PRICES, MARKUPS AND TRADE REFORM

Jan De Loecker
Pinelopi K. Goldberg
Amit K. Khandelwal
Nina Pavcnik

Prices, Markups and Trade Reform
Jan De Loecker, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik
NBER Working Paper No. 17925
March 2012, Revised September 2012
JEL No. F1,L1

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 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. Long-term gains to consumers may be higher to the extent that higher firm profits lead to new product introductions and growth. Indeed, firms with larger increases in markups had a higher propensity to introduce new products during this period.

Jan De Loecker
Department of Economics
307 Fisher Hall
Princeton University
Princeton, NJ 08544-1021
and NBER
jdeloeck@princeton.edu

Amit K. Khandelwal
Graduate School of Business
Columbia University
Uris Hall 606, 3022 Broadway
New York, NY 10027
and NBER
ak2796@columbia.edu

Pinelopi K. Goldberg
Yale University
Department of Economics
37 Hillhouse Ave.
P.O. Box 208264
New Haven, CT 06520-8264
and NBER
penny.goldberg@yale.edu

Nina Pavcnik
Department of Economics
6106 Rockefeller Center
Dartmouth College
Hanover, NH 03755
and NBER
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) and 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 developing a unified framework to estimate jointly markups and marginal costs from production data, and examine how prices, and their underlying markup and cost components, adjust during India's comprehensive trade liberalization.

Our paper makes three main contributions. The first contribution is towards measurement. In order to infer markups, one requires estimates of production functions. Typically, these estimates have well-known biases if researchers use revenue data rather than quantity data to estimate production functions. 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)). De Loecker (2011) demonstrates that controlling for demand shocks substantially attenuates the productivity increases in response to trade reforms in the European Union textile industry. As a result, approaches to infer markups from production data may be problematic if one uses revenue information. We alleviate this concern by inferring markups from a quantity-based production function using data that contain the prices and quantities of firms' products over time.

Second, we develop a unified framework to estimate markups, marginal costs and productivity of multi-product firms across a broad set of manufacturing industries. Since these performance measures are unobserved, we must impose some structure on the data. However, our approach requires substantially fewer assumptions than is typically required in industrial organization studies. Crucially, our approach does not require assumptions on consumer demand, market structure or the nature of competition. This flexibility enables us to infer the full distribution of markups across firms and products over time in different manufacturing sectors. Moreover, since prices are observed, we can directly recover marginal costs from our markup estimates. Our approach is quite general and since data containing this level of detail are becoming increasingly available, 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.

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., see 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, declines 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 (e.g., Bernard et al. (2003), Melitz and Ottaviano (2008), Mayer et al. (2011), Feenstra and Weinstein (2010), Arkolakis et al. (2012) and Edmonds et al. (2011)). 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. 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 De Loecker and Warzynski (2012), but we extend their methodology to account for multi-product firms and to take advantage of observable price data. The key assumption we need to infer markups is simply 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.[1] We obtain this output elasticity from estimates of production functions across many industries. In contrast to many studies, we utilize physical quantity data rather than revenues to estimate the production functions.[2] This alleviates the concern that the production function estimation is contaminated by prices, yet presents different challenges that we discuss in detail in Section 4. 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. We also confront the potential concern that (unobserved) input prices vary across firms. This concern is usually ignored in earlier work, but we show how to deal with this issue by exploiting variation in observable firm and market characteristics.

This approach calls for an explicit treatment of multi-product firms. We show 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 in the production function estimation stage is that we do not require assumptions on how firms allocate inputs across products, something we do not observe in our data.[3] While this approach may seem strong, it simply assumes that the *physical* relation-

---

[1] Our framework explicitly accounts for production that relies on fixed factors that may be costly to adjust over time. These include machinery and capital goods, and potentially also labor. This is particularly important in a country like India that has relatively strong labor regulations (Besley and Burgess (2004)) as well as capital market distortions. Our approach requires at least one flexible input that is adjustable in the short run; in our case, this variable input is materials.

[2] Foster et al. (2008) also use quantity data in their analysis of production functions, but they focus on a set of homogeneous products.

[3] Suppose a firm manufactures three products using raw materials, labor and capital. To our knowledge, no dataset covering

ship 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 a rickshaw as a multi-product firm. This assumption 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)). Moreover, this assumption of the same physical production structure does not rule out economies of scope, which can operate through lower input prices for multi-product firms (for example, due to their size), or higher (factor-neutral) productivity of multi-product firms. Once we estimate the production functions from the single-product firms, we show how to back out the productivity of multi-product firms and their allocation of inputs across their products. We obtain the markups for each product manufactured by firms by dividing the output elasticity of materials by the materials share of total revenue.[4] Finally, we divide prices by the markups to obtain marginal costs.

Our estimation of the output elasticities of the production function provides reasonable results. Notably, many firms exhibit economies of scale, a finding consistent with Klette and Griliches (1996) and De Loecker (2011) who show that estimation based on revenue data leads to a downward bias in the estimated returns to scale.

The performance measures are correlated in intuitive ways and are consistent with recent heterogeneous models of multi-product firms. Markups (marginal costs) are positively (negatively) correlated with a firm's underlying productivity. More productive firms manufacture more products, but firms have lower markups (higher marginal costs) on products that are farther from their core competency. 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).

We then analyze 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. The combination of an important trade reform that has statistical benefits and the ability to observe prices and their underlying components is a unique contribution of our analysis. 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 16.8 percent despite average tariff declines of 62 percentage points. Marginal costs, however, decline on average by 40.3 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. Output tariffs have only a small and statistically insignificant effect on costs, suggesting that reductions of X-inefficiencies are modest in comparison to reductions due to imported inputs.[5] Since our prices decompose exactly into

---

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 are appropriate for the particular homogenous 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.

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

[5]The relative importance of input and output tariffs is consistent with Amiti and Konings (2007) and Topalova and Khandelwal (2011) who find that firm-level productivity changes in Indonesia and India, respectively, were predominantly driven by input tariff declines.

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 declines was because markups increased: on average, the trade reform raised relative markups by 23.4 percent. The results imply that firms offset the cost declines from input tariff reductions by raising markups, and the net effect is that the reform has an attenuated impact on prices. The increases in markups do not imply that the trade reforms caused firms to collude or engage in less competitive behavior. Rather, the results simply show 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 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. This evidence suggests that the higher profits may have financed the development of new products that contributed to long run gains to consumers. A more detailed investigation of this channel 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 (Bloom and Van Reenen (2007)) 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 increases productivity. India's reforms are an excellent context to study these questions because of the nature of the reform 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 productivity gains. 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, productivity and marginal costs. In Section 4, we explain

the estimation and identification strategy employed in our analysis, paying particular attention to issues that arise specifically in the context of multi-product firms. Section 5 presents the results and Section 6 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. First, unlike the repeated cross section in the ASI, Prowess is a panel that tracks firm performance over time. Second, the data span India's trade liberalization from 1989-2003. 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.

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 the (independently) reported manufacturing sales.[6] We refer the reader to 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. There are a total of 1,501 products in the sample for estimation.[7] 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.

---

[6]We deflate all nominal values for our analysis. Materials are deflated by the wholesale price index of primary articles. Wages are deflated by the overall wholesale price index for manufacturing. Capital stock is fixed assets deflated by sector-specific wholesale price indexes. Unit values are also deflated by sector-specific wholesale price indexes.

[7]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.

6

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.[8]

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 combined 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)).[9] The 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 lib-

---

[8]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).

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

eralization 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 Production Technology, Markups and Costs

This section describes the framework to estimate markups and marginal costs from firm-level production data. Our approach to recovering markups follows De Loecker and Warzynski (2012).[10] The main difference from their empirical setting is that we use product-level quantity and price information, rather than firm-level deflated sales. This enables us to identify markups for each firm-product-year triplet observation. We are forced to confront a number of new issues in order to implement our approach. Once we obtain markups, we compute marginal costs by simply dividing the observed prices by the estimated markups.

Consider the following production function for firm f that manufactures a product:

$$Q_{ft} = F_t(X_{ft}) \exp(\omega_{ft}) \tag{1}$$

where Q is physical output and X is a vector of inputs. To simplify the exposition, we assume for now that this producer is a single-product firm. In Section 4, we explicitly deal with the empirical challenges that arise with multi-product firms. The production function in (1) is general; we can consider alternative functional forms for F and allow the production technology to vary over time. There are only two assumptions that are essential for the subsequent analysis. First, productivity ω enters in log-additive form and is Hicks-neutral. Second, we assume that productivity is firm-specific. This second assumption follows a tradition in the trade literature that models productivity along these lines (e.g., Bernard et al. (2011)). For single-product firms, this assumption is of course redundant.

Although this production function is entirely standard, the advantage of our setting is that we directly observe output (Q) at the product level. This distinguishes our approach of estimating a production function from the traditional literature where researchers rely on deflated sales data. We discuss the estimation and the identification issues in Section 4.

We assume that producers minimize costs. Let $\mathbf{V}_{ft}$ denote the vector of variable inputs used by the firm. We use the vector $\mathbf{K}_{ft}$ to denote dynamic inputs of production. Any input that faces adjustment costs will fall into this category; capital is an obvious one, and our framework will include labor as well. We consider

---

[10]This approach to infer markups from production data began with Hall (1986). An alternative approach to estimating markups would be to assume a particular market structure, firm behavior and utility function of consumers, which jointly determine markups (e.g., Berry et al. (1995) or Goldberg (1995)). However, we explicitly want to avoid making the many functional form assumptions required to implement this approach. This approach is also less feasible when inferring markups across the entire manufacturing sector. We therefore believe that inferring markups from production data is the appropriate path in this setting, and one that can be easily implemented in a variety of contexts.

the firm's conditional cost function where we condition on the set of dynamic inputs $\mathbf{K}_{ft}$. The associated Lagrangian function is:

$$L(\mathbf{V}_{ft}, \mathbf{K}_{ft}, \lambda_{ft}) = \sum_{v=1}^{V} P_{ft}^v V_{ft}^v + \sum_{d=1}^{D} r_{ft}^d K_{ft}^d + \lambda_{ft} \left[Q_{ft} - Q_{ft}(\mathbf{V}_{ft}, \mathbf{K}_{ft}, \omega_{ft}) \right] \tag{2}$$

where P_{ft}^v and r_{ft}^d denote the firm's input prices for the variable inputs $v = 1, \ldots, V$ and the prices of dynamic inputs $d = 1, \ldots, D$, respectively. The first order condition for any variable input free of adjustment costs is

$$\frac{\partial L_{ft}}{\partial V_{ft}^v} = P_{ft}^v - \lambda_{ft} \frac{\partial Q_{ft}(.)}{\partial V_{ft}^v} = 0 \tag{3}$$

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

$$\frac{\partial Q_{ft}(.)}{\partial V_{ft}^v} \frac{V_{ft}^v}{Q_{ft}} = \frac{1}{\lambda_{ft}} \frac{P_{ft}^v V_{ft}^v}{Q_{ft}} \tag{4}$$

The left-hand side of the above equation represents the elasticity of output with respect to variable input V_{ft}^v (the "output elasticity"). The approach simply requires one freely adjustable input into production. This becomes important in settings, such as ours, where there are frictions in adjusting capital and labor. Define the markup μ_{ft} as $\mu_{ft} \equiv \frac{P_{ft}}{\lambda_{ft}}$. As De Loecker and Warzynski (2012) show, the cost-minimization condition can be rearranged to write the markup as:

$$\mu_{ft} = \theta_{ft}^v (\alpha_{ft}^v)^{-1} \tag{5}$$

where θ_{ft}^v denotes the output elasticity on variable input V^v and $\alpha_{ft}^v = \frac{P_{ft}^v V_{ft}^v}{P_{ft} Q_{ft}}$ is its expenditure share of revenue. This expression forms the basis for our approach. To compute the markup, we need the output elasticity and the share of the input's expenditure in total sales. We obtain the output elasticity by estimating the production function in (1). We compute the input's revenue share for single-product firms directly from the data.

