ITI
7/28/08
11:00 am

The Quality-Complementarity Hypothesis: Theory and New Evidence from Colombia*

Maurice Kugler
Eric Verhoogen

First version: Jan. 30, 2008
This version: July 20, 2008

Abstract

This paper presents a tractable formalization and an empirical investigation of the *quality-complementarity hypothesis*, the hypothesis that input quality and plant productivity are complementary in generating output quality. We embed this complementarity in a general-equilibrium trade model with heterogeneous, monopolistically competitive firms, extending Melitz (2003), and show that it generates distinctive implications for two simple, observable within-sector correlations — between output prices and plant size and between input prices and plant size — and for how those correlations vary across sectors. Using uniquely rich and representative data on the unit values of outputs and inputs of Colombian manufacturing plants, we then document three facts: (1) output prices are positively correlated with plant size within industries on average; (2) input prices are positively correlated with plant size within industries on average; and (3) both correlations are more positive in industries with more scope for quality differentiation, as measured by the advertising and R&D intensity of U.S. industries. The correlations between export status and input and output prices are similar to those for plant size. These facts are consistent with the predictions of our model and difficult to reconcile with alternative models. We interpret the results as broadly supportive of the quality-complementarity hypothesis.

---

* We would like to thank, without implicating, Abhijit Banerjee, Oeindrila Dube, Esther Duflo, Jonathan Eaton, Marcela Eslava, Garth Frazer, Andrew Foster, Penny Goldberg, John Haltiwanger, James Harrigan, Kate Ho, David Hummels, Amit Khandelwal, Wojciech Kopczuk, Adriana Kugler, David Lee, Nathan Nunn, Rohini Pande, Cristian Pop-Eleches, Esteban Rossi-Hansberg, Rafael Santos, Chad Syverson, Kensuke Teshima, Jim Tybout, Miguel Urquiola, David Weinstein and many seminar participants for helpful conversations; Kensuke Teshima, Hitoshi Shigeoka and Juan Ovalle for excellent research assistance; and Juan Francisco Martínez, Luis Miguel Suárez, German Pérez and Beatriz Ferreira of DANE for help with the plant data. Verhoogen thanks the Program for Economic Research at Columbia University for funding and the Center for Health and Well-Being at Princeton University and the Abdul Latif Jameel Poverty Action Lab at MIT for hospitality while the paper was being written. An earlier version of this paper was circulated under the title, “Product Quality at the Plant Level: Plant Size, Exports, Output Prices and Input Prices in Colombia.” Kugler: Wilfrid Laurier University and the Center for International Development, Harvard University, email: maurice_kugler@harvard.edu. Verhoogen (corresponding author): Columbia University, BREAD, NBER, CEPR, and IZA, email: eric.verhoogen@columbia.edu. Address: 420 W. 118th St. Room 1022, New York NY 10027.

1 Introduction

The increasing availability of microdata on manufacturing plants has revealed extensive heterogeneity across plants, even within narrowly defined sectors. A number of stylized facts have been established, among them that exporting plants are larger than non-exporters, that they have higher measured total factor productivity, and that they pay higher wages, on average within industries.[1] Melitz (2003) provides a general-equilibrium trade model that allows for heterogeneity in productivity across firms,[2] is consistent with the first two facts, and accounts elegantly for re-allocations of production within industries in response to trade liberalization. The framework has rightly become the workhorse model for analyzing the behavior of heterogeneous firms under monopolistic competition.

The elegance of the Melitz model comes at the cost of a number of simplifications, however. The treatment of inputs in particular is highly stylized: the lone input, labor, is assumed to be homogeneous. The model thus has relatively little to say about the input choices of firms and cannot account for the third stylized fact above, the greater average wage among exporters. In addition, although the model can be interpreted in terms of quality-differentiated "outputs," as we discuss below, the version of the model that has become standard assumes symmetric outputs. At the same time, plant-level datasets typically lack product-level information, particularly on prices, and even the few datasets that contain information on output prices have little information on the prices of material inputs. As a result, it has been difficult to evaluate the importance of quality differences at the product level, and much of the literature has carried along the assumptions of homogeneous inputs and symmetric outputs with little discussion.

