firm profitability and innovation (e.g., see Aghion et al. (2005)). In Goldberg et al. (2010a), we show that firms introduced many new products–accounting for about a quarter of output growth–during this period. If the cost reductions (and associated markup increases) induced by the trade reform spurred this product growth, the long-run benefits to consumers are potentially substantially larger. We also observe a positive correlation between changes in firm markups and product introductions (results available upon request).[54] This suggests that firms used the input tariff reductions and associated profit increases to finance the development of new products, implying potential long-term gains to consumers. A complete analysis of this mechanism and the impact on welfare lies beyond

[54]These findings are consistent with Peters (2012) who develops a model with imperfect competition that generates heterogeneous markups which determine innovation incentives.

the scope of this current paper.

5 Conclusion

This paper examines the adjustment of prices, markups and marginal costs in response to trade liberalization. We take advantage of detailed price and quantity information to estimate markups from quantity-based production functions. Our approach does not require any assumptions on the market structure or demand curves that firms face. This feature of our approach is important in our context since we want to analyze how markups adjust to trade reforms without imposing *ex ante* restrictions on their behavior. An added advantage of our approach is that since we observe firm-level prices in the data, we can directly compute firms' marginal costs once we have estimates of the markups.

Estimating quantity-based production functions for a broad range of differentiated products introduces new methodological issues that we must confront. We propose an identification strategy based on estimating production functions on single-product firms. The advantage of this approach is that we do not need to take a stand on how inputs are allocated across products within multi-product firms. We also demonstrate how to correct for a bias that arises when researchers do not observe input price variation across firms, an issue that becomes particularly important when estimating quantity-based production functions.

The large variation in markups suggests that trade models that assume constant markups may be missing an important channel when quantifying the gains from trade. Furthermore, our results highlight the importance of analyzing the effects of both output and input tariff liberalization. We observe large declines in marginal costs, particularly due to input tariff liberalization. However, prices do not fall by as much. This imperfect pass-through occurs because firms offset the cost declines by raising markups. Conditional on marginal costs, we find pro-competitive effects of output tariffs on markups. Our analysis is based on data representative of larger firms, so our results are representative of these larger firms. Our results suggest that trade liberalization can have large, yet nuanced effects, on marginal costs and markups. Understanding the welfare consequences of these results using models with variable markups is an important topic for future research.

Our results have broader implications for thinking about the trade and productivity across firms in developing countries. The methodology produces quantity-based productivity measures that can be compared with revenue-based productivity measures. Hsieh and Klenow (2009) discuss how these measures can inform us about distortions and the magnitude of misallocation within an economy. Importantly, our methodology can deliver quantity-based productivity measures purged of substantial variation in markups across firms, which potentially improves upon our understanding of the role of misallocation in generating productivity dispersion. We leave the analysis of the role of misallocation on the distribution of these performance measures for future research.

References

Ackerberg, D. A., K. Caves, and G. Frazer (2006). Structural identification of production functions. *Mimeo, UCLA*.

Aghion, P., N. Bloom, R. Blundell, R. Griffith, and P. Howitt (2005). Competition and Innovation: An Inverted-U Relationship. *The Quarterly Journal of Economics 120*(2), 701–728.

Amiti, M. and J. Konings (2007). Trade liberalization, intermediate inputs, and productivity: Evidence from Indonesia. *American Economic Review 97*(5), 1611–1638.

Arkolakis, C., A. Costinot, D. Donaldson, and A. Rodríguez-Clare (2012). The elusive pro-competitive effects of trade. *mimeo, Yale University*.

Arkolakis, C., A. Costinot, and A. Rodríguez-Clare (2012). New trade models, same old gains? *American Economic Review 102*(1), 94–130.

Arkolakis, C. and M.-A. Muendler (2010). The extensive margin of exporting products: A firm-level analysis. *NBER Working Paper 16641*.

Atkeson, A. and A. Burstein (2008). Pricing-to-Market, Trade Costs, and International Relative Prices. *American Economic Review 98*(5), 1998–2031.

Atkin, D. (2013). Trade, Tastes, and Nutrition in India. *American Economic Review 103*(5), 1629–63.

Atkin, D. and D. Donaldson (2014). Who is getting globalized? the size and implications of intranational trade costs. *Mimeo, MIT*.

Balakrishnan, P., K. Pushpangadan, and M. S. Babu (2000). Trade liberalisation and productivity growth in manufacturing: Evidence from firm-level panel data. *Economic and Political Weekly 35*(41), 3679–3682.

Baumol, W. J., J. C. Panzar, and R. D. Willig (1983). Contestable markets: An uprising in the theory of industry structure: Reply. *American Economic Review 73*(3), 491–96.

Bernard, A. B., J. Eaton, J. B. Jensen, and S. Kortum (2003). Plants and productivity in international trade. *American Economic Review 93*(4), 1268–1290.

Bernard, A. B., S. J. Redding, and P. K. Schott (2010). Multiple-product firms and product switching. *American Economic Review 100*(1), 70–97.

Bernard, A. B., S. J. Redding, and P. K. Schott (2011). Multi-product firms and trade liberalization. *Quarterly Journal of Economics 126*(3), 1271–1318.

Berry, S. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics 25*(2), 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*(4), 841–90.

Besley, T. and R. Burgess (2004). Can labor regulation hinder economic performance? evidence from India. *The Quarterly Journal of Economics 119*(1), 91–134.

Bloom, N. and J. Van Reenen (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics 122*(4), 1351–1408.

Bloom, N. and J. Van Reenen (2010). Why do management practices differ across firms and countries? *Journal of Economic Perspectives 24*(1), 203–24.

Corden, W. M. (1966). The structure of a tariff system and the effective protective rate. *Journal of Political Economy 74*(3), 221–237.

De Loecker, J. (2011). Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica 79*(5), 1407–1451.

De Loecker, J. (2013). Detecting learning by exporting. *American Economic Journal: Microeconomics 5*(3), 1–21.

De Loecker, J. and P. K. Goldberg (2014). Firm performance in a global market. *Annual Review of Economics 6*(1).

De Loecker, J. and F. Warzynski (2012). Markups and firm-level export status. *American Economic Review 102*(6), 2437–2471.

Debroy, B. and A. Santhanam (1993). Matching trade codes with industrial codes. *Foreign Trade Bulletin 24*(1), 5–27.

Eaton, J. and S. Kortum (2002). Technology, geography, and trade. *Econometrica 70*(5), 1741–1779.

Eckel, C. and J. P. Neary (2010). Multi-product firms and flexible manufacturing in the global economy. *Review of Economic Studies 77*(1), 188–217.

Edmonds, C., V. Midrigan, and Y. Xu (2011). Competition, markups and the gains from international trade. *Mimeo, Duke University*.

Feenstra, R. C. and D. Weinstein (2010). Globalization, markups, and the U.S. price level. *NBER Working Paper 15749*.

Foster, L., J. Haltiwanger, and C. Syverson (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review 98*(1), 394–425.

Freund, C. and M. D. Pierola (2011). Export superstars. *Mimeo, The World Bank*.

Goldberg, P. K. (1995). Product differentiation and oligopoly in international markets: The case of the u.s. automobile industry. *Econometrica 63*(4), 891–951.

Goldberg, P. K. and R. Hellerstein (2013). A Structural Approach to Identifying the Sources of Local Currency Price Stability. *Review of Economic Studies 80*(1), 175–210.

Goldberg, P. K., A. K. Khandelwal, N. Pavcnik, and P. Topalova (2009). Trade liberalization and new imported inputs. *American Economic Review 99*(2), 494–500.

Goldberg, P. K., A. K. Khandelwal, N. Pavcnik, and P. Topalova (2010a). Imported intermediate inputs and domestic product growth: Evidence from India. *The Quarterly Journal of Economics 125*(4), 1727–1767.

Goldberg, P. K., A. K. Khandelwal, N. Pavcnik, and P. Topalova (2010b). Multiproduct firms and product turnover in the developing world: Evidence from India. *The Review of Economics and Statistics 92*(4), 1042–1049.

Goldberg, P. K. and F. Verboven (2005). Market integration and convergence to the Law of One Price: evidence from the European car market. *Journal of International Economics 65*(1), 49–73.