Markups for multi-product firms are obtained using the exact same expression; the markup is computed for each product j produced by firm f at time t using:

$$\mu_{fjt} = \theta_{fjt}^v (\alpha_{fjt}^v)^{-1} \tag{6}$$

The only difference between expressions (5) and (6) is the presence of the product-specific subscript j. This seemingly small difference complicates the analysis considerably. In contrast to the single product setting, α_{fjt}^v is not directly observed in the data because firms do not report input expenditures by product.[11] Moreover, we require an estimate of the output elasticity for each product manufactured by each multi-product firm. In the next section, we discuss the identification strategy we employ in order to obtain

---

[11]To our knowledge, no dataset reports this information for all inputs.

the output elasticities for both single- and multi-product firms, and recover the firm-product specific input expenditures (the α^v_{fjt}) for the multi-product firms.

This approach provides markups that vary at the firm-product-year level. Since prices are observed at this same level of disaggregation, we can simply calculate the marginal costs from the observed price data:

$$mc_{fjt} = \frac{P_{fjt}}{\mu_{fjt}}. \tag{7}$$

We believe that this approach to recover markups is quite flexible and suitable for a setting where one wants to estimate markups for many industries. The approach requires the presence of at least one input that can be freely adjusted, but allows for frictions in the adjustment of capital and labor. This is important in a country like India that has had heavily regulated labor markets and is often associated with labor and capital market distortions. It does not impose assumptions on the returns to scale nor assumptions on the demand and market structure for each industry. Moreover, we do not require knowledge of the user cost of capital, a factor price that is difficult to measure in a country like India with distorted capital markets. This flexible approach allows us to be *a priori* agnostic about how prices, markups and marginal costs change with trade liberalization.

4   Empirical Framework

This section discusses our empirical framework. We first discuss the identification strategy and then describe the estimation routine.

Consider the log version of the general production function given in equation (1):

$$q_{fjt} = f^j(\boldsymbol{x}_{fjt}; \boldsymbol{\beta}) + \omega_{ft} + \epsilon_{fjt} \tag{8}$$

where lower case letters denote logs. We introduce the subscript *j* since many firms in our sample manufacture more than one product. The quantity of product j in firm f at time t, q_{fjt}, is produced using a set of firm-product-year specific inputs, x_{fjt}. The error term ϵ_{fjt} captures measurement error in recorded output as well unanticipated shocks to output. As noted earlier, the productivity term ω_{ft} is assumed to vary at the firm level. Our goal is to estimate the parameters of the production function – the output elasticities $\boldsymbol{\beta}$ – for each product.

In the subsequent subsections, we discuss the restrictions we impose on equation (8) given the nature of our data and how we identify the output elasticities. The key challenge is that the data do not record how inputs are allocated across outputs within a firm. For single-product firms, this concern is of course not an issue. The challenge is how to deal with the unobserved input allocation among multi-product firms while also controlling for unobserved productivity shocks.

4.1 Identification Strategy: The Use of Single-Product Firms

By writing the production function in equation (8) in terms of physical output rather than revenue, we exploit our ability to observe separate information on quantities and prices for each product manufactured by firms. This advantage eliminates the concern about a "price bias" that arises when one must deflate sales by an industry-level price index to obtain firm output (for a detailed discussion, see De Loecker (2011)). For single-product firms, the inputs are allocated to the only product these firms manufacture and so we are not concerned about how to allocate inputs across products. The typical concern that arises in the productivity estimation is correlation between the unobserved productivity shock and inputs. Removing this bias has been the predominant focus of the production function estimation literature and following the insights of Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2006), we use proxy estimators to deal with this correlation.[12]

For multi-product firms, a *new* identification problem arises since the data do not record how the inputs are allocated across the products within a firm. To understand this, denote the log of the share of input X in the production of product j as $\rho^X_{fjt} = x_{fjt} - x_{ft}$, for any input $X = \{L, M, K\}$, where L is labor, M is materials and K is capital. We only observe firm-level inputs X_{ft} and not how each of them is allocated across products. Substituting this expression into equation (8) yields:

$$q_{fjt} = f^j(\boldsymbol{x}_{ft}; \boldsymbol{\beta}) + \omega_{ft} + A_{fjt}(\rho^X_{fjt}, \boldsymbol{x}_{ft}, \beta) + \epsilon_{fjt} \qquad (9)$$

where x_{ft} denotes the log of inputs X_{ft}. For multi-product firms, the production function contains an additional component in the error term, $A(.)$, that will generally be a function of the unobserved input shares (ρ^X_{fjt}), the firm level inputs (x_{ft}) and the production function coefficients, $\boldsymbol{\beta}$. The expression (9) clearly demonstrates that $A(.)$ is correlated - by construction - with the inputs on the right-hand-side of the production function which results in biased estimates of $\boldsymbol{\beta}$.[13] We could deal with this bias by taking a stand on the underlying demand function for each product and model how firms compete in that market, but as discussed earlier, we wish to avoid these assumptions in this paper.[14]

We propose an identification strategy that does not require assumptions on the input allocation across products. The strategy is to obtain estimates of the production function using a sample of single-product firms, since as discussed above, the input allocation problem does not exist for these firms. This sample

---

[12] Unobserved input prices might still bias estimates of the production function. We discuss this issue in detail in Section 4.3.2.

[13] To illustrate the bias, consider a translog production function with a single factor of production, labor. The production function in equation (9) would be: $q_{fjt} = \beta_l l_{fjt} + \beta_{ll} l^2_{fjt} + \omega_{ft} + \epsilon_{fjt}$. We observe labor at the firm level, l_{ft}. The labor allocated for the production of j can be written as: $l_{fjt} = \rho_{fjt} + l_{ft}$. Substitution yields:

$$q_{fjt} = \beta_l l_{ft} + \beta_{ll} l^2_{ft} + \underbrace{\beta_l \rho_{fjt} + \beta_{ll} \left(\rho_{fjt}\right)^2 + 2\beta_{ll}(\rho_{fjt} l_{ft})}_{=A_{fjt}(\rho_{fjt}, \boldsymbol{x}_{ft}, \boldsymbol{\beta})} + \omega_{ft} + \epsilon_{fjt} \qquad (10)$$

Clearly, the component $A_{fjt}(\rho_{fjt}, \boldsymbol{x}_{ft}, \boldsymbol{\beta})$ is unobserved and would be subsumed in the productivity term, ω_{ft}. This would result in biased estimates of β_l since l_{ft} would be correlated with this composite error.

[14] The few productivity studies that have explicitly dealt with multi-product firms had to make assumptions on how inputs are allocated. For example, De Loecker (2011) allocates inputs equally across products and Foster et al. (2008) allocate inputs according to revenue share. These approaches rest on implicit assumptions about markups, i.e., markups are the same across the products manufactured by a firm. These assumptions are innocuous in the above studies since the authors are not interested in recovering markups, but they would be inappropriate in our setting since estimation of markups is the primary objective of this paper.

consists of those firms that produce a single product *at a given point in time;* it includes firms that may eventually add additional products later in the sample period. This feature of the sample is important since many firms start off as single product firms and add products during our sample. Our analysis uses these firms in conjunction with firms that always manufacture a sole product.

This approach implies that we use an unbalanced panel of firms for estimation; the imbalance occurs by removing firms from the estimation if they add a product over the sample period. The unbalanced panel is important because it allows for the non-random event that a firm becomes a multi-product producer based on productivity shocks. This non-random event 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 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. In addition, we use a sample selection correction procedure to account for the possibility that the productivity threshold determining the transition of a firm from single- to multi-product status is correlated with the right-hand side variables of the production function (in particular, capital). We describe the sample selection correction procedure in 4.3.3.[15]

4.2   Economies of Scope and Relationship to Cost Function Estimation

In this subsection, we discuss the assumptions underlying the identification strategy laid out in the previous subsection. We explain that despite relying on single-product firms in our estimation, our strategy does not rule out economies of scope and relate our approach to cost function estimation.

Our identification strategy assumes that the production technology is product-specific rather than specific to the firm. Once we obtain the coefficients of the production function from the sample of single-product firms, we use those on the multi-product firms that produce the same product. This identification strategy rules out *physical* synergies across products. 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. This assumption is not unusual. In fact, it is implicitly used in all productivity studies when researchers pool single- and multi-product firms in the same industry and estimate an industry-level production function. We feel that this restriction is mild, especially given the current practice within the production function estimation literature and weighed against the costs of assuming a market structure/demand system that would dictate how to allocate inputs across products.

This approach does not rule out economies of scope, which may be important for multi-product firms. To be precise, Baumol et al. (1983) speak of economies of scope in production, if the cost function is subadditive: $c_{ft}(q_1, q_2) \leq c_{ft}(q_1) + c_{ft}(q_2)$ where $c_{ft}(.)$ is a firm's cost curve. While our framework rules out *production* synergies, it allows for economies of scope through *cost* synergies. For example, economies of

---

[15] 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, since Prowess covers the medium and large firms in India which are relatively stable.

scope can emerge if multi-product firms face lower input prices than single-product firms due to their size or the negotiation of better deals with their suppliers. Furthermore, we allow for potential differences in productivity between the two types of firms that might be caused by differences in management practices or organizational structures.

An alternative way of explaining the assumptions underlying our approach is to express them in terms of the cost function rather than the production function. A multi-product firm faces the short-run cost function, written in a general form as:

$$C(\mathbf{Q}) = \Phi(\Omega)C(\mathbf{Q}, \mathbf{W}, \beta) + F(\iota(\mathbf{Q}))$$

where C denotes the total costs for a firm producing a vector of outputs $\mathbf{Q}$, $\Phi(\Omega)$ denotes the impact of factor-neutral productivity on costs, $\mathbf{W}$ denotes a vector of input prices, F are the fixed costs (which would be zero in long-run cost function), and $\iota(.)$ is an indicator that takes the value of 1 if a firm produces a particular product in the vector $\mathbf{Q}$ and is zero otherwise. The assumption we impose is that the function $C(\mathbf{Q}, \mathbf{W}, \beta)$is the same across single- and multi-product firms. However, costs between the two types of firms can still differ because of: 1) factor-neutral productivity differences reflected in $\Phi(\Omega)$; 2) differences in factor prices reflected by $\mathbf{W}$; and 3) (in the short run) the amortization of fixed costs F across more products for multi-product firms.

The above representation raises the natural question of why we do not exploit the duality between production and cost function and estimate a multi-product cost function as done in an earlier literature. The main reason for focusing on the production function is that we do not have information on wages and the firm-specific user cost of capital, which we would need in order to estimate a cost function. Further, 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 during our sample period 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, our approach to estimate production functions from single-product firms while accounting for the potential selection bias is an appealing alternative.

4.3   Estimation

This section discusses the production function estimation on the sub-sample of single-product firms.

4.3.1   The Basic Setup

Since we focus on a subsample of single-product firms for the estimation routine, we drop the subscript j to simplify notation.[16] Since we have a physical measure of output, we control for differences in units across

---

[16]To be clear, we observe many firms manufacturing the same product j, but we subsume the subscript j when describing firms that only manufacture one product.

products by using unit fixed effects throughout the estimation procedure.

The production function we estimate is of the form:[17]

$$q_{ft} = f(\boldsymbol{x}_{ft}; \boldsymbol{\beta}) + \omega_{ft} + \epsilon_{ft} \tag{12}$$

where we subsume the constant term in productivity, collect all inputs in $\boldsymbol{x}_{ft}$ and $\boldsymbol{\beta}$ is the vector of coefficients describing the transformation of inputs into output. In the case of a translog production function, the vector of log inputs $\boldsymbol{x}_{ft}$ are labor, material and capital, their squares, and their interaction terms; the vector of coefficients is $\boldsymbol{\beta} = (\beta_l, \beta_m, \beta_k, \beta_{ll}, \beta_{mm}, \beta_{kk}, \beta_{lm}, \beta_{lk}, \beta_{mk}, \beta_{lmk})$.

We deal with the potential correlation between unobserved productivity shock and input choices by relying on the approach taken by Ackerberg et al. (2006), who modify Olley and Pakes (1996) and Levinsohn and Petrin (2003).[18]

The law of motion for productivity is:

$$\omega_{ft} = g_{t-1}(\omega_{ft-1}, \tau_{it-b}^{\text{output}}, \tau_{it-b}^{\text{input}}, I_{ft-1}^x) + \xi_{ft} \tag{13}$$

where $b = \{0, 1\}$ and I_{ft-1}^x is a dummy indicating whether the firm is exporting. Output tariffs enter the law of motion for productivity because we expect the competitive pressure associated with an increase in import competition to induce firms to undertake changes (e.g., reorganization; reduction of X-inefficiencies) that increase productivity. Input tariffs enter because they lead firms to use new imported input varieties that affect measured productivity (see Halpern et al. (2011) and Goldberg et al. (2010a)). The notation implies that we allow trade liberalization to impact a firm's productivity instantaneously or with a lag.[19]

We assume that the dynamic inputs–labor and capital–are chosen prior to observing ξ_{ft}. Materials m_{ft} are chosen when the firm learns its productivity. These timing assumptions treat labor, in addition to capital, as an input that faces adjustment costs, which we feel is appropriate in a country like India where labor markets are not flexible (see Besley and Burgess (2004)).