This paper places product-level heterogeneity at the center of the analysis, and investigates a particular hypothesis linking heterogeneity in productivity, heterogeneity in input quality, and heterogeneity in output quality, which we refer to as the *quality-complementarity hypothesis.* The hypothesis posits that input quality and plant productivity are complements in the production of output quality. The theoretical part of the paper embeds this complementarity in a general-equilibrium trade model with heterogeneous firms, extending the Melitz (2003) framework, and shows that the resulting model carries distinctive implications for two simple, observable within-sector correlations — between output prices and plant size and between input prices and plant

---

[1] For a review of this literature, see Tybout (2003).

[2] In the Colombian data, we do not know which plants belong to which firms, and must conduct the analysis at the level of plants. Hereafter we will treat plants as single-establishment firms, and use the terms plant and firm interchangeably.

size — and for how those correlations vary across sectors. An attractive feature of our model is that it remains tractable in a general-equilibrium context while explicitly considering firms' choices of input and output quality.

The empirical part of the paper draws on uniquely rich and representative product-level data from Colombian manufacturing plants to test the cross-sectional predictions of the model. The data, from yearly plant censuses over the period 1982-2005, contain detailed information on the unit values and physical quantities of both inputs and outputs. To our knowledge, these data represent the most complete source of product-level information in a nationally representative plant dataset in any country. Using these data, we document three facts. First, within narrowly defined sectors, on average, output prices are positively correlated with plant size. Second, within narrowly defined sectors, on average, input prices are positively correlated with plant size. Third, using a measure of the scope for quality differentiation from the work of Sutton (1991, 1998), the advertising and R&D intensity of industries in the U.S., both of the above correlations are more positive in industries with more scope for quality differentiation.[3] The correlations between export status and input and output prices are similar to those for plant size. We also present evidence that market power of either final-good producers or their input suppliers is unlikely to be a full explanation for the observed patterns. These empirical findings together are consistent with the predictions of our model and are difficult to reconcile with models that impose symmetry or homogeneity of either set of goods. We interpret the results as broadly supportive of the quality-complementarity hypothesis.

An important caveat is that we do not observe product quality directly but must make inferences about quality from prices and quantities.[4] In this sense, we follow Hummels and Klenow (2005), who use information on prices and volumes of bilateral trade flows — in particular, the result that richer countries export greater volume at higher prices within 6-digit trade categories — to infer that richer countries are likely to be producing higher-quality goods within categories. Our finding that output prices and plant size are positively correlated within narrow sectors is the plant-level analogue of that Hummels-Klenow result. As we discuss below, however, a number of plausible alternative models can account for this correlation without appealing to quality

---

[3] We express the main results in terms of correlations between prices and plant size, rather than prices and physical quantities at the product level, because the possibility of spurious relationships due to measurement error in physical quantities (discussed in Section 6) would require cumbersome qualifications to the statement of results. However, we show in Section 9.4 below that we find similar results when using physical quantities, appropriately instrumented.

[4] Although detailed product attributes can sometimes be observed directly within narrow sectors (see e.g. Goldberg and Verboven (2001)), such information is not available in nationally representative datasets.

differences and are *observationally equivalent* to quality models in trade-flow data or plant-level data containing only output prices and quantities. An advantage of our data is that we observe prices of *inputs* as well as outputs, and can use the relationship between input prices and plant size, as well as variation in that relationship across sectors, to distinguish between the competing models.

A second caveat is that our model uses functional forms that, although standard in the trade literature, have a number of special properties. While these functional forms are particularly useful in deriving closed-form solutions in general equilibrium, the complementarity between input quality and plant productivity in generating output quality would be likely to yield similar cross-sectional predictions under a variety of different demand and production specifications. It is also worth emphasizing that our empirical strategy does not rely on the particular functional forms used in the model, and the resulting estimates are likely to be similarly robust.