Goyal, S. (1996). Political economy of India's economic reforms. *Institute for Studies in Industrial Development Working Paper*.

Hall, R. E. (1986). Market structure and macroeconomic fluctuations. *Brookings Papers on Economic Activity 17*(2), 285–338.

Hall, R. E. (1988). The Relation between Price and Marginal Cost in U.S. Industry. *Journal of Political Economy 96*(5), 921–47.

Halpern, L., M. Koren, and A. Szeidl (2011). Imports and productivity. *CEPR Discussion Paper 5139*.

Harrison, A. E. (1994). Productivity, imperfect competition and trade reform: Theory and evidence. *Journal of International Economics 36*(1-2), 53–73.

Hasan, R., D. Mitra, and K. Ramaswamy (2007). Trade reforms, labor regulations, and labor-demand elasticities: Empirical evidence from India. *The Review of Economics and Statistics 89*(3), 466–481.

Hellerstein, R. (2008, September). Who bears the cost of a change in the exchange rate? Pass-through accounting for the case of beer. *Journal of International Economics 76*(1), 14–32.

Hsieh, C.-T. and P. J. Klenow (2009). Misallocation and manufacturing TFP in China and India. *The Quarterly Journal of Economics 124*(4), 1403–1448.

Katayama, H., S. Lu, and J. R. Tybout (2009). Firm-level productivity studies: Illusions and a solution. *International Journal of Industrial Organization 27*(3), 403–413.

Khandelwal, A. K. (2010). The long and short (of) quality ladders. *Review of Economic Studies 77*(4), 1450–1476.

Klette, T. J. and Z. Griliches (1996). The inconsistency of common scale estimators when output prices are unobserved and endogenous. *Journal of Applied Econometrics 11*(4), 343–61.

Kremer, M. (1993). The o-ring theory of economic development. *The Quarterly Journal of Economics 108*(3), 551–75.

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *American Economic Review 70*(5), 950–59.

Kugler, M. and E. Verhoogen (2011). Prices, plant size, and product quality. *The Review of Economic Studies 79*, 307–339.

Levinsohn, J. (1993). Testing the imports-as-market-discipline hypothesis. *Journal of International Economics 35*(1-2), 1–22.

Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies 70*(2), 317–341.

Mayer, T., M. J. Melitz, and G. I. P. Ottaviano (2014). Market size, competition, and the product mix of exporters. *American Economic Review 104*(2), 495–536.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica 71*(6), 1695–1725.

Melitz, M. J. and G. I. P. Ottaviano (2008). Market size, trade, and productivity. *Review of Economic Studies 75*(1), 295–316.

Nakamura, E. and D. Zerom (2010). Accounting for Incomplete Pass-Through. *Review of Economic Studies 77*(3), 1192–1230.

Olley, G. S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–1297.

Panagariya, A. (2008). *India: The Emerging Giant*. OUP Catalogue. Oxford University Press.

Panzar, J. C. (1989, 00). Technological determinants of firm and industry structure. In R. Schmalensee and R. Willig (Eds.), *Handbook of Industrial Organization*, Volume 1 of *Handbook of Industrial Organization*, Chapter 1, pp. 3–59. Elsevier.

Pavcnik, N. (2002). Trade liberalization, exit, and productivity improvement: Evidence from Chilean plants. *Review of Economic Studies 69*(1), 245–76.

Peters, M. (2012). Heterogeneous mark-ups and endogenous misallocation. *Mimeo, Massachussets Institute of Technology*.

Sivadasan, J. (2009). Barriers to competition and productivity: Evidence from India. *The B.E. Journal of Economic Analysis & Policy 9*(1), 42.

Topalova, P. (2010). Factor immobility and regional impacts of trade liberalization: Evidence on poverty from India. *American Economic Journal: Applied Economics 2*(4), 1–41.

Topalova, P. and A. K. Khandelwal (2011). Trade liberalization and firm productivity: The case of India. *The Review of Economics and Statistics 93*(3), 995–1009.

Verhoogen, E. A. (2008). Trade, quality upgrading, and wage inequality in the Mexican manufacturing sector. *The Quarterly Journal of Economics 123*(2), 489–530.

Wooldridge, J. M. (2009, September). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics Letters 104*(3), 112–114.

Tables and Figures

Table 1: Summary Statistics

	Share of Sample		Single-Product	
	Output	All Firms	Firms	Products
Sector	(1)	(2)	(3)	(4)
15 Food products and beverages	9%	302	135	135
17 Textiles, Apparel	10%	303	161	78
21 Paper and paper products	3%	77	56	32
24 Chemicals	26%	434	194	483
25 Rubber and Plastic	5%	139	85	83
26 Non-metallic mineral products	7%	110	74	60
27 Basic metals	16%	212	115	101
28 Fabricated metal products	2%	74	48	45
29 Machinery and equipment	7%	160	80	186
31 Electrical machinery, communications	5%	89	52	102
34 Motor vehicles, trailers	9%	71	47	95
Total	100%	1,970	1,047	1,400

Notes: Table reports summary statistics for the average year in the sample. The first column reports the share of output by sector in the average year. Columns 2 and 3 report the number of firms and number of single-product firms manufacturing products in the average year. Column 4 reports the number of products by sector.

Table 2: Example of Sector, Industry and Product Classifications

Examples of Industries, Sectors and Products		
NIC Code		Description
27		Basic Metal Industries (Sector *s*)
2710		Manufacture of Basic Iron & Steel (Industry *i*)
	130101010000	Pig iron
	130101020000	Sponge iron
	130101030000	Ferro alloys
	130106040800	Welded steel tubular poles
	130106040900	Steel tubular structural poles
	130106050000	Tube & pipe fittings
	130106100000	Wires & ropes of iron & steel
	130106100300	Stranded wire
2731		Casting of iron and steel (Industry *i*)
	130106030000	Castings & forgings
	130106030100	Castings
	130106030101	Steel castings
	130106030102	Cast iron castings
	130106030103	Maleable iron castings
	130106030104	S.G. iron castings
	130106030199	Castings, nec

Products (*j*)

Notes: This table is replicated from Goldberg et al. (2010b). For NIC 2710, there are a total of 111 products, but only a subset are listed in the table. For NIC 2731, all products are listed in the table.

Table 3: Average Output Elasticities, by Sector

Sector	Observations in Production Function Estimation (1)	Labor (2)	Materials (3)	Capital (4)	Returns to Scale (5)
15 Food products and beverages	795	0.13	0.71	0.15	0.99
		[0.17]	[0.22]	[0.14]	[0.28]
17 Textiles, Apparel	1,581	0.11	0.82	0.08	1.01
		[0.02]	[0.04]	[0.08]	[0.06]
21 Paper and paper products	470	0.19	0.78	0.03	1.00
		[0.12]	[0.10]	[0.05]	[0.06]
24 Chemicals	1,554	0.17	0.79	0.08	1.03
		[0.08]	[0.07]	[0.06]	[0.08]
25 Rubber and Plastic	705	0.15	0.69	-0.02	0.82
		[0.39]	[0.29]	[0.35]	[0.89]
26 Non-metallic mineral products	633	0.16	0.67	-0.04	0.79
		[0.26]	[0.12]	[0.40]	[0.36]
27 Basic metals	949	0.14	0.77	0.01	0.91
		[0.09]	[0.11]	[0.06]	[0.18]
28 Fabricated metal products	393	0.18	0.75	0.03	0.96
		[0.04]	[0.08]	[0.17]	[0.17]
29 Machinery and equipment	702	0.20	0.76	0.18	1.13
		[0.08]	[0.05]	[0.05]	[0.14]
31 Electrical machinery & communications	761	0.09	0.78	-0.06	0.81
		[0.11]	[0.11]	[0.22]	[0.28]
34 Motor vehicles, trailers	386	0.25	0.63	0.11	1.00
		[0.26]	[0.20]	[0.20]	[0.25]

Notes: Table reports the output elasticities from the production function. The first column reports the number of observations for each production function estimation. Columns 2-4 report the average estimated output elasticity with respect to each factor of production for the translog production function for all firms. Standard deviations (not standard errors) of the output elasticities are reported in brackets. The 5th column reports the average returns to scale, which is the sum of the preceding three columns.