We proxy for productivity by inverting the materials demand function:

$$m_{ft} = m_t(l_{ft}, k_{ft}, \omega_{ft}, \mathbf{z}_{ft}). \tag{14}$$

The vector $\mathbf{z}_{ft}$ contains all additional variables that affect a firm's demand for materials. It includes output prices, product dummies, product market shares, and the input (τ_{it}^{input}) and output tariffs ($\tau_{it}^{\text{output}}$) that the firm faces on the product. The subscript i on the tariff variables denotes an industry to indicate that tariffs

---

[17]Our main specification is the translog and is given by

$$q_{ft} = \beta_l l_{ft} + \beta_{ll} l_{ft}^2 + \beta_k k_{ft} + \beta_{kk} k_{ft}^2 + \beta_m m_{ft} + \beta_{mm} m_{ft}^2 + \beta_{lk} l_{ft} k_{ft} + \beta_{lm} l_{ft} m_{ft} + \beta_{mk} m_{ft} k_{ft} + \beta_{lmk} l_{ft} m_{ft} k_{ft} + \omega_{ft} \tag{11}$$

[18]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. In developing countries, many firms report zero investment, which further complicates the use of investment as a proxy.

[19]De Loecker (2010) discusses the importance of including all variables which may affect productivity in the law of motion $g_t(.)$.

vary at a higher level of aggregation than products.[20] The vector also includes product fixed effects and the price of the product.[21] These variables capture demand shocks that affect a firm's output, which in turn affects a firm's demand for materials. For example, the product's price affects the quantity produced which in turn affects a firm's input expenditure. 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, capital and labor, productivity and the aforementioned variables.

As is usual in the proxy approach, our estimation proceeds in two steps.[22] We begin by inverting equation (14) to obtain the proxy for firm productivity, $\omega_{ft} = h_t(m_{ft}, l_{ft}, k_{ft}, \mathbf{z}_{ft})$. Following Ackerberg et al. (2006), in the first stage, we run:

$$q_{ft} = \phi_t(l_{ft}, k_{ft}, m_{ft}, \mathbf{z}_{ft}) + \epsilon_{ft} \tag{15}$$

to obtain estimates of expected output $(\widehat{\phi}_{ft})$ and an estimate for the residual ϵ_{ft}.[23]

The first stage separates the effect of ω_{ft} on output from the effects of unanticipated shocks ϵ_{ft} on output. After the first stage, we compute productivity as $\widehat{\omega}_{ft} = \widehat{\phi}_{ft} - f(\boldsymbol{x}_{ft}; \boldsymbol{\beta})$ for any vector of β. The second stage provides estimates of all production function coefficients by relying on the law of motion for productivity. By non-parametrically regressing $\widehat{\omega}_{ft}(\boldsymbol{\beta})$ on its lag $\widehat{\omega}_{ft-1}(\boldsymbol{\beta})$, the set of tariff variables $(\boldsymbol{\tau}_{it})$, and the firm export indicator I^x_{ft-1}, we recover the innovation to productivity $\xi_{ft}(\boldsymbol{\beta})$ for a given β.

To estimate the parameter vector β, we use moments that are now standard in this literature based on degree of variability of a given input. 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 material expenditure is the flexible input, we construct its moments using lagged materials.[24] For labor and capital, we construct moments using current and lagged values. The moments are:

$$E\left(\boldsymbol{\xi}_{ft}(\boldsymbol{\beta})\mathbf{Y}_{ft}\right) = 0 \tag{16}$$

where $\mathbf{Y}_{ft}$ contains lag materials, the one-year lag labor, current capital, and their higher order terms[25]:

$$\boldsymbol{Y}_{ft} = \{l_{ft-1}, l^2_{ft-1}, m_{ft-1}, m^2_{ft-1}, k_{ft}, k^2_{ft}, l_{ft-1}m_{ft-1}, l_{ft-1}k_{ft}, m_{ft-1}k_{ft}, l_{ft-1}m_{ft-1}k_{ft}\} \tag{17}$$

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

---

[20] For example, all tea products (industry 1549) face the same output and input tariffs.

[21] As we discuss below, we estimate the production functions at the two-digit sector level; the product fixed effects are identified since there are many products within each sector.

[22] In principle, we could estimate both stages jointly using a system GMM estimator, but in practice this is difficult to implement because the first stage contains many parameters including product fixed effects (see De Loecker (2011) for more details). We therefore adopt a two-stage approach.

[23] The set of estimated unit fixed effects are included in our measure of ϵ_{ft}.

[24] In our setting, input tariffs are serially correlated and since they affect input prices, input prices are serially correlated over time.

[25] We have also experimented with including current labor and the one-year lag in capital in $\mathbf{Y}_{ft}$ and the results were virtually unchanged.

respond to these shocks; moreover, the degree of adjustment can vary across firms and time.

We estimate the model using a GMM procedure on a sample of firms that manufacture a single product for at least three consecutive years. We choose three years since the moment conditions require at least two years of data because of the lagged values; we add an extra year to allow for potential measurement error in the precise timing of a new product introduction. In principle, one could run the estimation separately for each product. In practice, our sample size is such that we estimate (12) at the two-digit sector level.

4.3.2 Input Price Variation Across Firms

Although we observe quantities and prices for firm outputs, we do not have this information for firm inputs and so we follow the standard practice of deflating input expenditures using price indexes. If input prices vary across firms, this practice results in unobserved input price variation across firms generating an analogous concern that arises if output revenue is deflated by a common deflator. This concern is present even when data on worker wages are available, as researchers typically do not have information on firm-specific materials prices, and never observe the firm-specific user cost of capital. But while this issue always arises when researchers do not observe input-level prices and quantities, the problem is exacerbated in our setting because we use physical quantity data rather than deflated revenues.

To understand this concern, suppose we want to estimate the productivity of two shirt producers. Imagine that the only difference between the firms is that one firm manufactures shirts made from expensive silk while the other manufactures shirts made from cheaper cotton. Otherwise, the two firms have identical productivity: they utilize the same quantity of inputs (balls of yarn, number of workers and number of sewing machines) to produce the same quantity of shirts. In our data, we would not observe the price of the materials (or workers or capital) and so we would deflate the firms' input expenditures by a common deflator. Since we do not account for variation in input prices, we will estimate a lower productivity for the silk shirt manufacturer because it sells the same quantity as its cotton producer counterpart, but has higher (deflated) input expenditures. Our estimation procedure would yield downward biased estimates of the production function coefficients and find productivity differences between the two firms, even though they are exactly identical. Ultimately, the bias in the production function coefficients would lead to biased markup estimates.[26]

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, **?** document that producers of more expensive products also use more expensive inputs. Formally, we begin by modifying our notation to account for the fact that we observe deflated input expenditures, rather than input quantities. We use $\tilde{\mathbf{x}}_{ft}$ to denote the firm's vector of deflated input expenditures and re-write the production function in (12) as:

$$q_{ft} = f(\tilde{\boldsymbol{x}}_{ft}; \boldsymbol{\beta}) + \omega_{ft} + B(\mathbf{W}_{ft}; \mathbf{W}_{ft} \times \tilde{\mathbf{x}}_{ft}; \boldsymbol{\beta}) + \epsilon_{ft}. \tag{18}$$

---

[26]See also the related discussion in Katayama et al. (2009).

Let $B_{ft} \equiv B(\mathbf{W}_{ft}; \mathbf{W}_{ft} \times \tilde{\mathbf{x}}_{ft}; \boldsymbol{\beta})$. The term B_{ft} appears because the input deflators are sector-, rather than firm-specific. In our translog specification, this term captures deviations of the firm input prices from the sector deflators, $\mathbf{W}_{ft}$, their interactions with the deflated expenditures, $\mathbf{W}_{ft} \times \tilde{\mathbf{x}}_{ft}$ and the parameter vector $\boldsymbol{\beta}$. It is clear from (18) that B_{ft} is correlated with the deflated inputs and will generate biased estimates of the production function.[27]

The presence of B_{ft} is not a concern for the first stage of the estimation since it is not correlated with random unanticipated shocks to production ϵ_{ft}. The purpose of the first stage is simply to estimate expected output $\widehat{\phi}_{ft}$, net of unanticipated shocks ϵ_{ft}.

However, in the second stage of estimation, we need to recover an estimate of the true innovation to productivity ξ_{ft} in order to form moments in equation (17). The presence of B_{ft} in (18) violates these orthogonality conditions. To see this problem, define $\tilde{\omega}_{ft}$ as the composite term:

$$\tilde{\omega}_{ft} = \omega_{ft} + B_{ft}. \tag{19}$$

The expected output from the first stage will yield a second-stage estimate of productivity $\tilde{\omega}_{ft}$ that contains the unobserved variation in input prices. Using the law of motion for productivity in equation (13), let the *measured* innovation to productivity be $\tilde{\xi}_{ft} = \tilde{\omega}_{ft} - g_{t-1}(\tilde{\omega}_{ft-1}, \tau_{it-b}^{\text{output}}, \tau_{it-b}^{\text{input}}, I_{ft-1}^{x})$ where $b = \{0, 1\}$. Using (19) and the law of motion for productivity (13), we can express the measured innovation to productivity $\tilde{\xi}_{ft}$ as a function of true innovation in productivity ξ_{ft}:

$$\tilde{\xi}_{ft} = \xi_{ft} + B_{ft} - e(B_{ft-1}, \tau_{it-b}^{\text{output}}, \tau_{it-b}^{\text{input}}, I_{ft-1}^{x}) \tag{20}$$

where $e(B_{ft-1}, \tau_{it-b}^{\text{output}}, \tau_{it-b}^{\text{input}}, I_{ft-1}^{x}) = g_{t-1}(\tilde{\omega}_{ft-1}, \tau_{it-b}^{\text{output}}, \tau_{it-b}^{\text{input}}, I_{ft-1}^{x}) - g_{t-1}(\omega_{ft-1}, \tau_{it-b}^{\text{output}}, \tau_{it-b}^{\text{input}}, I_{ft-1}^{x})$. Equation (20) illustrates that $\tilde{\xi}_{ft}$ is a function of current and lagged input prices, which are obviously correlated with the current and lagged input expenditures. The term $B_{ft} - e(B_{ft-1}, \tau_{it-b}^{\text{output}}, \tau_{it-b}^{\text{input}}, I_{ft-1}^{x})$ creates a correlation between the measured innovation $\tilde{\xi}_{ft}$ and measured input use, so the moment conditions are violated.

Our approach for dealing with this bias is to explicitly control for the B_{ft} term in the second-stage. To implement this approach, we need to take a stand on the factors that drive input price variation across firms. Our premise is that, conditional on observable variables, the only reason input prices vary across firms manufacturing the same product is due to quality differences. If two firms produce the same quality, they face the same input prices. If they produce different qualities, their input prices will differ. While there

---

[27] To illustrate this bias, consider again the translog production function with a single factor of production, labor, that we specified in Footnote 13. Here, we focus on the bias that arises on single product firms due to unobserved input price variation. We observe the total wage bill of the firm which we deflate with a sector-specific index to obtain the firm's deflated (log) labor expenditure, $\tilde{l}_{ft}$. Let $\tilde{l}_{ft} = l_{ft} + w_{ft}^{L}$, where w_{ft}^{L} is the (log) *firm*-specific deviation of the price of labor (e.g., the firm's wage) from the sectoral average. Since we do not observe w_{ft}^{L}, we would estimate:

$$q_{ft} = \beta_l \tilde{l}_{ft} + \beta_{ll} \tilde{l}_{ft}^2 \underbrace{- \beta_l w_{ft}^{L} + \beta_{ll} \left(w_{ft}^{L}\right)^2 - 2\beta_{ll}(w_{ft}^{L}\tilde{l}_{ft})}_{=B_{ft}(\mathbf{W}_{ft}; \mathbf{W}_{ft} \times \tilde{\mathbf{x}}_{ft}; \boldsymbol{\beta})} + \omega_{ft}.$$

Clearly, the component $B_{ft}(\mathbf{W}_{ft}; \mathbf{W}_{ft} \times \tilde{\mathbf{x}}_{ft}; \boldsymbol{\beta})$ is unobserved and would be subsumed in the productivity term, ω_{ft}. This would result in biased estimates of β_l since $\tilde{l}_{ft}$ would be correlated with this composite error.

may be other factors that cause input prices to vary across firms, we believe that quality differences are the first-order concern in our setting and the primary factor generating input price variation.[28]

We provide a formal model that rationalizes our approach in the Appendix. Here, we sketch the main argument and provide the economic intuition underlying our empirical strategy. Our main premise is that manufacturing high quality products requires high quality inputs, and 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 **?**). This complementarity is important in our framework because it allows us to express the prices of *all* inputs facing a firm as a function of a single index of product quality.[29] The appendix 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. Before we explain how to control for output quality, it is important to note that our approach does not assume that products are only vertically differentiated. It can allow for costless horizontal differentiation, but 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)).