The quality-complementarity hypothesis carries a number of potentially important implications. Notably, the hypothesis suggests new channels through which international integration may affect industrial evolution in developing countries. If export markets carry higher quality requirements than domestic markets, then as producers increase exports and upgrade product quality we would expect them to increase quality demands on input suppliers, which may in turn induce those suppliers to upgrade various aspects of their production processes. Alternatively, a reduction of tariffs on imported inputs is likely to reduce the price of high-quality inputs, which may in turn lead firms to upgrade the quality of final goods they produce. Both of these mechanisms are likely to affect the distribution of gains from international integration, even within narrow industries, and hence the extent of political support for trade liberalization. Another potentially important implication of the quality-complementarity hypothesis is that it provides a partial explanation for a familiar and well-established finding in labor economics, the positive correlation between establishment size and wages, known as the "employer size-wage effect" (Brown and Medoff, 1989; Oi and Idson, 1999). To our knowledge, our paper is the first to present evidence from broadly representative data that this correlation also holds for material inputs. The fact that the pattern holds for material inputs as well as labor lends weight to the argument that the size-wage correlation at least in part reflects differences in labor quality, and not solely institutions that are specific to the labor market. In addition, together with the observation that higher-productivity plants are more likely to select into export markets, the quality-complementarity hypothesis offers a natural explanation for the fact that exporters pay higher wages than non-exporters, the third stylized fact above.

The paper is organized as follows. The next section discusses related literature. Section 3 presents a brief look at the data from two detailed sectors, to illustrate the main ideas of the paper. Section 4 develops our model, embedding the quality-complementarity hypothesis in a general-equilibrium setting with heterogeneous, monopolistically competitive firms. Section 5 describes the dataset. Section 6 describes our econometric strategy. Section 7 presents the main results, examining all sectors together (Section 7.1) and how the price-plant size correlations vary across sectors (Section 7.2). In Section 8, we consider alternative explanations for the empirical patterns and undertake additional analyses that suggest that these hypotheses are not the full explanation for the empirical patterns we observe. Section 9 presents a number of additional robustness checks, and Section 10 concludes.

2 Related Literature

This paper is related to a number of existing literatures, in addition to the work mentioned above. Complementarities among inputs in production have been emphasized by Sattinger (1979), Milgrom and Roberts (1990), Kremer (1993), Grossman and Maggi (2000), and Jones (2007), among others, although these papers do not focus on complementarities between inputs and plant productivity draws in the sense of Melitz (2003), as we do here. In explicitly considering the quality choices of firms, and in using R&D and advertising intensity as a measure of the scope for quality differentiation to examine variation across sectors, we follow the lead of Shaked and Sutton (Shaked and Sutton, 1982, 1987; Sutton, 1991, 1998, 2007), although we focus on the variable costs rather than the fixed costs of producing high quality.[5]

Verhoogen (2008) hypothesizes a complementarity between plant productivity and labor quality in producing output quality, but in the context of a partial-equilibrium model in which a number of key relationships — notably, the wage-labor quality schedule and the extent of consumer demand — are taken to be exogenous. The theoretical advance of the current paper, beyond

---

[5] We focus on variable costs — input prices — in large part because they are observable in the Colombian data, unlike fixed costs of raising quality. In the long run, it would clearly be desirable to combine an initial stage of sunk investment in raising quality along the lines of Shaked and Sutton (1982, 1987) and Sutton (1991, 1998, 2007) with the tractability of the Melitz (2003) framework in analyzing the behavior of large numbers of heterogeneous firms. (Costantini and Melitz (2007) take a step in this direction by modeling an endogenous investment choice, but in the context of a model that does not explicitly incorporate quality differences.) It is also worth noting that the our constant-elasticity-of-substitution (CES) demand framework with no fixed costs of raising quality is poorly suited to analyzing market concentration, because it is difficult to reconcile with the fact that in many industries the number of market players remains fixed even as the market grows large. (See Sutton (1991, pp. 70-71).)

generalizing the hypothesis to material inputs, is to set it in an tractable, general-equilibrium framework.[6] Empirically, the important advantage of the current paper is that we observe output and input prices and can therefore conduct more stringent and direct tests of the complementarity hypothesis.

Several studies use product price information from the U.S. Census of Manufactures for a limited number of relatively homogeneous sectors for which unit values can be calculated on a consistent basis. Roberts and Supina (1996, 2000), Syverson (2007), Hortaçsu and Syverson (2007), and Foster, Haltiwanger, and Syverson (2008) all report *negative* correlations between plant size and output prices.[7] A subset of the studies mentioned above use information on unit values of material inputs, which are available on a consistent basis for an even more limited number of inputs (Dunne and Roberts, 1992; Roberts and Supina, 1996, 2000; Syverson, 2007; Hortaçsu and Syverson, 2007); none of these papers explicitly reports cross-sectional correlations of material input prices with plant size. The only study we are aware of that explicitly considers the correlation between non-labor input prices and plant size is Davis, Grim, Haltiwanger, and Streitwieser (2006), which focuses on electricity prices and shows that prices paid by manufacturing plants are *decreasing* in purchase volume.[8] An important advantage of our study over this previous work is that we have access to consistently defined output and input prices for a much broader set of sectors — indeed, for the universe of manufacturing plants with 10 or more workers in Colombia. Below we show that our results are consistent with the U.S. findings for the most homogeneous sectors, but also that the most homogeneous sectors are not representative of the manufacturing sector as a whole.