Table 4: Median Output Elasticities, by Sector

Sector	Labor (1)	Materials (2)	Capital (3)	Returns to Scale (4)
15 Food products and beverages	0.12	0.75	0.20	1.09
17 Textiles, Apparel	0.11	0.82	0.09	1.02
21 Paper and paper products	0.18	0.79	0.03	0.98
24 Chemicals	0.16	0.79	0.06	1.02
25 Rubber and Plastic	0.21	0.75	0.04	1.03
26 Non-metallic mineral products	0.18	0.69	0.04	0.88
27 Basic metals	0.14	0.78	0.02	0.96
28 Fabricated metal products	0.17	0.75	0.02	0.94
29 Machinery and equipment	0.17	0.75	0.16	1.08
31 Electrical machinery & communications	0.10	0.80	0.01	0.91
34 Motor vehicles, trailers	0.23	0.64	0.10	0.97

Notes: Table reports the median output elasticities from the production function. Columns 2-4 report the median estimated output elasticity with respect to each factor of production for the translog production function for all firms. The 5th column reports the median returns to scale.

Table 5: Output Elasticities, Input Price Variation and Sample Selection

	Estimates without Correcting for Input Price Variation				Estimates without Correcting for Sample Selection			
Sector	Labor (1)	Materials (2)	Capital (3)	Returns to Scale (4)	Labor (1)	Materials (2)	Capital (3)	Returns to Scale (4)
15 Food products and beverages	0.03	0.75	0.82	1.78	0.22	0.63	0.14	1.03
17 Textiles, Apparel	-0.07	0.70	-0.07	0.52	0.11	0.83	0.09	1.03
21 Paper and paper products	-0.13	0.23	-0.19	-0.23	0.17	0.77	0.03	0.98
24 Chemicals	0.38	0.69	-0.72	0.26	0.16	0.79	0.04	0.99
25 Rubber and Plastic	-0.10	0.30	-0.15	0.21	0.17	0.75	-0.05	0.94
26 Non-metallic mineral products	0.08	0.64	0.81	1.50	0.12	0.71	0.11	0.93
27 Basic metal	-0.18	1.11	-0.33	0.69	0.12	0.80	0.02	0.94
28 Fabricated metal products	-1.17	-0.28	1.60	0.28	0.15	0.74	0.04	0.95
29 Machinery and equipment	-0.72	1.18	-0.50	-0.10	0.16	0.76	0.15	1.06
31 Electrical machinery, communications	-1.59	0.57	-0.13	-0.47	0.10	0.84	0.02	0.95
34 Motor vehicles, trailers	-0.23	-0.39	1.23	0.44	0.20	0.70	0.04	0.94

Notes: The left table reports the median output elasticities from production function estimations that do not account for input price variation. The right panel reports the median output elasticities from production function estimations that do not account for sample selection (transition from single-product to multi-product firms).

Table 6: Markups, by Sector

	Markups	
Sector	Mean	Median
15 Food products and beverages	1.78	1.15
17 Textiles, Apparel	1.57	1.33
21 Paper and paper products	1.22	1.21
24 Chemicals	2.25	1.36
25 Rubber and Plastic	4.52	1.37
26 Non-metallic mineral products	4.57	2.27
27 Basic metals	2.54	1.20
28 Fabricated metal products	3.70	1.36
29 Machinery and equipment	2.48	1.34
31 Electrical machinery, communications	5.66	1.43
34 Motor vehicles, trailers	4.64	1.39
Average	2.70	1.34

Notes: Table displays the mean and median markup by sector for the sample 1989-2003. The table trims observations with markups that are above and below the 3rd and 97th percentiles within each sector.

Table 7: Pass-Through of Costs to Prices

	Log $Price_{fjt}$		
	(1)	(2)	(3)
Log Marginal $Cost_{fjt}$	0.337 ***	0.305 ***	0.406 †
	0.041	0.084	0.247
Observations	21,246	16,012	12,334
Within R-squared	0.27	0.19	0.09
Firm-Product FEs	yes	yes	yes
Instruments	-	yes	yes
First-Stage F-test	-	98	5

Notes: The dependent variable is (log) price. Column 1 is an OLS regression on log marginal costs. Column 2 instruments marginal costs with input tariffs and lag marginal costs. Column 3 instruments marginal costs with input tariffs and two-period lag marginal costs. The regressions exclude outliers in the top and bottom 3rd percent of the markup distribution. All regressions include firm-product fixed effects. The regressions use data from 1989-1997. The standard errors are bootstrapped and are clustered at the firm level. Significance: † 10.1 percent, * 10 percent, ** 5 percent, *** 1 percent.

Table 8: Prices and Output Tariffs, Annual Regressions

	Log $Prices_{fjt}$	
	(1)	(2)
Output $Tariff_{it}$	0.136 **	0.167 ***
	0.056	0.054
Within R-squared	0.00	0.02
Observations	21,246	21,246
Firm-Product FEs	yes	yes
Year FEs	yes	no
Sector-Year FEs	no	yes
Overall Impact of Trade Liberalization	-8.4 **	-10.4 ***
	3.4	3.3

Notes: The dependent variable is a firm-product's (log) price. Column 1 includes year fixed effects and Column 2 includes sector-year fixed effects. The regressions exclude outliers in the top and bottom 3rd percent of the markup distribution. All regressions include firm-product fixed effects and use data from 1989-1997. Standard errors are clustered at the industry level. The final row uses the average 62% decline in output tariffs from 1989-1997 to compute the mean and standard error of the impact of trade liberalization on prices. That is, for each column the mean impact is equal to the -0.62*100*{coefficient on output tariffs}. Significance: * 10 percent, ** 5 percent, *** 1 percent.

Table 9: Prices, Costs and Markups and Tariffs

	Log Prices$_{fjt}$	Log Marginal Cost$_{fjt}$	Log Markup$_{fjt}$
	(1)	(2)	(3)
Output Tariff$_{it}$	0.156 ***	0.047	0.109
	0.059	0.084	0.076
Input Tariff$_{it}$	0.352	1.160 **	-0.807 ‡
	0.302	0.557	0.510
Within R-squared	0.02	0.01	0.01
Observations	21,246	21,246	21,246
Firm-Product FEs	yes	yes	yes
Sector-Year FEs	yes	yes	yes
Overall Impact of Trade Liberalization	-18.1 **	-30.7 **	12.6
	7.4	13.4	11.9

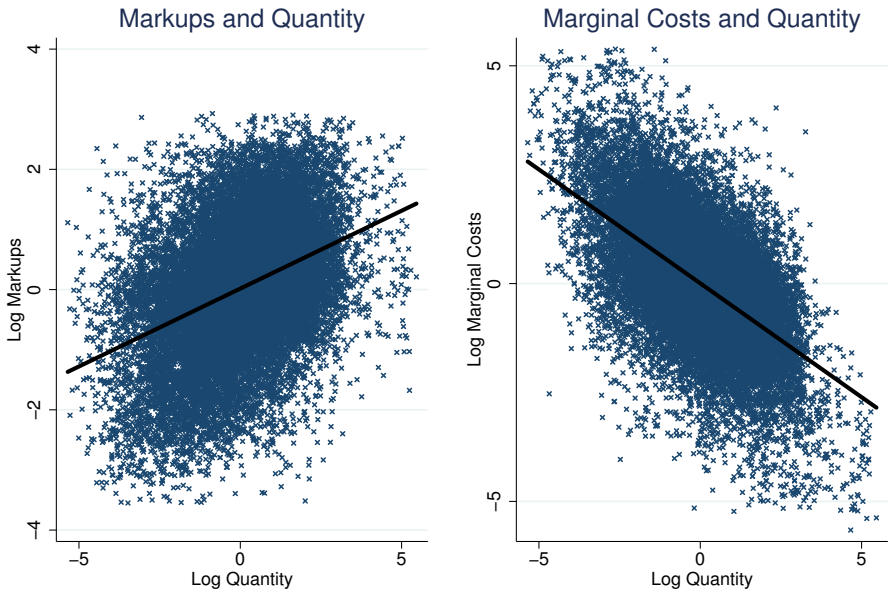
Notes: The dependent variable is noted in the columns. The sum of the coefficients from the markup and marginal costs regression equals their respective coefficient in the price regression. The regressions exclude outliers in the top and bottom 3rd percent of the markup distribution, and include firm-product fixed effects and sector-year fixed effects. The final row uses the average 62% and 24% declines in output and input tariffs from 1989-1997, respectively, to compute the mean and standard error of the impact of trade liberalization on each performance measure. That is, for each column the mean impact is equal to the -0.62*100*{coefficient on output tariff} + -0.24*100*{coefficient on input tariff}. The regressions use data from 1989-1997. The table reports the bootstrapped standard errors that are clustered at the industry level. Significance: ‡ 11.3 percent, * 10 percent, ** 5 percent, *** 1 percent.