Our control for quality is a flexible function of output prices, market share and product dummies. This specification does not force 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 controls for market share or product characteristics. In a simple logit model, quality will be a function of output prices and market shares (see Khandelwal (2010) for a detailed exposition). In more general models, such as a nested logit or random coefficients model, quality will be a function of additional variables, such as product characteristics, conditional market shares, etc. While we do not observe product characteristics, product dummies accommodate these more general demand specifications. We also note that 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 will be dictated by the (unspecified) underlying demand and market structures and behavioral assumptions. Accordingly, our specification is consistent with *any* degree of pass-through between input and output prices. The empirical strategy uses a flexible function of output price, market share and product dummies to control for output quality, which reflects input quality and in turn, input prices.

Formally, let v_{ft} denote output quality. We can re-write the production function as

$$q_{ft}(v_{ft}) = f(\tilde{\boldsymbol{x}}_{ft}(v_{ft}); \boldsymbol{\beta}) + \omega_{ft} + B(\mathbf{W}_{ft}(v_{ft}); \mathbf{W}_{ft}(v_{ft}) \times \tilde{\mathbf{x}}_{ft}(v_{ft}); \boldsymbol{\beta}) + \epsilon_{ft}. \tag{21}$$

---

[28]Note that this approach allows for input prices to vary across geographic locations; the crucial assumption is that *conditional on location*, which is observed, the only reason that input prices vary across two firms producing the same product is that the qualities of their products are different.

[29]We maintain the assumption of perfectly competitive input markets throughout the analysis. We assume that conditional on a particular input quality, all firms face the same input price. In principle, we could allow input prices to vary across regions (while maintaining the assumption that firms within a region producing the same quality face the same input prices), but in practice we do not have reliable information on location of production in the data.

The specification makes explicit that both the deflated firm input expenditures $\tilde{x}_{ft}$ and the (unobserved) firm input prices $\mathbf{W}_{ft}$ are functions of product quality v_{ft}. Product quality is expressed as a flexible function of output price, market share, and product dummies: $v_{ft} = v_t(p_{ft}, ms_{ft}, I_{ft})$. Since input prices are an increasing function of output quality, input prices are also a function of these variables, i.e., $\mathbf{W}_{ft} = w_t(p_{ft}, ms_{ft}, I_{ft})$. We can then write the term B_{ft} as a flexible polynomial in output price, market share, product dummies, and interaction of these variables with the deflated input expenditures. That is, we specify B_{ft} as a polynomial d_t, so that: $B_{ft} \equiv d_t[p_{ft}, ms_{ft}, I_{ft}, (p_{ft}, ms_{ft}, I_{ft}) \times \tilde{x}_{ft}]$. The production function that we take to the data becomes:

$$q_{ft} = f(\tilde{x}_{ft}; \boldsymbol{\beta}) + \omega_{ft} + d_t[p_{ft}, ms_{ft}, I_{ft}, (p_{ft}, ms_{ft}, I_{ft}) \times \tilde{x}_{ft}] + \epsilon_{ft} \qquad (22)$$

In the first stage, we estimate:

$$q_{ft} = \phi_t(\tilde{x}_{ft}, z_{ft}) + \epsilon_{ft} \qquad (23)$$

where z_{ft} includes p_{ft}, ms_{ft} and I_{ft}. Productivity is given by $\omega_{ft} = \phi_{ft} - f(\tilde{x}_{ft}; \boldsymbol{\beta}) - d_t(.) = \phi_{ft} - f(\tilde{x}_{ft}; \boldsymbol{\beta}) - \mathbf{b}_{ft}\boldsymbol{\delta}$, where $\mathbf{b}_{ft}$ contains the price, market share, their interactions with input expenditures, and product dummies.[30] The coefficients on the variables in $\mathbf{b}_{ft}$ are themselves not of direct interest since we compute markups using the estimates of $\boldsymbol{\beta}$. Next, we obtain the innovation in productivity, $\xi_{ft}(\boldsymbol{\beta}, \boldsymbol{\delta})$, in the usual way as the difference between realized and expected productivity. The moments that identify the parameters become:

$$E\left(\xi_{ft}(\boldsymbol{\beta}, \boldsymbol{\delta})\mathbf{Y}_{ft}^*\right) = 0 \qquad (24)$$

where $\mathbf{Y}_{ft}^*$ now contains in addition to the variables mentioned in the previous subsection, lagged output prices, lagged market shares, (double) lagged input tariffs, and their appropriate interactions with the input terms.[31] The reason we use these variables to form additional moment conditions, is that we jointly identify the production function coefficients and the coefficients capturing the input price variation, δ, which describe the relationship between the elements of $d_t(.)$ and input prices (although the latter coefficients are not of direct interest). 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.

4.3.3 The Sample Selection Correction

In this subsection, we discuss the correction procedure we employ to account for the potential sample selection bias arising from using only single-product firms in the estimation.

As noted earlier, rather than relying on firms that remain single-product producers for the entire sam-

---

[30]We do not interact the product dummies with the other elements of $\mathbf{b}_{ft}$ to reduce the dimensionality of the estimation.

[31]We form moments using the actual innovation ξ_{ft} since we use $d_t(.)$ to proxy for the unobserved input price variation. We therefore recover physical productivity shocks (ω_{ft}) given parameters ($\boldsymbol{\beta}, \boldsymbol{\delta}$). If input prices or quality do not vary across firms, the estimates from this approach would be identical to those obtained when using the method described in the Section 4.3.1.

ple, our estimation uses an unbalanced panel of single-product firms. As in Olley and Pakes (1996), the unbalanced panel improves on the selection problem since it includes firms that may eventually become multi-product producers in response to productivity shocks. Even with an unbalanced panel, a selection bias may still arise if the productivity threshold determining a firm's decision to become multi-product is correlated with the inputs (in particular, capital). We address this bias by introducing a correction for sample selection and modifying the law of motion for productivity following a similar approach to Olley and Pakes (1996).

The underlying model behind our sample selection correction is one where the number of products manufactured by firms increases with productivity. There are several multi-product firm models that generate this correlation and the one that matches our setup most closely is Mayer et al. (2011). 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.[32]

The cutoff point that is 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, there is no selection bias and we obtain consistent estimates of production function coefficients (as long as we use the unbalanced panel). 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 and the firm's information set at time $t - 1$ (we assume the decision to add a product is made in the previous period). The state variables in our setting include capital, labor, productivity and all variables in $\mathbf{z}_{ft}$ that are serially correlated (i.e., input and output tariffs, output prices and product dummies).[33] 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 (S) and *0* otherwise, and let $\bar{\omega}_{ft}$ be the productivity threshold a firm has to clear in order to produce more than one product.

---

[32] Alternative models of multi-product firms, 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 in our setup since product dropping is not a prominent feature of our data (Goldberg et al. (2010b)). Moreover, in the empirical section we find strong support that firms' marginal costs are lower on their core competent products (products that have higher sales shares).

[33] If we modeled the behavior of multi-product firms that shed products to become single-product producers, the number of products would be a natural state variable to include in *z*. So while we abstract away from product deletions since they are rare in our data, our approach can accommodate this case as well.

The selection rule can be rewritten as:

$$
\begin{aligned}
\Pr(\chi_{ft} = 1) &= \Pr\left[\omega_{ft} \leq \bar{\omega}_{ft}(l_{ft}, k_{ft}, \mathbf{z}_{ft}) | \bar{\omega}_{ft}(l_{ft}, k_{ft}, \mathbf{z}_{ft}), \omega_{ft-1}\right] \qquad (25) \\
&= \kappa_{t-1}(\bar{\omega}_{ft}(l_{ft}, k_{ft}, \mathbf{z}_{ft}), \omega_{ft-1}) \\
&= \kappa_{t-1}(l_{ft-1}, k_{ft-1}, i_{ft-1}, \mathbf{z}_{ft-1}, \omega_{ft-1}) \\
&= \kappa_{t-1}(l_{ft-1}, k_{ft-1}, i_{ft-1}, \mathbf{z}_{ft-1}, m_{ft-1}) \equiv S_{ft-1}
\end{aligned}
$$

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[34], and $\omega_{ft-1} = h_t(l_{ft-1}, k_{ft-1}, \mathbf{z}_{ft-1}, m_{ft-1})$ to arrive at the last equation.[35] The law of motion for productivity now becomes:

$$
\omega_{ft} = g'_{t-1}(\omega_{ft-1}, \tau_{it-b}^{\text{input}}, \tau_{it-b}^{\text{output}}, I_{ft-1}^x, \bar{\omega}_{ft}) + \xi_{ft} \qquad (26)
$$

where $b = \{0, 1\}$.

As in Olley and Pakes (1996), we have two different indexes of firm heterogeneity, the productivity and the productivity cutoff point. Note that $S_{ft-1} = \kappa_{t-1}(\omega_{ft-1}, \bar{\omega}_{ft})$ and therefore $\bar{\omega}_{ft} = \kappa_{t-1}^{-1}(\omega_{ft-1}, S_{ft-1})$. Plugging this last expression into the modified law of productivity gives:

$$
\omega_{ft} = g'_{t-1}(\omega_{ft-1}, \tau_{it-b}^{\text{input}}, \tau_{it-b}^{\text{output}}, I_{ft-1}^x, S_{ft-1}) + \xi_{ft} \qquad (27)
$$

This is the law of motion we use to form the moments in the second stage of the estimation. To obtain S_{ft-1}, we estimate by industry, a probit that regresses a firm's single-product status on materials, capital, labor, the tariff variables, the output price of the product the firm currently produces, product and time dummies. Once the probit is estimated, we construct the propensity score $\hat{S}_{ft-1}$–the predicted probability that the firm remains single-product at time t given the firm's information set at time $t-1$–and insert it in the modified law of motion for productivity.

4.4 Markups and Marginal Costs

4.4.1 Single-Product Firms

We can now apply our framework from Section 3 to compute markups and marginal costs using the estimates of the production function. Using equation (5), we compute markups $\hat{\mu}_{ft} = \hat{\theta}_{ft}^M (\alpha_{ft}^M)^{-1}$, where the estimated output elasticities $\hat{\theta}_{ft}^M$ on materials are computed using the estimated coefficients of the production function[36] and α_{ft}^M, the revenue share of materials, is data. We then use the markup definition in conjunction with the price data to recover the firm's marginal cost, mc_{ft}, at each point in time as

---

[34]The accumulation equation for capital is: $k_{ft} = (1-\delta)k_{ft-1} + i_{ft-1}$, where δ is the depreciation rate of capital.

[35]This specification 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 $\mathbf{z}$ and ω. So all variables entering the non-parametric function $\kappa_{t-1}(.)$ help predict the firm's employment at time t.

[36]The expression for the materials output elasticity is: $\hat{\theta}_{ft}^M = \hat{\beta}_m + 2\hat{\beta}_{mm}m_{ft} + \hat{\beta}_{lm}l_{ft} + \hat{\beta}_{mk}k_{ft} + \hat{\beta}_{lmk}l_{ft}k_{ft}$.

$$\widehat{mc}_{ft} = \frac{P_{ft}}{\widehat{\mu}_{ft}} .^{37}$$

4.4.2 Multi-Product Firms

We now discuss how we obtain markups and marginal costs for the multi-product firms.

As shown in equations (6) and (7), computing markups and marginal costs requires the product-specific output elasticity on materials and product-specific revenue shares for materials. We obtain the output elasticity from the 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_{fjt}^M.