In independent work, Hallak and Sivadasan (2008) find positive output price-plant size and output price-export status correlations in Indian data, similar to the our first finding mentioned above.[9] An advantage of the Colombian data is that they contain information on the unit values of material inputs; as we mentioned above and discuss in more detail in Section 8 below, the

---

[6] Verhoogen (2008) uses a non-homothetic, multinomial-logit demand specification in order to make transparent the dependence of consumer willingness to pay for quality on income. While the CES demand specification in this paper makes no such link, it has the advantages of tractability and greater comparability to the existing trade literature.

[7] Aw, Batra, and Roberts (2001) use two cross-sections of plant-level data on output unit values from the Taiwanese electronics sector to investigate plant-level price differences between goods sold on the export and domestic market, but do not present evidence on cross-sectional price-plant size correlations.

[8] In an interesting case study of a tractor manufacturer in Pakistan, Andrabi, Ghatak, and Khwaja (2006) find that suppliers with greater relationship-specific assets receive both smaller orders and lower prices. The authors argue that this is because such suppliers are also supplying lower-quality inputs. Given the nature of their data, however, the authors are not able to compare input prices paid by different final-good producers.

[9] Hallak and Sivadasan (2008) also develop a theoretical framework with two dimensions of heterogeneity and minimum quality requirements for entering the export market; see also footnote 19.

information on inputs is crucial for distinguishing the implications of quality models from a number of competing explanations.[10]

In addition to the Hummels and Klenow (2005) paper discussed above, a number of recent papers have used information on unit values in trade-flow data to draw inferences about product quality. Schott (2004) documents that imports into the U.S. from richer countries have higher unit values than imports from poorer ones, within narrow product categories. Other notable contributions in this vein include Hallak (2006), Hallak and Schott (2005), Choi, Hummels, and Xiang (2006), and Khandelwal (2007). A general issue with this set of studies is that it is not clear whether the patterns in trade-flow data reflect quality variation across individual *firms* or variation across sub-sectors, for instance at the (unobserved) 12-digit level. As a consequence, it is difficult to know what implications the studies carry for analysis at the plant level.[11]

Finally, our work is also related to two recent papers relating unit values in trade-flow data to extensions of the Melitz (2003) model: Baldwin and Harrigan (2007) and Johnson (2007). Both papers find that exports to more distant or difficult-to-reach markets have higher unit values on average. This fact is difficult to explain with the standard interpretation of the Melitz (2003) model, which would suggest that more productive firms both charge lower prices and enter more distant markets than less productive firms. The fact is consistent with the hypothesis that more productive firms produce higher-quality goods and charge higher prices, a hypothesis that is explicitly present in the Verhoogen (2004, 2008) model as well as in the variants of the Melitz model that Baldwin and Harrigan (2007) and Johnson (2007) present. The hypothesis is also implicitly present in the Melitz (2003) model itself, given a suitable redefinition of quality units — a redefinition alluded to (albeit not fully developed) in the Melitz's original paper (Melitz, 2003, p. 1699). Appendix A.2 spells out the "quality" version of the Melitz (2003) model, shows how it relates to the model we present in the next section, and shows that it is isomorphic to the Baldwin and Harrigan (2007) model if one abstracts from differences in distance between

---

[10] We are aware of two other independent projects using producer-level output price information. In Mexican data, Iacovone and Javorcik (2008) document that plants raise output prices in preparation for exporting, which suggests that the quality-upgrading process highlighted by Verhoogen (2004, 2008) begins prior to entry into the export market. Crozet, Head, and Mayer (2007) use price information and direct quality ratings on French wines to test the implications of a quality-sorting model of trade. Neither of these projects has access to data on the unit values of material inputs.