Table 10: Pro-Competitive Effects of Output Tariffs

	Log Markup$_{fjt}$			
	(1)	(2)	(3)	(4)
Output Tariff$_{it}$	0.143 ***	0.150 **	0.129 **	0.149 **
	0.050	0.062	0.052	0.062
Output Tariff$_{it}$ x Top$_{fp}$			0.314 **	0.028
			0.134	0.150
Within R-squared	0.59	0.65	0.59	0.65
Observations	21,246	16,012	21,246	16,012
2nd-Order Marginal Cost Polynomial	yes	yes	yes	yes
Firm-Product FEs	yes	yes	yes	yes
Sector-Year FEs	yes	yes	yes	yes
Instruments	no	yes	no	yes
First-stage F-test	-	8.6	-	8.6

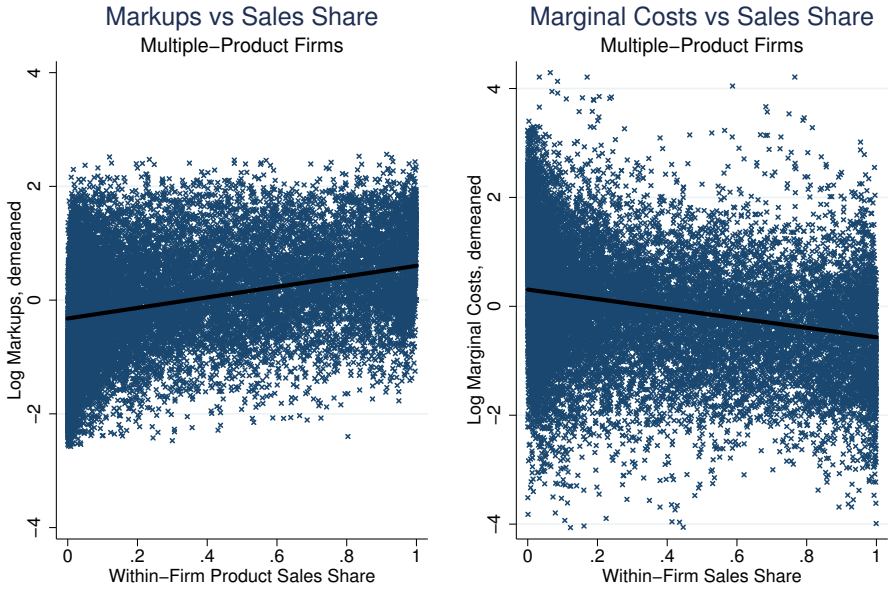
Notes: The dependent variable is (log) markup. All regressions include firm-product fixed effects, sector-year fixed effects and a second-order polynomial of marginal costs (these coefficients are suppressed and available upon request). Columns 2 and 4 instrument the second-order polynomial of marginal costs with second-order polynomial of lag marginal costs and input tariffs. Columns 3 interacts output tariffs and the second-order marginal cost polynomial with an indicator if a firm-product observation was in the top 10 percent of its sector's markup distribution when it first appears in the sample. The regressions exclude outliers in the top and bottom 3rd percent of the markup distribution. The table reports the bootstrapped standard errors that are clustered at the industry level. Significance: * 10 percent, ** 5 percent, *** 1 percent.

Figure 1: Marginal Costs and Quantities



Variables demeaned by product–year FEs.
Markups, cost and quantity outliers are trimmed below and above 3rd and 97th percentiles.

Figure 2: Markups, Costs and Product Sales Share



Markups and marginal costs are demeaned by product–year and firm–year FEs.
Markup and marginal cost outliers are trimmed below and above 3rd and 97th percentiles.

Figure 3: Distribution of Prices in 1989 and 1997

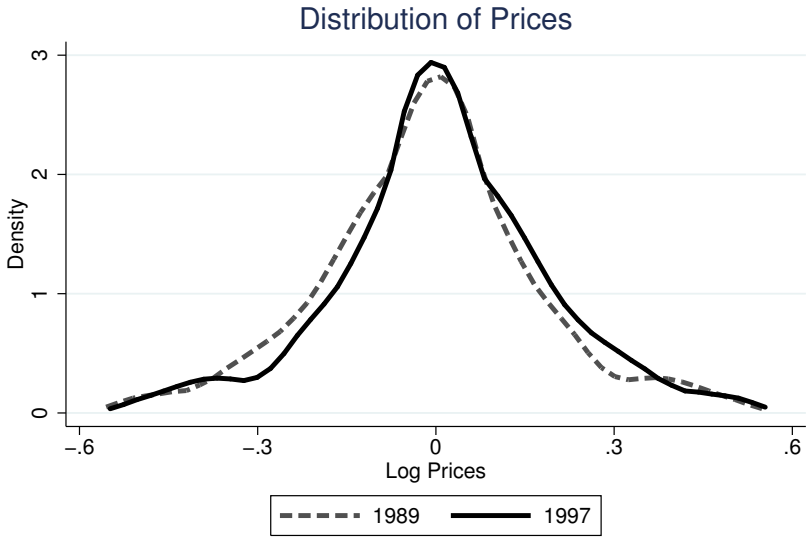
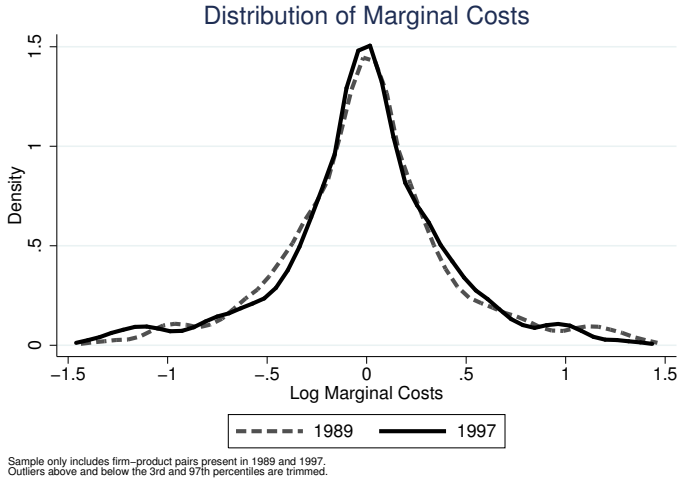
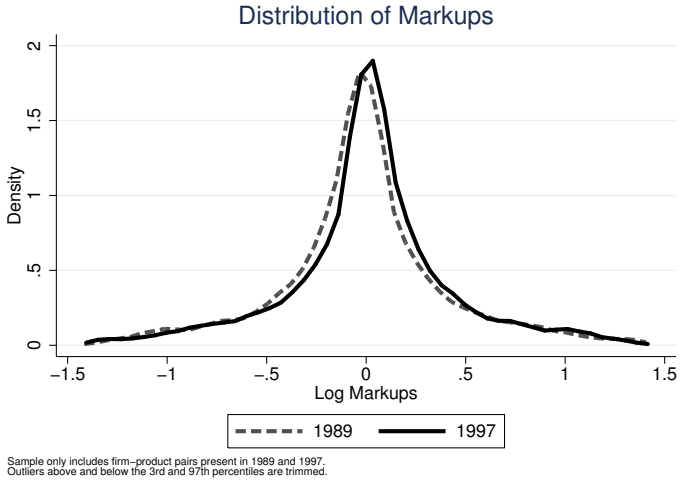


Figure 4: Distribution of Marginal Costs and Marginal Costs in 1989 and 1997





Appendices

A A Formal Model of Input Price Variation

This appendix provides a formal economic model that rationalizes the use of a flexible polynomial in output price, market share and product dummies to control for input prices. The model is a more general version of the models considered in Kremer (1993) and Verhoogen (2008).