Let $\rho_{fjt} = \ln\left(\frac{\tilde{X}_{fjt}}{\tilde{X}_{ft}}\right)$ be product j's input cost share, where $\tilde{X}_{ft}$ denotes total deflated expenditures on each input by firm f at time t. We assume that this share does not vary across inputs. We solve for ρ_{fjt} as follows. We first eliminate unanticipated shocks and measurement error from the output data by following the same procedure as in the first stage of our estimation routine in Section 4.3.1 for the single-product firms. We project output quantity, q_{fjt}, on interactions of all inputs, output and input tariffs, the output price, product dummies and time dummies and obtain the predicted values. We next compute a firm-product-specific term $\widehat{\omega}_{fjt} : \widehat{\omega}_{fjt} \equiv E(q_{fjt}) - f(\tilde{x}_{ft}; \hat{\beta}).^{38}$ From (9), this becomes:

$$\widehat{\omega}_{fjt} = \omega_{ft} + A_{fjt}(\rho_{fjt}, x_{ft}, \hat{\beta}) \tag{28}$$

$$= \omega_{ft} + \hat{a}_{ft}\rho_{fjt} + \hat{b}_{ft}\rho_{fjt}^2 + \hat{c}_{ft}\rho_{fjt}^3 \tag{29}$$

where the second equation follows from applying our translog functional form. The terms $\hat{a}_{ft}$, $\hat{b}_{ft}$, and $\hat{c}_{ft}$ are functions of the estimated parameter vector $\hat{\beta}.^{39}$

With an estimate of $E(q_{fjt})$, we can construct $\widehat{\omega}_{fjt}$ for each multi-product firm observation (firm-year-product triplet). For each year, we obtain the firm's productivity and input allocations, the $J+1$ unknowns

---

[37] Output elasticities will vary across sectors because we estimate β separately for each sector. Moreover, given our translog specification, even single-product firms *within* a sector will have different output elasticities.

[38] The estimated parameter vector $\hat{\beta}$ is already purged of the bias arising from input price variation discussed in Section 4.3.2.

[39] For the translog production function:

$$\hat{a}_{ft} = \hat{\beta}_l + \hat{\beta}_m + \hat{\beta}_k + 2\left(\hat{\beta}_{ll}l_{ft} + \hat{\beta}_{mm}m_{ft} + \hat{\beta}_{kk}k_{ft}\right) + \hat{\beta}_{lm}\left(l_{ft} + m_{ft}\right)$$
$$+ \hat{\beta}_{lk}\left(l_{ft} + k_{ft}\right) + \hat{\beta}_{mk}\left(m_{ft} + k_{ft}\right) + \hat{\beta}_{lmk}(l_{ft}m_{ft} + l_{ft}k_{ft} + m_{ft}k_{ft})$$
$$\hat{b}_{ft} = \hat{\beta}_{ll} + \hat{\beta}_{mm} + \hat{\beta}_{kk} + \hat{\beta}_{lm} + \hat{\beta}_{lk} + \hat{\beta}_{mk} + \hat{\beta}_{lmk}\left(l_{ft} + m_{ft} + k_{ft}\right)$$
$$\hat{c}_{ft} = \hat{\beta}_{lmk}$$

$\left(\omega_{ft}, \rho_{f1t}, \ldots, \rho_{fJt}\right)$, by solving a system of $J+1$ equations:

$$\widehat{\omega}_{1ft} = \omega_{ft} + \hat{a}_{ft}\rho_{f1t} + \hat{b}_{ft}\rho_{f1t}^2 + \hat{c}_{ft}\rho_{f1t}^3 \tag{30}$$

$$\ldots \quad \ldots \tag{31}$$

$$\widehat{\omega}_{fJt} = \omega_{ft} + \hat{a}_{ft}\rho_{fJt} + \hat{b}_{ft}\rho_{fJt}^2 + \hat{c}_{ft}\rho_{fJt}^3 \tag{32}$$

$$\sum_{j=1}^{J} \exp\left(\rho_{fjt}\right) = 1, \qquad \exp(\rho_{fjt}) \leq 1 \forall j \tag{33}$$

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.[40]

We now have all the ingredients to calculate markups and the implied marginal costs for the multi-product firms according to equation (6):

$$\hat{\mu}_{fjt} = \hat{\theta}_{fjt}^M \frac{P_{fjt}Q_{fjt}}{\exp\left(\hat{\rho}_{fjt}\right)P_{ft}^M V_{ft}^M} \tag{34}$$

The product-specific output elasticity for materials $\hat{\theta}_{fjt}^M$ is a function of the production function coefficients, the product-specific output (which is data) and the materials allocated to product j. Hence, it can be easily computed once the allocation of inputs across products has been recovered.[41] Marginal costs for the products made by multi-product firms are then recovered by dividing prices by the markup according to equation (7).

5 Results

5.1 Output Elasticities

In this subsection, we present the output elasticities recovered from the estimation procedure outlined in Section 4. We also describe how failing to correct for input price variation or account for the selection bias affects the parameters.

The output elasticities for the single-product firms used in the estimation are reported in Table 3.[42] A 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

---

[40]We find that the input allocations across products are highly correlated with, but not identical to, allocating inputs according to product revenue shares. We experiment with various starting values for the unknowns and find that conditional on converging to an inside solution (e.g., all the product's input shares are between 0 and 1, non-inclusive), the solution is unique. Out of the total 13,394 multi-product firm-year pairs, we hit a corner solution in 894 cases. We exclude these observations from from the analysis in Section 5.

[41]The expression for the materials output elasticity for product j at time t is: $\widehat{\theta}_{fjt}^M = \hat{\beta}_m + 2\hat{\beta}_{mm}\left[\hat{\rho}_{fjt} + m_{ft}\right] + \hat{\beta}_{lm}\left[\hat{\rho}_{fjt} + l_{ft}\right] + \hat{\beta}_{mk}\left[\hat{\rho}_{fjt} + k_{ft}\right] + \hat{\beta}_{lmk}\left[\hat{\rho}_{fjt} + l_{ft}\right]\left[\hat{\rho}_{fjt} + k_{ft}\right]$. For single product firms, $exp(\rho_{fjt}) = 1$ and the output elasticity becomes the one reported in Footnote 36.

[42]The output elasticity for materials is reported in Footnote 36. The output elasticities for labor and capital are $\widehat{\theta}_{ft}^L = \hat{\beta}_l + 2\hat{\beta}_{ll}l_{ft} + \hat{\beta}_{lm}m_{ft} + \hat{\beta}_{lk}k_{ft} + \hat{\beta}_{lmk}m_{ft}k_{ft}$ and $\widehat{\theta}_{ft}^K = \hat{\beta}_k + 2\hat{\beta}_{kk}k_{ft} + \hat{\beta}_{lk}l_{ft} + \hat{\beta}_{mk}m_{ft} + \hat{\beta}_{lmk}l_{ft}k_{ft}$.

the elasticities across sectors. The last column in the table reports the returns to scale. The main message of the table is that there is a distribution around the average and returns to scale exceed one in multiple industries. In each sector, approximately 25 percent of the firms produce under increasing returns to scale. These findings on increasing returns to scale are consistent with Klette and Griliches (1996) and De Loecker (2011) who showed that unobserved firm-level price variation can lead to a downward bias of the returns to scale.

The left panel of Table 4 re-runs the estimation without implementing the correction for the unobserved input price variation discussed in Section 4.3.2. It is clear that the uncorrected procedure yields nonsensical estimates of the production function. For example, the labor output elasticities are often negative and range from -3.01 to 3.17 across sectors, while the returns to scale are either very low or extremely high. As we discussed earlier, these results are not surprising given that we estimate a quantity-based production function using deflated input expenditures. It is clear that failing to account for input price variation yields distorted estimates. For example, a negative labor elasticity stems from the fact that there are firms that produce similar quantities of output using more expensive labor, resulting in a downward biased labor elasticity. In general, it is difficult to sign this bias because there are three inputs which interact in complicated ways in the translog, but it is clear that one needs to correct for input price variation across firms using the procedure described above.

The right panel of Table 4 presents the output elasticities from estimation of the production function on a balanced panel of single-product firms. This estimation does not incorporate the sample selection correction described in Section 4.3.3. As discussed earlier, we expect the output elasticity for capital, and to a lesser extent for labor, to change if we use a balanced panel. In several instances, the output elasticities of capital and labor change. The output elasticity of materials does not change as drastically, which is consistent with our premise that it is the correlation between the productivity threshold and capital and labor that may generate a selection bias.

5.2   Prices, Markups and Marginal Costs Patterns

Our data and methodology deliver measures of firm performance–prices, markups, marginal costs, and productivity, which are usually not simultaneously considered in the existing literature. This section describes the relationship between prices, markups, marginal costs, and productivity.

The markup estimates are reported in Table 5. The mean and median markups are 2.13 and 1.10, respectively, but there is considerable variation across sectors.

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. We would therefore expect to observe an inverse relationship between a product's marginal cost and quantity produced.[43] In Appendix Figure A.1 we plot marginal costs against production quantities (we de-mean each variable by product-year fixed effects in order to facilitate comparisons across firms). The figure shows

---

[43] It is important to note here that we do not derive marginal cost based on the duality of production and cost functions. Instead, we bring in additional information on prices and derive marginal cost based on its definition as the ratio of price to markup. This implies that even though we find returns to scale in the production function estimation, it does not follow automatically that our estimated marginal costs will be declining in quantity.

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 next examine the relationships between the performance measures. Table 6 reports the raw pair-wise correlation matrix across prices, markups, marginal costs and firm-level productivity. As we expect, prices are positively correlated with marginal costs and markups, but negatively correlated with firm productivity. Marginal costs are negatively correlated with markups and productivity, and there is a positive correlation between markups and productivity.

Figure A.2 in the Appendix presents an alternative way to see the inter-relatedness of the firm performance measures by plotting firms' productivity against the marginal costs and markups of their products. In order to compare these variables across firms, we regress the firm-level productivity measures on the firm's main industry-year pair fixed effects and recover the residuals. This allows us to compare productivity across firms within industries. We compare the markups and marginal costs across products by de-meaning each by product-year fixed effects. The figure plots the de-meaned markups and marginal costs against the de-meaned productivities, and removes outliers above and below the 97th and 3rd percentiles. Again, the variables are correlated in ways that one would expect: productivity is positively correlated with markups and strongly negatively correlated with marginal costs.

The previous tables and figures demonstrate correlation patterns that arise across firms. 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 the multi-product firm extension of Melitz and Ottaviano (2008) developed by Mayer et al. (2011). A key assumption in that model 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". A firm's product scope is determined by the point at which the marginal revenue of a product is equal to the marginal cost of manufacturing that product. This structure on the production technology is also shared by the multi-product firm models of Eckel and Neary (2010) and Arkolakis and Muendler (2010). In all three models, more productive firms manufacture more products, a pattern clearly confirmed by Figure A.3.

Mayer et al. (2011) also assume a linear demand system which implies that firms have non-constant markups across products. Further, firms have their highest markups on their "core" products with markups declining as they move away from their main product. Figures A.4 and A.5 in the Appendix provide evidence supporting both implications. In these two figures, 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. Figure A.5 plots the de-meaned markups and marginal costs against the sales share of the product within each firm, and A.4 plots the same variables against the product's rank. We observe that the 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 Mayer et al. (2011).

Foreshadowing the results in the next subsection, we also find evidence of imperfect pass-through of costs on prices because of variable markups. In column 1 of Table 7, we regress prices on firm-product fixed effects and marginal costs and find a pass-through coefficient of 0.29. Since marginal costs are estimated, they are measured with error. We instrument marginal costs with firm-level productivity and input tariffs in column 2. These are valid instruments since they affect prices *only* through their effect on marginal costs. That is, neither productivity nor input prices have a *direct* effect on prices; they only affect prices through their effects on marginal costs. The coefficients in the first stage have the expected signs: productivity (input tariffs) is negatively (positively) correlated with costs and the first stage F-statistic is large. The pass-through coefficient increases to 0.60 in the IV specification, which is in line with long-run pass-through estimates in the exchange rate literature. This imperfect pass-through means that any 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.