We proceed in the following steps. We first show that under the assumptions of the model, the

quality of *every* input is an increasing function of output quality. Next, we show that this implies that the price of *every* input will be an increasing function of output quality. In the final step, we show that output quality can be expressed as a flexible function of output price, market share and a set of product dummies. Having established a monotone relationship between input prices and output quality, this implies that the price of every input can also be expressed as a function of the above variables.

A.1 Production Function for Output Quality

Let v_j indicate quality of product j and ψ_i indicate the quality of input i used to produce product j.[55] The production function for output quality is given by:

$$v_j = \prod_{i=1}^{n} [\psi_i]^{\kappa_i} \omega_j \qquad with \qquad \sum \kappa_i < 1 \tag{A.1}$$

For example, with three inputs, the above production function takes the form:

$$v_j = \psi_K^{\kappa_K} \psi_L^{\kappa_L} \psi_M^{\kappa_M} \omega_j$$

This function belongs to the class of 'O-Ring' production functions discussed in Kremer (1993) and Verhoogen (2008). The particular (multiplicative) functional form is not important; the important feature is that $\frac{\partial v_j}{\partial \psi_i \partial \psi_k} > 0$, $\forall i, k$ and $i \neq k$. This cross-derivative implies complementarity in the quality of inputs. A direct consequence is that higher output quality requires high quality of *all* inputs (e.g., high quality material inputs are used by high-skill workers operating high-end machinery). The production function for quality can vary across industries, but we assume that all firms producing in the same industry face the same quality production function.

In addition to the production function for quality, we assume that higher quality inputs are associated with higher input prices. Let $\overline{W}_i$ denote the sectoral average of the price of input i (e.g., sectoral wage) and $W_i(\psi_i)$ the price of a specific quality ψ of input i. Then,

$$W_i(\psi_i) - \overline{W}_i = z_i \psi_i \quad \text{and} \quad z_i > 0. \tag{A.2}$$

In our framework that postulates perfectly competitive input markets, this assumption is tantamount to assuming that input markets are characterized by vertical differentiation only. So while high quality inputs are expensive, all firms pay the same input prices *conditional* on input quality.

A.2 Demand

The indirect utility V_{nj} that consumer n derives from consuming one unit of product j can be written in general form as:

[55] Here, the subscript j denotes a particular product produced by a firm.

$$V_{nj} = \theta_n v_j - \alpha p_j + \varepsilon_{nj} \tag{A.3}$$

where p_j is output price, θ_n denotes the willingness to pay for quality and ε_{nj} denotes an idiosyncratic preference shock. This specification is general and encompasses the demand models commonly used in the literature. In its most general formulation, the specification above corresponds to the random coefficients model. In models of pure vertical differentiation, the utility will be given by the above expression with $\varepsilon_{nj} = 0$. A simple logit sets $\theta_n = \theta = 1$ (i.e., no observable consumer heterogeneity) and ε_{nj} is assumed to follow the extreme value distribution. In the nested logit, $\theta_n = \theta = 1$ and ε_{nj} follows the generalized extreme value distribution. Following the Industrial Organization literature, it is convenient to define the mean utility δ_j of product j as $\delta_j = v_j - \alpha p_j$. The output quality v_j is typically modeled as a function of observable and unobservable product characteristics. For example, in Berry (1994) 's notation with X_j denoting observable product characteristics,ξ_j denoting unobservable quality, and a specification of mean utility that is linear in characteristics, output quality is given by $v_j = X_j\beta + \xi_j$.

We now show how to control for quality variation across firms using observable characteristics using the specification in (A.3). Berry (1994) shows that the actual market share of a product (ms_j) is a function of product characteristics and output price:

$$ms_j = s_j(\boldsymbol{\delta}, \boldsymbol{\sigma}) = s_j(\mathbf{v}, \mathbf{p}, \boldsymbol{\vartheta}) \tag{A.4}$$

where $\boldsymbol{\sigma}$ denotes a vector of density parameters of consumer characteristics and $\boldsymbol{\vartheta}$ denotes a parameter vector. While the exact functional form is determined by choice of a particular demand structure, the general insight is that market shares are a function of product characteristics (i.e., quality) and prices. Berry (1994) shows that equation (A.4) can be inverted to obtain the mean utilities $\boldsymbol{\delta}$ as a function of the observed market shares and the density parameters to be estimated.[56] With the $\delta's$ in hand, one can obtain quality as a function of output price and the mean utility. This insight is exploited by Khandelwal (2010) who uses a nested logit model to express quality as a function of output price and conditional and unconditional market shares. In a simple logit model, quality is a function of only output prices and unconditional market shares. Here, we use a general formulation that specifies quality as a function of output price, a vector of (conditional and unconditional) market shares and a set of product dummies:

$$v_j = v(p_j, \boldsymbol{ms}_j, \boldsymbol{D}) \tag{A.5}$$

The product dummies are used in lieu of product characteristics (which are not available in our data) and can accommodate more general demand specifications such as the nested logit and even the random coefficients model in cases where it is reasonable to assume that product characteristics do not change from year to year.

[56] In the random coefficients model, the δ's are solved numerically. In simpler models, one can solve for the mean utilities analytically.

A.3 The Firm's Maximization Problem

Without loss of generality, we assume that firms use prices and quality as strategic variables to maximize profits. Conditional on exogenous (to the firm) input prices that are determined in competitive input markets, firms choose input qualities. These choices determine the output quality according to the quality production function in (A.1). Let mc_j denote the marginal cost of producing a product j of quality v_j. The marginal cost can be written as a function of quantity produced q_j, quality v_j, a parameter vector $\boldsymbol{\gamma}$ and productivity ω_j, $mc_j(q_j, v_j, \gamma, \omega_j)$.

The profit function for a firm producing product j is:

$$\pi_j = N \cdot s_j[p - mc_j(q_j, v_j(\boldsymbol{\psi}, \omega_j), \gamma, \omega_j)] \tag{A.6}$$

where N denotes the market size (number of potential consumers). Output quality v_j is now explicitly written as a function of a vector of input qualities $\boldsymbol{\psi}$ and productivity ω_j using the production function for quality in (A.1).

The first order condition with respect to price is

$$p_j = mc_j(q_j, v_j, \gamma, \omega_j) + \frac{s_j}{|\partial s_j/\partial p_j|}. \tag{A.7}$$

The term $s_j/|\partial s_j/\partial p_j|$ represents the markup, and as shown in Berry (1994), p. 254, it equals $\frac{1}{\alpha}[s_j/(\partial s_j/\partial \delta_j)]$.

The first order condition with respect to the quality of each input i, ψ_i, is:

$$(p_j - mc_j) \cdot \frac{\partial s_j}{\partial \psi_i} - s_j \frac{\partial mc_j}{\partial \psi_i} = 0 \tag{A.8}$$

From the first order condition with respect to price, we have

$$(p_j - mc_j) = \frac{s_j}{|\partial s_j/\partial p_j|} = \frac{1}{\alpha}\frac{s_j}{\partial s_j/\partial \delta_j}. \tag{A.9}$$

Substituting this latter expression for the markup into the first order condition for input quality, we obtain:

$$s_j \frac{1}{\alpha}[1/(\partial s_j/\partial \delta_j)]\frac{\partial s_j}{\partial \psi_i} - s_j \frac{\partial mc_j}{\partial \psi_i} = 0 \tag{A.10}$$

or

$$\frac{1}{\alpha}[1/(\partial s_j/\partial \delta_j)]\left[\frac{\partial s_j}{\partial v_j}\frac{\partial v_j}{\partial \psi_i}\right] = \frac{\partial mc_j}{\partial \psi_i} \tag{A.11}$$

From $\delta_j = v_j - \alpha p_j$ follows that $\frac{\partial s_j}{\partial v_j} = \frac{\partial s_j}{\partial \delta_j}$, and the above first order condition simplifies to:

$$\frac{1}{\alpha}\frac{\partial v_j}{\partial \psi_i} = \frac{\partial mc_j}{\partial \psi_i} \tag{A.12}$$

Using the production function for quality to obtain the derivative $\frac{\partial v_j}{\partial \psi_i}$ and substituting into (A.12), we obtain

$$\psi_i = \frac{1}{\alpha}\kappa_i v_j \left[1/\frac{\partial mc_j}{\partial \psi_i}\right] \quad \forall i \tag{A.13}$$

This expression is similar to the one derived in Verhoogen (2008), but with two differences. First, as we have shown above, the above expression can be derived from a very general demand system and market structure. Second, we did not assume a Leontief production technology. The last feature of the model complicates the analysis slightly. With a Leontief production technology, the derivative $\frac{\partial mc_j}{\partial \psi_i}$ is constant, and it will be positive given the assumption that higher quality inputs demand higher prices. However, with more general production technologies, this derivative will itself depend on quality. We therefore need to show explicitly that ψ_i is an increasing function of v_j. The latter can be established using the second order conditions associated with profit maximization:

$$\begin{aligned} \frac{1}{\alpha}\kappa_i \frac{\partial v_j}{\partial \psi_i}\frac{1}{\psi_i} - \frac{1}{\alpha}\kappa_i v_j \frac{1}{(\psi_i)^2} - \frac{\partial^2 mc_j}{\partial \psi_i^2} &< 0 \\ \frac{1}{\alpha}\kappa_i^2 \frac{v_j}{(\psi_i)^2} - \frac{1}{\alpha}\kappa_i \frac{v_j}{(\psi_i)^2} - \frac{\partial^2 mc_j}{\partial \psi_i^2} &< 0 \end{aligned} \tag{A.14}$$

Let us define function $F \equiv \psi_i \left(\frac{\partial mc_j}{\partial \psi_i}\right) - \frac{1}{\alpha}\kappa_i v_j$. From the implicit function theorem, $\frac{\partial \psi_i}{\partial v_j} = -\frac{F_j}{F_i}$ where

$$F_j = -\frac{1}{\alpha}\kappa_i < 0 \tag{A.15}$$

and by virtue of the second order condition,

$$F_i = \frac{\partial mc_j}{\partial \psi_i} + \psi_i \frac{\partial^2 mc_j}{\partial \psi_i^2} - \frac{1}{\alpha}\kappa_i^2 \frac{v_j}{\psi_i} = \frac{1}{\alpha}\kappa_i v_j \frac{1}{\psi_i} + \psi_i \frac{\partial^2 mc_j}{\partial \psi_i^2} - \frac{1}{\alpha}\kappa_i^2 \frac{v_j}{\psi_i} > 0 \tag{A.16}$$

It follows that $\frac{\partial \psi_i}{\partial v_j} = -\frac{F_j}{F_i} > 0$. That is, input quality is an increasing function of output quality for every input.

Given the assumption that higher input quality demands a higher input price, it immediately follows that input prices will also be an increasing function of output quality for all inputs. From equation (A.2):

$$W_i(\psi_i) = \overline{W}_i + z_i \psi_i = \overline{W}_i + z_i \frac{1}{\alpha}\kappa_i v_j \left[1/\frac{\partial mc_j}{\partial \psi_i}\right]$$

In light of the above discussion, each input price facing a particular firm can be expressed as a function of the firm's output quality, $W_i = g_i(v_j)$. Moreover, given that output quality is a function of output price, market share and product dummies, we have: $W_i = w_i(p_j, \boldsymbol{ms}_j, \boldsymbol{D})$. The input price function will be in general input-specific, as the indexation by i indicates. When estimating the

production function, we can allow for input-specific input price functions and the coefficients $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ will be still identified. However, in this general case, we are not able to identify the coefficients of each input price function separately, which are required for computing the firm-specific input prices $\widehat{w}_{fjt}$ needed in the computation of the input allocations ρ_{fjt} in section 3.3. Therefore, we impose the same function $W_i = w(p_j, \boldsymbol{ms}_j, \boldsymbol{D})$ across all inputs in which case the firm-specific input prices reduce to a scalar that we can identify once the parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ have been estimated. We note however that in other applications that do not require the computation of the ρ's , it is possible to consistently estimate the parameters of quantity-based production functions using input-specific input price control functions. Furthermore, even in applications that require the estimation of firm-specific input prices like ours, it would be possible to allow for input-specific input price control functions if one had data on input prices for a subset of inputs. For example, in many data sets there is information on firm-specific wages and sometimes there is even information on firm-specific materials prices. In such cases, one would not need to estimate input price control functions for labor and materials (since the input prices are observed in these cases), so that one could allow of an input price control function specific to capital.

B Estimation Procedure under a Special Case: Cobb-Douglas Production Function

We present our estimation procedure under the predominantly used production function specification in applied work: the Cobb-Douglas (CD) production function. While restrictive on the input-substitution patterns and the output elasticities, it greatly simplifies the estimation routine and the recovery of the input allocation terms (ρ). In addition, it helps to highlight the fundamental identification forces as the input price correction term does not include (interactions of) deflated expenditures.

We follow the structure of the main text (Section 3) and impose the CD functional form:

$$f(\boldsymbol{x}_{fjt}) = \beta_l l_{fjt} + \beta_m m_{fjt} + \beta_k k_{fjt}. \tag{B.1}$$

Following the same steps as in the main text we get the following estimating equation for the single-product firms corresponding to equation (10). We omit the product subscript j given that the firms used in the estimation produce a single product:

$$q_{ft} = \beta_l \tilde{l}_{ft} + \beta_m \tilde{m}_{ft} + \beta_k \tilde{k}_{ft} - \Gamma w_{ft} + \omega_{ft} + \epsilon_{ft}, \tag{B.2}$$

where $\Gamma w(.)$ is a special case of the function $B(.)$ in the main text, $\Gamma = \beta_l + \beta_m + \beta_k$ is the returns to scale parameter, and as before $w_{ft} = \tilde{x}_{ft} - x_{ft}\ \forall x = \{l, m, k\}$.

After running the first stage

$$q_{ft} = \phi_t(\tilde{\boldsymbol{x}}_{ft}, \boldsymbol{z}_{ft}) + \epsilon_{ft}, \tag{B.3}$$

with $\tilde{\boldsymbol{x}}_{ft} = \{\tilde{l}_{ft}, \tilde{m}_{ft}, \tilde{k}_{ft}\}$, we have an estimate of predicted output ($\hat{\phi}_{ft}$). It is then immediate that the input price correction term $B(.)$ enters in equation (20) in a separate and additive fashion:

$$\omega_{ft}(\boldsymbol{\beta}, \boldsymbol{\delta}) = \hat{\phi}_{ft} - \beta_l \tilde{l}_{ft} - \beta_m \tilde{m}_{ft} - \beta_k \tilde{k}_{ft} - \Gamma w(p_{ft}, \boldsymbol{ms}_{ft}, \boldsymbol{D}, \mathbf{G}_{ft}), \tag{B.4}$$

where $-\Gamma w(.)$ is a special case of the function $B(.)$ in the main text. If one assumes a vertical differentiation model of demand, then the input price control function $w(.)$ will take only output price as its argument, and the last term in (B.4) becomes $\Gamma w(p_{ft})$. We form moments on $\xi_{ft}(\boldsymbol{\beta}, \boldsymbol{\delta})$ by exploiting the same law of motion of productivity in equation (18), and the same timing assumptions as in the main text.

In the special case where $w(.)$ is a function of output price only, we can easily demonstrate how the assumption of a common $w(.)$ across inputs helps identify the coefficients of the single input control function. Suppose that $w(p_{ft}) = \gamma p_{ft}$. In this case $\delta = \Gamma\gamma = (\beta_l + \beta_m + \beta_k)\gamma$; therefore, once the parameters of the production function, $\beta_l, \beta_m, \beta_k$, and δ are estimated, the coefficient γ is identified. But suppose we had allowed the input price control function to vary by inputs so that: $w_l(p_{ft}) = \gamma_l p_{ft}$; $w_m(p_{ft}) = \gamma_m p_{ft}$; and $w_k(p_{ft}) = \gamma_k p_{ft}$. Then: $\delta = (\beta_l\gamma_l + \beta_m\gamma_m + \beta_k\gamma_k)$. Given our timing assumptions, we would still be able to consistently estimate the coefficients of the production function and δ, but we would not be able to separately identify the coefficients γ_l,γ_m, and γ_k. Hence in this case, we would not be able to obtain the firm-specific input prices.