5.3 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 prices in 1989 and 1997 in Figure 1. 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 and plotting the residuals. As before, we remove outliers in the bottom and top 3rd percentiles.[44] 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 is virtually *unchanged* between 1989 and 1997. This is 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 does not control for macroeconomic factors that could influence prices beyond the trade reforms. We use the entire sample and control for macroeconomic trends in the following specification:[45]

$$p_{fjt} = \delta_{fj} + \delta_{st} + \delta_1 \tau_{it}^{\text{output}} + \eta_{fjt}. \tag{35}$$

We exploit variation in prices and output tariffs within a firm-product over time through the firm-product fixed effects (δ_{fj}) and control for macroeconomic fluctuations through sector-year fixed effects δ_{st}. Since the trade policy measure varies at the industry level, we cluster our standard errors at this level. We report the price regression with just year fixed effects in column 1 of Table 8. The coefficient on the output tariff

---

[44]We trim to ensure that the results are not driven by outliers. Our results are robust to alternative trims (e..g, the top and bottom 1st or 2nd percentiles) and to not trimming at all. These results are available upon request.

[45]One could try to capture the net impact of tariff reforms using the effective rate of protection measure proposed by Corden (1966). However, this measure is derived in a setting with perfect competition and an 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.

is positive implying that a 10 percentage point decline is associated with a very small–0.94 percent–decline in prices.[46] Between 1989 and 1997, output tariffs fall on average by 62 percentage points; this results in a precisely estimated average price decline of 5.9 percent (=62*0.094). This is a small effect of the trade reform on prices and it is consistent with the raw distributions plotted in Figure 1. 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 8.0 (=62*.128) 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 (35), but separately include input and output tariffs:

$$p_{fjt} = \delta_{fj} + \delta_{st} + \delta_1 \tau_{it}^{\text{output}} + \delta_2 \tau_{it}^{\text{input}} + \eta_{fjt}. \tag{36}$$

We refer the reader to the weighted regression in column 4 of Table 9 (the reason for weighing the regressions is discussed below). 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 1.36 percent decline in prices. On the other hand, the coefficient on input tariffs is small and very noisy. In a counterfactual analysis that holds input tariffs fixed and reduces output tariffs, we would observe a fairly precisely estimated decline in prices. However, the overall changes of prices taking into account both types of tariff declines is less precise (although still statistically significant at the 10 percent level). 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 16.8 percent.

We use the estimates of markup 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 2. Like Figure 1, this figure considers only firm-product pairs that appear in both 1989 and 1997 and de-means the observations by their time average. The figure demonstrates that between 1989 and 1997, the marginal cost distribution shifted left indicating an efficiency gain. However, this marginal cost decline is almost exactly 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 is virtually no change on prices.

We re-run specification (36) using marginal costs and markups as the dependent variables. Since marginal costs and markups are estimated variables, we weigh the regressions by the number of observations in each

---

[46]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.

sector's production function estimation reported in Table 3. Some sectors recover the production function using more observations than others (for example, Table 3 reports that the production function estimation for the apparel sector has 301 observations compared with 1,681 observations for the food products sector). We are more confident in our estimates of the output elasticities, and consequently the estimates of markups and marginal costs, for sectors where we have more data. This weighting procedure should reduce noise. Since prices decompose exactly to the sum of marginal costs and markups, the coefficients in columns 5 and 6 sum to their respective coefficients in column 4. We first focus on marginal costs regressions reported in column 5. The coefficient on output tariffs is positive, suggesting that marginal costs decline with output tariff liberalization but the decline is small and statistically insignificant. However, the coefficient on input tariffs is both positive and large in magnitude, indicating that improved access to cheaper and more variety of imported inputs results in large cost declines; this coefficient is significant at the 11% level. 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 40.3 percent.

This magnitude of the marginal costs decline is fairly sizable and would translate to larger prices declines if markups are constant. However, Figure 2 suggests that markups rose during this period, and in column 6 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 suggest 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 23.4 percent. This increase offsets about half of the average decline in marginal costs, and as a result, the overall effect of the trade reform on prices is moderated. We obtain a qualitatively similar pattern of results in the unweighted regressions in columns 1-3, but not surprisingly, the coefficients are slightly less precisely estimated.

Although tempting, it is misleading to draw conclusions about the pro-competitive effects (or lack thereof) of the trade reform using the markup regressions in column 6 of Table 9. The reason is that output tariff liberalization can affect marginal costs through changes in X-inefficiencies and scale. As we have shown, firms raise markups when costs fall and this imperfect pass-through attenuates the coefficient on output tariffs in the markup regressions. In order to isolate pro-competitive effects, we need to control for any simultaneous shocks to marginal costs. We do this by re-running the markup regression but 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 weighted results in the right panel of Table 10 (the unweighted results are reported in the left panel).[47] Indeed, the coefficient on output tariffs in column 4 is positive and significant; this provides direct evidence that output tariff liberalization exerted pro-competitive effects on markups. The way to interpret the results in column 4 is to consider the markups on 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 1.30 percent relative decline in markups. Column 5 adds input

---

[47]To control for marginal costs as flexibly as possible, we use a third-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. These results are available upon request.

tariffs to 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. In contrast to column 5 of Table 9, the coefficient on input tariff falls basically to zero. We observe a similar pattern of results with the unweighted results in the left panel of Table 10. Our analysis demonstrate that India's trade reform led to large cost reductions leading firms to respond by raising markups. But once we control for these cost effects, output tariff reductions do exert pro-competitive effects by putting downward pressure on markups as we would expect.

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 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 columns 3 and 6 of Table 10. We remind the reader that the specification includes sector-year fixed effects to allow for changes in market conditions over time. The table shows a very strong effect of output tariffs on these high markup products. Using the numbers in the weighted specification in column 6, a 10 percentage point decline in output tariffs leads to a 1.18 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 a 4.27 percent decline a 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.

In sum, 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 suggests that the trade reforms benefited producers relatively more than consumers, at least in the short run. However, this does not imply that the reform lowered consumer welfare, especially in the long run. 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 benefits to consumers are potentially substantially larger. In Table 11, we report results that relate a firm's product additions to changes in its average markup across its products. In column 1, we regress an indicator if a firm adds a product in period t on the change in (log) average markups between $t-1$ and t, while controlling for firm and year fixed effects. In column 2, we use the change in the (log) number of products as the dependent variable. In both cases, increases in markups are strongly correlated with new product introductions.[48] This suggests that firms used the input tariff reductions and associated profit increases to finance the development of

---

[48]These findings are consistent with Peters (2012) who develops a model with imperfect competition that generates heterogeneous markups which determine innovation incentives. We note however that we did not obtain these results by committing to a particular structure (demand, market structure and competition) - our findings could also be consistent with alternative mechanisms.

new products, implying potential long-term gains to consumers. A complete analysis of this mechanism and the impact on welfare lies beyond the scope of this current paper.

6   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 to estimate 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 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 differences between these two measures can inform us about distortions and the magnitude of misallocation within an economy. Importantly, our quantity-based productivity measure is purged of substantial variation in markups across firms which potentially improves upon our understanding of the impact of misallocation on 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*.

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*.

Atkin, D. (2010). Trade, tastes and nutrition in India. *Economic Growth Center Working Paper 986*.

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., 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.

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. (2010). A note on detecting learning by exporting. *NBER Working Paper 16548*.

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. and F. Warzynski (2012). Markups and firm-level export status. *American Economic Review*. forthcoming.

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., 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.

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.

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.

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. Ottaviano (2011). Market size, competition, and the product mix of exporters. *NBER Working Paper 16959*.

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.

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.

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.

Tables and Figures

Table 1: Summary Statistics

Sector	Share of Sample Output (1)	All Firms (2)	Single-Product Firms (3)	Products (4)
15 Food products and beverages	8%	288	137	134
17 Textiles	7%	323	196	78
18 Wearing apparel	0%	22	20	14
21 Paper and paper products	2%	75	59	30
23 Coke, refined petroleum products	24%	26	11	22
24 Chemicals	19%	436	216	479
25 Rubber and Plastic	4%	144	102	82
26 Non-metallic mineral products	5%	112	88	59
27 Basic Metal	11%	213	143	99
28 Fabricated metal products	1%	73	55	45
29 Machinery and equipment	5%	150	84	182
31 Electrical machinery and apparatus	3%	81	55	97
32 Radio, TV and communication	2%	46	39	86
34 Motor vehicles, trailers	7%	57	40	94
Total	100%	2,046	1,245	1,501

Notes: Table reports summary statistics for the sample. The first column reports the share of output by sector in 1995. Columns 2 and 3 report the number of firms and number of single-product firms manufacturing products in the sector in 1995. Column 4 reports the number of products over the full sample, 1989-2003.

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

The products for NIC 2710 are bracketed as Products (*j*). The products for NIC 2731 are bracketed as 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: Output Elasticities of Translog Production Function, by Sector

Sector	Observations in Production Function Estimation (1)	Labor (2)	Materials (3)	Capital (4)	Returns to Scale (5)
15 Food products and beverages	1,681	0.29 [0.02]	0.60 [0.03]	0.08 [0.03]	0.98 [0.04]
17 Textiles	2,054	0.31 [0.01]	0.57 [0.02]	0.11 [0.03]	0.98 [0.04]
18 Wearing apparel	301	0.39 [0.09]	0.62 [0.09]	0.09 [0.07]	1.10 [0.20]
21 Paper and paper products	761	0.29 [0.02]	0.59 [0.02]	0.10 [0.02]	0.97 [0.05]
23 Coke, refined petroleum products	172	0.30 [0.00]	0.60 [0.00]	0.10 [0.00]	1.00 [0.00]
24 Chemicals	2,367	0.31 [0.04]	0.57 [0.05]	0.14 [0.04]	1.02 [0.12]
25 Rubber and Plastic	1,072	0.29 [0.12]	0.66 [0.08]	0.11 [0.07]	1.06 [0.22]
26 Non-metallic mineral products	978	0.42 [0.09]	0.60 [0.05]	0.05 [0.08]	1.07 [0.13]
27 Basic Metal	1,392	0.12 [0.19]	0.51 [0.17]	0.02 [0.20]	0.66 [0.52]
28 Fabricated metal products	637	0.29 [0.03]	0.57 [0.06]	0.10 [0.01]	0.97 [0.09]
29 Machinery and equipment	964	0.25 [0.08]	0.61 [0.07]	0.07 [0.07]	0.92 [0.17]
31 Electrical machinery and apparatus	610	0.24 [0.08]	0.54 [0.08]	0.10 [0.09]	0.89 [0.24]
32 Radio, TV and communication	437	0.24 [0.09]	0.62 [0.12]	0.05 [0.08]	0.92 [0.23]
34 Motor vehicles, trailers	598	0.30 [0.06]	0.87 [0.31]	0.01 [0.06]	1.18 [0.34]

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 (single- and multiproduct firms). Standard deviations of the output elasticities reported in brackets are below the mean values. The 5th column reports the average returns to scale, which is the sum of the mean values from the preceding three columns.

Table 4: Output Elasticities, Input Price Variation and Sample Selection

Sector	Estimates without Correcting for Input Price Variation			Estimates on a Balanced Panel of Single-Product Firms		
	Labor (1)	Materials (2)	Capital (3)	Labor (4)	Materials (5)	Capital (6)
15 Food products and beverages	1.22	0.50	0.36	0.36	0.61	0.10
17 Textiles	3.17	-0.24	-2.50	0.32	0.52	0.05
18 Wearing apparel	0.74	0.14	0.33	0.27	0.63	0.12
21 Paper and paper products	0.00	1.30	1.19	0.31	0.57	0.10
23 Coke, refined petroleum products	0.53	0.10	-0.04	-0.01	0.49	0.45
24 Chemicals	0.00	1.36	0.42	0.29	0.66	0.11
25 Rubber and Plastic	-1.12	1.19	1.88	0.30	0.58	0.11
26 Non-metallic mineral products	-2.53	1.96	0.28	0.38	0.59	0.13
27 Basic Metal	-1.04	0.73	1.15	0.30	0.62	0.08
28 Fabricated metal products	2.01	0.44	-0.02	0.33	0.59	0.08
29 Machinery and equipment	0.37	-0.02	0.39	0.26	0.60	0.11
31 Electrical machinery and apparatus	-0.71	-0.03	-0.70	0.30	0.60	0.10
32 Radio, TV and communication	-3.01	0.54	2.20	0.29	0.60	0.02
34 Motor vehicles, trailers	0.43	0.47	0.25	0.27	0.62	0.12

Notes: The table reports the output elasticities from production function estimations that do not account for input price variation or sample selection. The first panel reports the elasticities from the estimation that does not account for input price variation. The second panel reports the output elasticities from the estimation that uses only a balanced set of single-product firms from 1989-2003. To save space, the standard deviations of the elasticities are not reported.

Table 5: Markups, by Sector

Sector	Markups	
	Mean	Median
15 Food products and beverages	2.02	0.97
17 Textiles	1.45	1.11
18 Wearing apparel	1.15	1.10
21 Paper and paper products	1.05	0.98
23 Coke, refined petroleum products	4.17	0.93
24 Chemicals	2.11	1.00
25 Rubber and Plastic	1.96	1.28
26 Non-metallic mineral products	4.40	2.05
27 Basic Metal	1.86	0.91
28 Fabricated metal products	3.35	1.07
29 Machinery and equipment	2.34	1.27
31 Electrical machinery	2.42	1.04
32 Radio, TV and communication	1.78	1.31
34 Motor vehicles	2.16	1.81
Average	2.13	1.10

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 6: Pairwise Correlation Matrix between Prices, Markups, Marginal Costs and Productivity

	Prices	Markups	Marginal Costs	Productivity
Prices	1.00			
Markups	0.16	1.00		
Marginal Costs	0.94	-0.19	1.00	
Productivity	-0.39	0.04	-0.40	1.00

Notes: Table reports the pairwise correlation matrix across four performance variables: prices, markups, marginal costs and productivity. All variables are expressed in logs. Prices, markups and marginal costs vary at the firm-product level while productivity varies at the firm level. 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)
Log Marginal Cost$_{fjt}$	0.293***	0.603***
	0.037	0.052
Observations	20,631	20,631
Firm-Product FEs	yes	yes

Notes: The dependent variable is (log) price. The regressions exclude outliers in the top and bottom 3rd percent of the markup distribution. All regressions include firm-product fixed effects. Column 1 is a OLS regression on log marginal costs. Column 2 instruments marginal costs with firm-level productivity and input tariffs. The coefficients on productivity and input tariffs in the first stage are -0.70 and 0.52, respectively, and are both statistically significant at the 1-percent level. These are the expected signs and the joint F-statistic is 137. The regressions are run from using data from 1989-1997 and the standard errors are clustered at the industry level. Significance: * 10 percent, ** 5 percent, *** 1 percent.

Table 8: Prices and Output Tariffs, Annual Regressions

	Log Prices$_{fjt}$	Log Prices$_{fjt}$
	(1)	(2)
Output Tariff$_{it}$	0.094*	0.128***
	0.056	0.043
R-squared	0.00	0.02
Observations	20,631	20,631
Firm-Product FEs	yes	yes
Year FEs	yes	no
Sector-Year FEs	no	yes
Overall Impact of Trade	-5.9*	-8.0***
Liberalization	3.5	2.7

Notes: The dependent variable is a firm-product's (log) price. The left column includes year fixed effects and the right column includes sector-year fixed effects; all specifications have firm-product 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 are run 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 and Tariffs, Annual Regressions

	Log Prices_{fjt}	Log Marginal Cost_{fjt}	Log Markup_{fjt}	Log Prices_{fjt}	Log Marginal Cost_{fjt}	Log Markup_{fjt}
	(1)	(2)	(3)	(4)	(5)	(6)
Output Tariff_{it}	0.118**	0.050	0.068	0.136**	0.085	0.051
	0.048	0.099	0.073	0.053	0.115	0.082
Input Tariff_{it}	0.313	1.054	-0.741	0.346	1.444	-1.098**
	0.408	0.751	0.493	0.449	0.888	0.549
R-squared	0.02	0.03	0.03	0.02	0.03	0.02
Observations	20,631	20,631	20,631	20,631	20,631	20,631
Firm-Product FEs	yes	yes	yes	yes	yes	yes
Sector-Year FEs	yes	yes	yes	yes	yes	yes
Weights (see footnote)	no	no	no	yes	yes	yes
Overall Impact of Trade Liberalization	-15.0*	-28.7	13.7	-16.8*	-40.3*	23.4
	8.5	19.4	13.9	9.1	22.1	15.2

Notes: The dependent variable is noted in the columns. Within each panel, the sum of the coefficients from the markup and marginal costs regression equals their respective coefficient in the price regression. The left panel reports the unweighted regressions. The right panel uses the number of observations from the production function estimation (see column 1 of Table 3) as weights for each sector. 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. 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}. Significance: * 10 percent, ** 5 percent, *** 1 percent.

Table 10: Pro-Competitive Effects of Output Tariffs

|	Log Markup_{fjt}			Log Markup_{fjt}		
	(1)	(2)	(3)	(4)	(5)	(6)
Output Tariff_{it}	0.119***	0.118***	0.109***	0.130***	0.129***	0.118***
	0.029	0.034	0.031	0.029	0.037	0.032
Input Tariff_{it}		0.034			0.023	
		0.273			0.296	
Output Tariff_{it} x Top_{fp}			0.317***			0.309***
			0.080			0.072
R-squared	0.594	0.594	0.597	0.622	0.622	0.624
Observations	20,631	20,631	20,631	20,631	20,631	20,631
Marginal cost controls	yes	yes	yes	yes	yes	yes
Firm-Product FEs	yes	yes	yes	yes	yes	yes
Sector-Year FEs	yes	yes	yes	yes	yes	yes
Weights (see footnote)	no	no	no	yes	yes	yes

Notes: The dependent variable is (log) markup. The left panel reports the unweighted regressions. The right panel uses the number of observations from the production function estimation (see column 1 of Table 3) as weights for each sector. Columns 3 and 6 interact output tariffs and the third-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. All regressions include firm-product fixed effects, sector-year fixed effects and a third-order polynomial of marginal costs (these coefficients are suppressed and available upon request). The regressions are run from 1989-1997 and standard errors are clustered at the industry level. Significance: * 10 percent, ** 5 percent, *** 1 percent.

Table 11: Markups and Product Scope

|	Add Dummy	Change in Log Products
	(1)	(2)
Change in Log Average Markup$_{ft}$	0.055***	0.045***
	0.006	0.004
Firm FEs	yes	yes
Year FEs	yes	yes
R-squared	0.30	0.23
Obserations	9,353	9,353

Notes: The right-hand-side variable is the change between t-1 and t of a firm's average (unweighted) log markup. The average is constructed by first dropping outlier markups below and above the 3rd and 97th percentile, respectively, and then computing the (unweighted) average markup across products for each firm and then taking first differences. The dependent variable in column 1 is an indicator if the firm adds a product between t-1 and t. The dependent variable in column 2 is the change in the firm's log number of products. All regressions include firm and year fixed effects. Significance: * 10 percent, ** 5 percent, *** 1 percent.

Figure 1: Distribution of Prices in 1989 and 1997
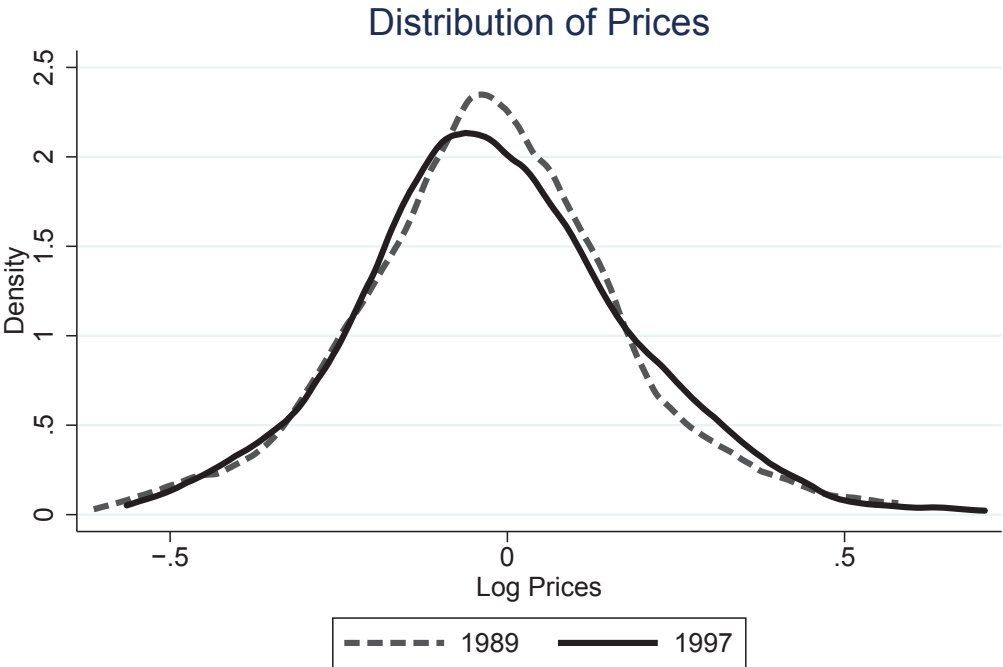
Distribution of Prices

Sample only includes firm−product pairs present in 1989 and 1997.
Observations are demeaned by their time average, and outliers above and below the 3rd and 97th percentiles are trimmed.

Figure 2: Distribution of Markups and Marginal Costs in 1989 and 1997

Sample only includes firm-product pairs present in 1989 and 1997.
Observations are demeaned by their time average, and outliers above and below the 3rd and 97th percentiles are trimmed.
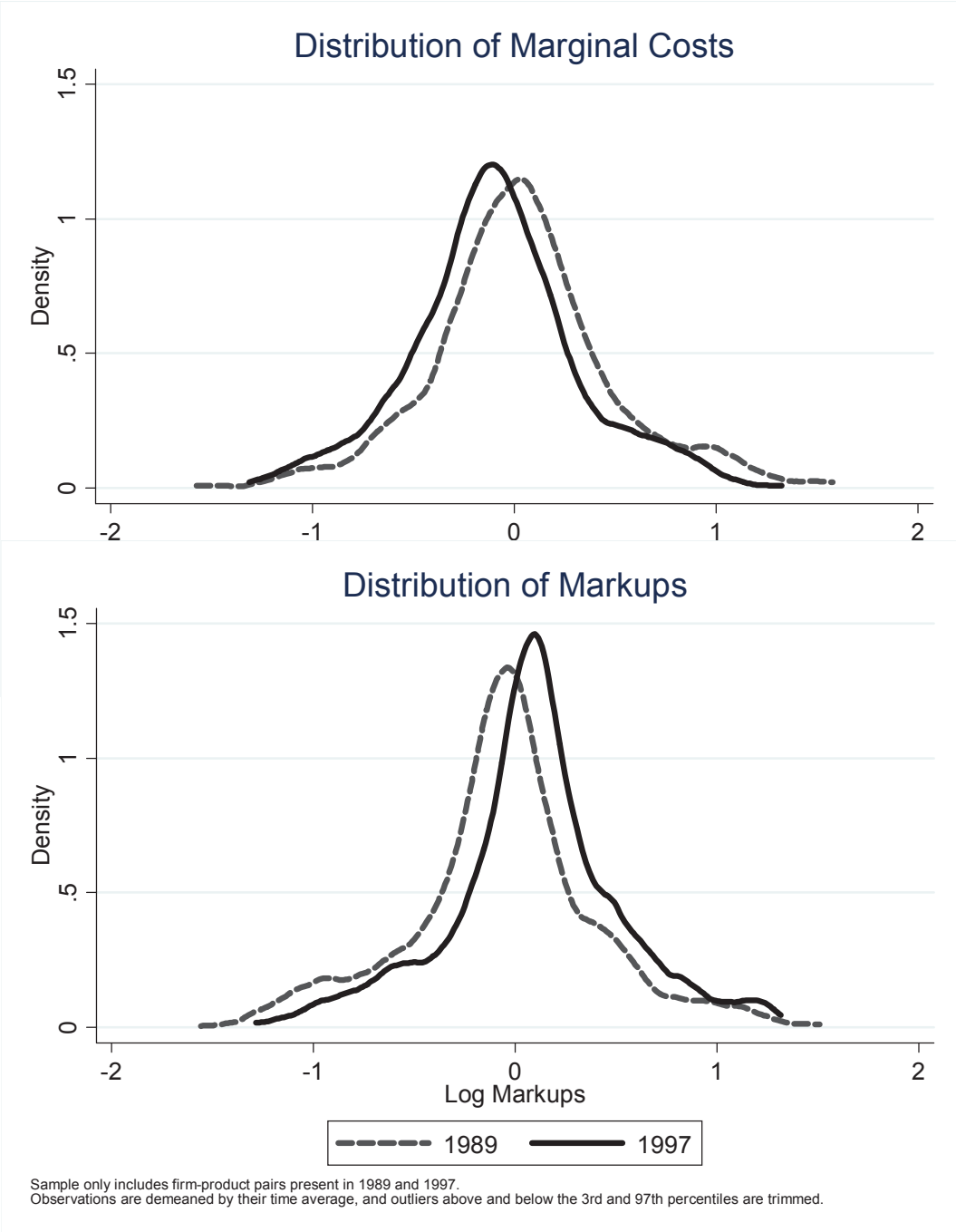
Appendix

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, and hence the B_{ft} term in equation (18), along the lines discussed in Section 4.3.2. 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

In order to proceed, we must specify the production function for quality. Let v_j indicate quality of product j and ψ_i indicate the quality of input i used to produce product j.[49] The production function for output quality is given by:

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

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

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

   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).