To estimate markups and marginal costs we need the input allocation terms ρ_{fjt}. In the case of the CD, their derivation is simplified to solving the system of equations given by:

$$\omega_{ft} + \Gamma\rho_{fjt}\widehat{w}_{fjt} = \hat{\phi}_{fjt} - \beta_l \tilde{l}_{ft} - \beta_m \tilde{m}_{ft} - \beta_k \tilde{k}_{ft} \tag{B.5}$$

where $\widehat{w}_{fjt}$ is the input price term that we compute based on the estimated function $w(.)$ and Γ is defined as above. Taking into account that $\sum_j \exp(\rho_{fjt}) = 1$, this results in a system of $J_{ft} + 1$ equations (one for each product j produced by firm f at time t, plus the summing up constraint for the input allocations) in $J_{ft} + 1$ unknowns (the J_{ft} input allocations for each firm-year pair and firm productivity) and we can solve for ρ_{fjt} and ω_{ft}.

We now have all we need to compute markups and marginal costs. The main difference from the translog is that $\theta^M_{fjt} = \beta_m$, so that all the variation in markups (and marginal costs) comes from the materials expenditure share α_{fjt}.

C Data Appendix

We use the Prowess data, compiled by the Centre for Monitoring the Indian Economy (CMIE), that spans the period from 1989 to 2003. In addition to standard firm-level variables, the data include annual sales and quantity information on firms' product mix. Although Prowess uses an internal product classification that is based on the Harmonized System (HS) and National Industry Classification (NIC) schedules, our version of Prowess did not explicitly link the product names

reported by the firms to this classification. We hired two research assistants, working independently, to map the codes to the product names reported by firms. The research assistants assigned product codes with identical NIC codes in 80% of the cases, representing 91% of output. A third research assistant resolved the differences between the mappings done by the first two research assistants by again manually checking the classifications.

To estimate the production function, we need firm-level labor, capital and materials. Prowess does not have reliable employment information, so we use the total wage bill (which includes bonuses and contributions to employees' provident funds) as our measure for labor. Materials are defined as the consumption of commodities by an enterprise in the process of manufacturing or transformation into product. It includes raw material expenses and consumption of stores and spares. Capital is measured by gross fixed assets, which includes movable and immovable assets. These variables are deflated by two-digit NIC wholesale price indexes.

We match the firm variables to tariff data. The tariff data are reported at the six-digit HS level and were compiled by Topalova (2010). We pass the tariff data through India's input-output matrix for 1993-94 to construct input tariffs. We concord the tariffs to India's NIC schedule developed by Debroy and Santhanam (1993). Formally, input tariffs are defined as $\tau_{it}^{\text{input}} = \sum_k a_{ki} \tau_{kt}^{\text{output}}$, where $\tau_{kt}^{\text{output}}$ is the tariff on industry k at time t, and a_{ki} is the share of industry k in the value of industry i.

D Markups and Monopsony Power

If firms have monopsony power, this would alter the first order conditions in Section 3.1 (equations 3-5). We briefly discuss under which conditions our main results, relating markups to tariff changes, are not affected.

Consider a firm that produces just one product, and suppose production requires just one flexible input V_{ft}^v. The Lagrangian in this case would be:

$$\mathcal{L} = W_{ft}^v V_{ft}^v + \lambda_{ft} \left(Q_{ft} - Q_{ft} \left(V_{ft}^v, \omega_{ft} \right) \right). \tag{D.1}$$

Taking first order conditions and allowing for monopsony power gives:

$$\frac{\partial \mathcal{L}}{\partial V_{ft}^v} = W_{ft}^v + \frac{\partial W_{ft}^v}{\partial V_{ft}^v} V_{ft}^v - \lambda_{ft} \frac{\partial Q(.)}{\partial V_{ft}^v} = 0. \tag{D.2}$$

If a firm has no monopsony power, $\frac{\partial W_{ft}^v}{\partial V_{ft}^v} = 0$. For firms with monopsony power, $\frac{\partial W_{ft}^v}{\partial V_{ft}^v} < 0$: the more the firm buys, the lower the price of the input. We can rearrange the FOC as:

$$W_{ft}^v + \frac{\partial W_{ft}^v}{\partial V_{ft}^v} V_{ft}^v = \lambda_{ft} \frac{\partial Q(.)}{\partial V_{ft}^v} \tag{D.3}$$

The Lagrange multiplier remains: $\lambda_{ft} = P_{ft}/\mu_{ft}$, We get:

$$\mu_{ft}\left(W^v_{ft} + \frac{\partial W^v_{ft}}{\partial V^v_{ft}} V^v_{ft}\right) = P_{ft}\frac{\partial Q(.)}{\partial V^v_{ft}}. \tag{D.4}$$

If we now compare a firm with and without monopsony power, ceteris paribus, the markup for the firm with monopsony power will be larger. This implies that we may be under-estimating the markup by ignoring potential monopsony power.

However, even if our estimates of the markup levels were biased due to the existence of monopsony power, it is still unlikely that our conclusions regarding the effects of tariffs on markups and costs would be affected. To see this, note that the above expression can be simplified to[57]

$$\mu_{ft} = (\theta_{ft}\alpha^{-1}_{ft})/(1 + \upsilon_{ft}). \tag{D.5}$$

where υ is the elasticity of the input price with respect to the quantity of the input purchased $\upsilon_{ft} = \frac{\partial W^v_{ft}}{\partial V^v_{ft}}\frac{V^v_{ft}}{W^v_{ft}}$, and the other variables are as defined in the main text. If there is no monopsony power, then $\upsilon_{ft} = 0$, and the markup expression corresponds to the one we use in the main text of the paper. Taking logs of the more general markup expression implies that in our trade regressions (see Section 4.3) we run $\ln \mu_{ft} + \ln(1 + \upsilon_{ft})$ against output and input tariffs (in multi-product firms, markups and input price elasticities would be indexed by both firm f and product j). The inclusion of firm-product fixed effects implies that we will only bias our results if the input price elasticity changed post-trade reforms. Moreover, we have two empirical pieces of evidence that our results are robust to monopsony power. We might expect that the firms that are most likely to have monopsony power are larger firms or firms that are parts of Indian business groups. However we do not find differential effects of the trade reform across initial firm sizes or if a firm belongs to a business group.[58] This leads us to believe that monopsony power is not a first order concern in our setting.

E Results from the Standard Approach (Online Appendix–Not for Publication)

In this appendix, we compare our results to what would be obtained if one followed a standard approach of working with typical firm-level data that captures inputs and sales at the firm level. We aggregate our data to the firm level and repeat both the production function estimation and the main specifications that relate prices, markups and costs to trade policy. In Appendix Table A3, we report estimates of input elasticities from a Cobb-Douglas production function that uses a standard firm-level deflated revenue-based production function and a standard control function proxy. In Appendix Figure A1, we plot these sectoral elasticities against the elasticities from our

[57] Dividing through by W^v, and dividing and multiplying the right-hand-side by (V^v/Q), and rearranging terms.
[58] Results are available upon request.

methodology reported in Table 4. Although there is a positive correlation for each of the factor elasticities (with the exception of capital), as well as the returns to scale, the estimates produced by the two approaches are not the same.

Qualitatively similar estimates from the two approaches suggest that the input-price bias is partly offset by the output-price bias when using standard firm-level data; that is, firms with higher input prices tend to have higher output prices. Thus, in estimation of production function based on standard firm-level revenue data, input-price price bias occurs simultaneously with another bias– the output price bias–which works in the opposite direction and makes the input-price bias less transparent. However, this does not mean that the two biases necessarily exactly cancel each other. The extent of the offset will depend on the setting (see De Loecker and Goldberg (2014) for an extensive discussion). And while the two biases are working in opposite directions to produce "reasonable" elasticities, we have no way of assessing the exact quantitative net bias. De Loecker and Goldberg (2014) discuss some conditions under which the two biases would exactly offset each other: 1) the industry is characterized by monopolistic competition; 2) firms produce a horizontally differentiated product and face the same CES demand system; 3) production is characterized by constant returns to scale; and 4) input price variation (across firms and time) is input neutral. These conditions are violated in our setting (as evidence by the fact that our elasticities are not identical in the two approaches).