   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,

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

   The equation above says that in order to use higher quality inputs, a firm needs to pay higher input prices. There are many ways to justify this relationship. For example, if input markets are competitive but

---

[49]Here, the subscript j denotes both firm and product.

have vertical differentiation, firms must pay higher prices for higher quality inputs. So while high quality inputs are expensive, all firms pay the same input prices *conditional* on input quality.

A.2 Demand

We close the model by specifying the demand and firms' behavioral assumptions.

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

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

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 all 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 product characteristics.

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:

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

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.[50] With the $\boldsymbol{\delta}'s$ in hand, quality is 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, market share and a set of product dummies:

$$v_j = v(p_j, ms_j, I) \tag{A.5}$$

The product dummies are used in lieu of product characteristics (which are not available in our data) and

---

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

can accommodate more general demand specifications such as the nested logit and random coefficients model.

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 γ and productivity ω_j: $mc_j(q_j, v_j, \gamma, \omega_j)$.

The profit function for a firm producing product j is:

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

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 ψ 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:

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

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

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

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

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

or

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

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:

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

47

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

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

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:

$$\frac{1}{\alpha} \cdot \kappa_i \cdot \frac{\partial v_j}{\partial \psi_i} \cdot \frac{1}{\psi_i} - \frac{1}{\alpha} \cdot \kappa_i \cdot v_j \cdot \frac{1}{(\psi_i)^2} - \frac{\partial^2 mc_j}{\partial \psi_i^2} \quad < \quad 0 \tag{A.14}$$

$$\frac{1}{\alpha} \cdot \kappa_i^2 \cdot \frac{v_j}{(\psi_i)^2} - \frac{1}{\alpha} \cdot \kappa_i \cdot \frac{v_j}{(\psi_i)^2} - \frac{\partial^2 mc_j}{\partial \psi_i^2} \quad < \quad 0$$

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

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

and by virtue of the second order condition,

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

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):

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

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, $W_i = g(v_j)$. Moreover, given that output quality is a function of output price, market share and product dummies, we have: $W_i = w(p_j, ms_j, I)$.

Appendix References

1. Berry, Steven (1994). "Estimating Discrete-Choice Models of Product Differentiation", *The Rand Journal of Economics 25(2)*, 242-262.
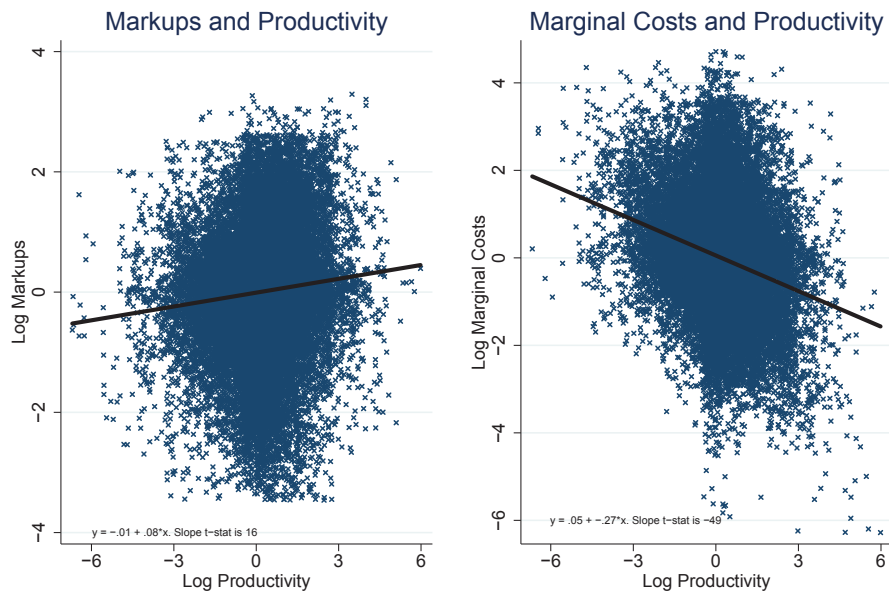
Appendix Figures
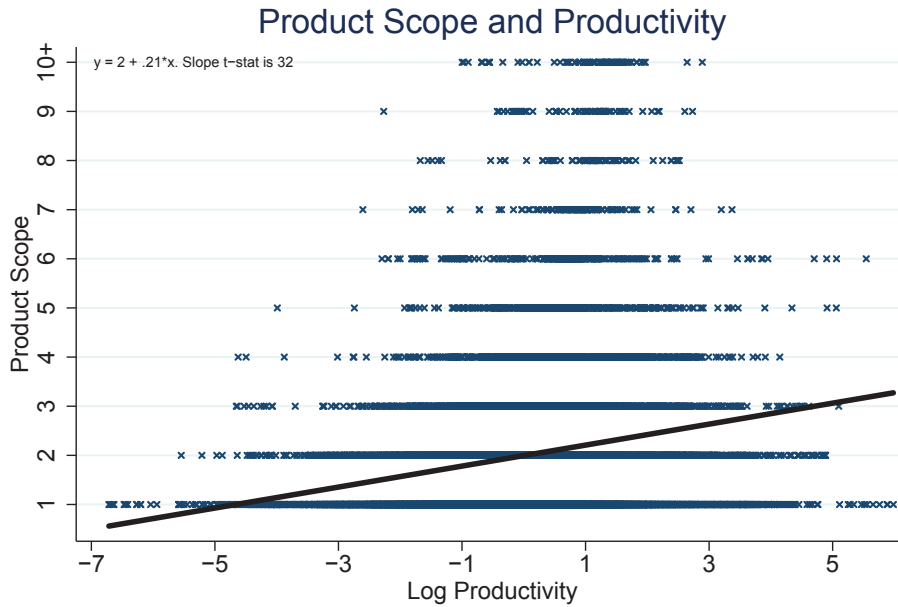
Figure A.1: Marginal Costs and Quantities
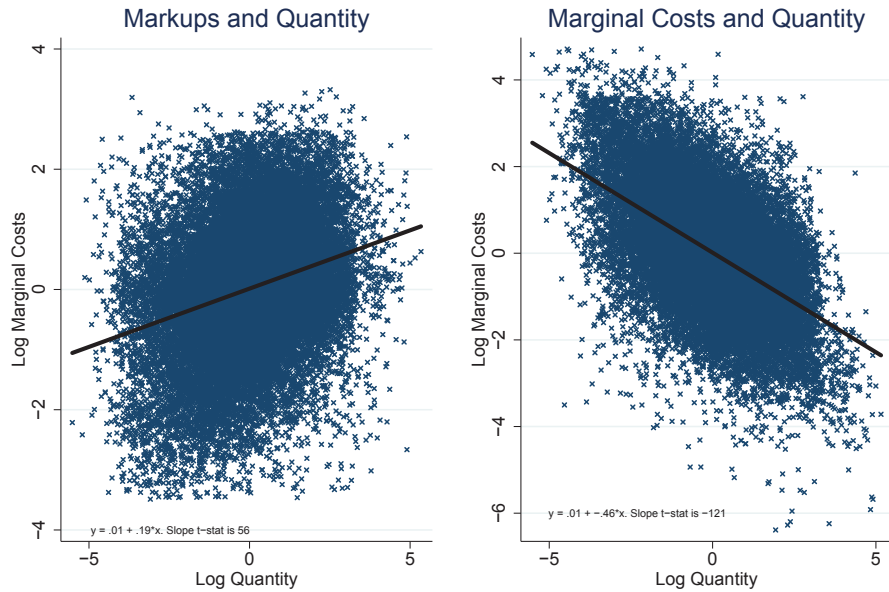
Variables demeaned by product−year FEs.
Markups, cost and quantity outliers are trimmed below and above 3rd and 97th percentiles.
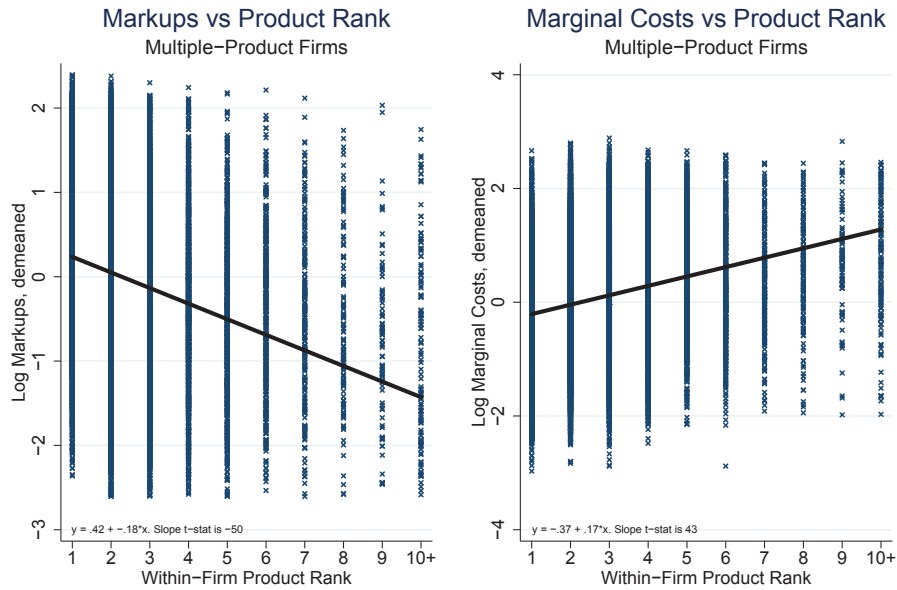
Figure A.2: Marginal Costs and Productivity

Markup and Marginal costs are demeaned by product−year FEs. Firm productivity is demeaned by the firm's main industry−year FE.
For each variable, outliers are trimmed below and above 3rd and 97th percentiles.

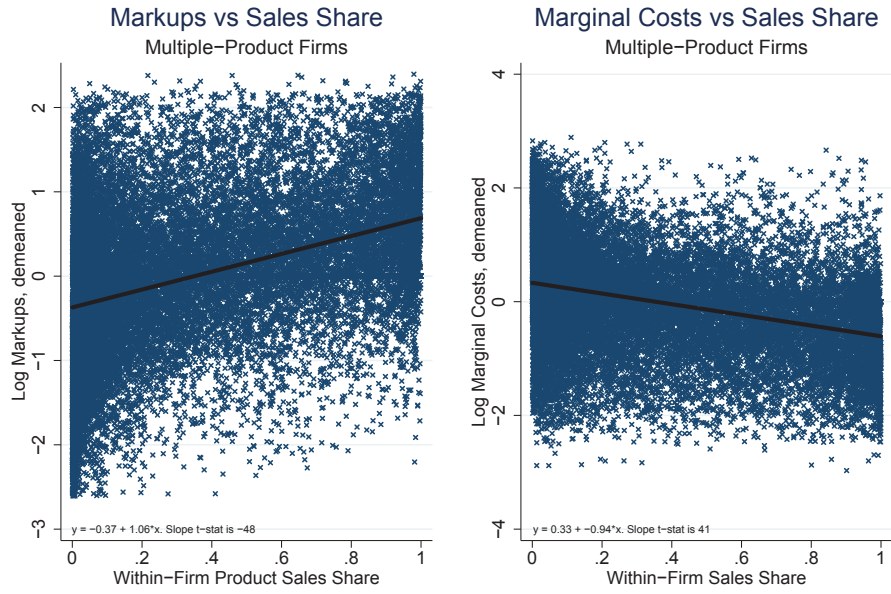Figure A.3: Product Scope and Productivity

Figure A.4: Markups, Costs and Product Rank

Figure A.5: 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.