We next use the production function estimates from the "standard approach" to re-examine one of our main results: how do markups change with the trade liberalization. Once we have an estimate of the production function coefficient on materials (θ^M), we can compute markups at the firm level $\mu_{ft} = \frac{\theta^M}{\alpha_{ft}}$ where α_{ft} is the firm's expenditure on materials divided by total sales. While we can compute markups at the firm level, we cannot compute marginal costs because it is not possible to construct a firm-level price without further information on demand. This immediately points out another limitation of not having product-level data: a markup estimate at the firm level cannot be decomposed into prices and costs.

Nevertheless, we can still examine how firm-level markups adjust to the trade reform. We regress the (log of) firm-level markups on output and input tariffs, both defined at the firm level using the firm's initial main industry, and year and firm fixed effects. We cluster standard errors at the industry level. Our results, shown in Appendix Table A4, are qualitatively similar to our main results reported in column 3 of Table 9. Output tariffs appear to have little effect on markups (recall that we cannot isolate pro-competitive effects in this regression since we cannot infer firm-level marginal costs). And although the estimates are somewhat noisy, we do find that a decline in input tariffs leads to an increase in markups. However, the standard errors are large, perhaps because we lose power by working at the firm-level rather than at the firm-product level.

These additional robustness checks suggest two implications. First, in practice, the input and output price biases are likely to offset each other, at least to some extent. This is related to higher input prices being associated with higher output prices. Second, working at the firm level means that it is not possible to decompose changes in prices into costs and markups. Many firm-level

data sets do not have information on prices, but even when information on prices is available, one would need to assume a demand system in order to create a firm-specific price index. Therefore, the standard practice of estimating revenue-based production functions using standard firm-level data is not sufficient for investigating how prices adjust to trade reforms and for explaining this adjustment by examining the response of marginal costs and markups.

Online Tables and Figures (Not for Publication)

Table A1: Controlling for Delicensing

	Log Prices$_{fjt}$	Log Marginal Cost$_{fjt}$	Log Markup$_{fjt}$
	(1)	(2)	(3)
Output Tariff$_{it}$	0.152 ***	0.042	0.110
	0.053	0.096	0.068
Input Tariff$_{it}$	0.344	1.158 *	-0.813 **
	0.506	0.693	0.402
Delicense$_{it}$	-0.012	0.010	-0.022
	0.046	0.093	0.072
Within R-squared	0.02	0.01	0.01
Observations	20,705	20,705	20,705
Firm-Product FEs	yes	yes	yes
Sector-Year FEs	yes	yes	yes

Notes: The dependent variable is noted in the columns. This table controls for whether or not the industry is delicensed at time *t*. The sum of the coefficients from the markup and marginal costs regression equals their respective coefficient in the price regression. The regressions exclude outliers in the top and bottom 3rd percent of the markup distribution, and include firm-product fixed effects and sector-year fixed effects. The regressions are run from 1989-1997 and standard errors are clustered at the industry level. Significance: * 10 percent, ** 5 percent, *** 1 percent.

Table A2: Prices, Markups and Costs and Effective Rate of Protection

	Log Prices$_{fjt}$	Log Marginal Cost$_{fjt}$	Log Markup$_{fjt}$
	(1)	(2)	(3)
Effective Rate of Protection$_{it}$	0.058 ***	0.024	0.034
	0.019	0.038	0.027
Within R-squared	0.02	0.01	0.01
Observations	21,246	21,246	21,246
Firm-Product FEs	yes	yes	yes
Sector-Year FEs	yes	yes	yes

Notes: The dependent variable is noted in the columns. The sum of the coefficients from the markup and marginal costs regression equals their respective coefficient in the price regression. The regressions exclude outliers in the top and bottom 3rd percent of the markup distribution, and include firm-product fixed effects and sector-year fixed effects. The regressions use data from 1989-1997. Standard errors that are clustered at the industry level. Significance: * 10 percent, ** 5 percent, *** 1 percent.

Table A3: Output Elasticities using Firm-Level Revenue Data and Cobb-Douglas Production Function

	Control Function, Cobb-Douglas Coefficients, Firm-Level			
Sector	Labor	Material	Capital	RTS
15 Food products and beverages	0.25	0.71	0.06	1.02
17 Textiles, Apparel	0.08	0.86	0.06	0.99
21 Paper and paper products	0.31	0.69	0.05	1.05
24 Chemicals	0.13	0.69	0.13	0.96
25 Rubber and Plastic	0.19	0.78	0.10	1.07
26 Non-metallic mineral products	0.38	0.18	0.31	0.87
27 Basic Metal	0.15	0.70	0.12	0.97
28 Fabricated metal products	0.17	0.76	0.03	0.96
29 Machinery and equipment	0.18	0.71	0.17	1.05
31 Electrical machinery, communications	0.22	0.69	0.15	1.05
34 Motor vehicles, trailers	0.21	0.68	0.22	1.11

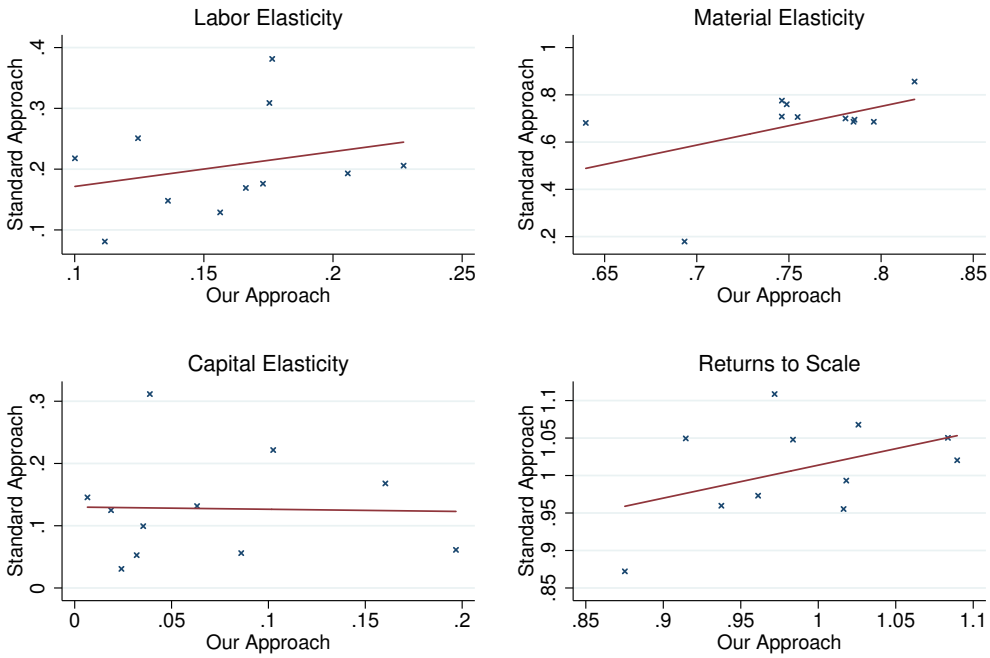
Notes: Table reports coefficients of a three-factor Cobb-Douglas production function: labor, materials, and capital. The estimations are run at the firm level using total revenues. Estimations are performed separately by sector using a control function approach (Levinsohn and Petrin, 2003).

Table A4: Firm-level Markups on Output and Input Tariffs

	Log Markup$_{ft}$
	(1)
Output Tariff$_{ft}$	-0.007
	0.032
Input Tariff$_{ft}$	-0.212
	0.290
R-squared	0.03
Observations	12,827
Firm FEs	yes
Sector-Year FEs	yes

Notes: Table reports the regression of (log) markups on output and input tariffs. Markups are constructed at the firm level using the materials output elasticity estimated from a firm-level deflated revenue production function estimation. Input and output tariffs are matched to the firm's initial main industry. Standard errors clustered at the industry level. Significance: * 10 percent, ** 5 percent, *** 1 percent.

Figure A1: Output Elasticities Comparison



Our Approach takes the elasticities reported in Table 4
Standard Approach uses elasticities from estimating a Cobb−Douglas function on firm−level data