[11] Bernard, Jensen, Redding, and Schott (2007), Halpern and Koren (2007) and and Eaton, Eslava, Kugler, and Tybout (2007) have recently developed datasets based on customs declarations for international transactions that include unit values at the plant level. These datasets open up a range of new research possibilities, but they have the disadvantage that they contains unit-value information only for firms that engage in international transactions, and only for the subset of transactions that cross borders. It is not clear to what extent the results for the minority of plants that export or import in each industry can be generalized to the industry as a whole, and no price comparisons can be made between firms that engage in international transactions and firms that do not.

countries.[12] As will become clear below, the key differences between our quality model and the quality Melitz/Baldwin-Harrigan model are the allowance for heterogeneity of inputs and the complementarity between plant productivity and input quality, which generate distinctive implications for input prices. An additional difference is that our framework treats product quality as a choice variable of plants, rather than a deterministic function of plants' productivity draws; this enables us to provide an account of how differences in output quality distributions emerge endogenously across sectors.

3 Illustrative Examples

In this section, we take a brief look at the data for two products, to illustrate the main ideas of the paper. The products have been selected both because they differ in the extent of quality differentiation and because they have relatively simple production processes.

First, consider hollow bricks *(ladrillos huecos)*, a common building material in Colombia, similar to cinderblocks but made primarily out of clay rather than concrete. The scope for quality differentiation in hollow bricks is arguably quite limited. Figure 1.A plots log real output unit value for each plant-year observation against plant size, as measured by log employment (for reasons discussed in Section 6 below), with both variables deviated from year means.[13] Small x's indicate non-exporting plants and small o's indicate exporting plants; consistent with the general pattern, exporting plants in this sector are larger than non-exporters. We observe a *negative* relationship between output price and plant size. Although less obvious in the graph, the relationship between output price and export status is also negative. It is important to note that the price data are noisy and the negative slope is only marginally significant; the large number of observations in the full dataset will be important in drawing statistical inferences with greater confidence below. Figure 1.B plots the log real unit value paid for common clay *(arcilla común)*, the main input into hollow bricks, by producers of hollow bricks. In this case, input prices are negatively correlated with plant size, suggesting that larger plants receive volume discounts or have greater bargaining power with suppliers.

Second, consider men's socks, a product for which there is arguably more scope for quality differentiation. Figure 2.A plots log real output unit value against log employment. In this case we

---

[12] The Johnson (2007) model also carries implications similar to the quality Melitz model if one abstracts from distance.

[13] The data are from the years 1982-1994, since those are the years for which export status is observed; for more details, see Section 5 and the data appendix.

observe a *positive* relationship between output prices and plant size. Although again less obvious, the correlation between output unit value and export status is also positive. Figure 2.B plots the log real price paid by producers of men's socks for raw cotton yarn, the main input into cotton socks. Here we see a strong *positive* relationship between input price and plant size.[14] Figure 2.C plots the log real price paid by producers of men's socks for another common input, cotton thread. Here again we see a strong positive relationship between input price and plant size.

The quality-complementarity hypothesis offers an intuitive explanation for these patterns: the most productive firms use the highest-quality inputs, produce the highest-quality outputs, and grow to be largest, but this mechanism only operates in sectors with sufficient scope for quality differentiation. The following theoretical section provides a general-equilibrium formalization of this intuition, and the empirical section investigates the corresponding patterns for the Colombian manufacturing sector as a whole, showing not only that the price-plant size slopes are greater in industries with more scope for quality differentiation, but also that, on average, the manufacturing sector is more like men's socks than like hollow bricks.

At this point, it is worth clarifying briefly what we mean by product quality. Implicitly, we are treating any product attribute that end-users value and hence are willing to pay for as a component of product quality. In the theoretical section below, we assume that in each country there is a single representative consumer (and a single production function for firms); we are thus sidestepping the issue that different consumers (and firms) may value a given product attribute differently. This may not be as restrictive as it seems, however: Anderson, de Palma, and Thisse (1992) show that the demand patterns generated by a Dixit-Stiglitz representative consumer (like the one in our model) can also be generated by a model with individual consumers making discrete choices on the basis of utility functions with a random consumer-product match term. In this view, the heterogeneous component of consumer tastes can be interpreted as a mean-zero draw, and quality can be interpreted as the subset of product attributes (or components of attributes) that are valued by all consumers.

It is important to note that *we are not equating product quality with the price paid for the product*, nor are we arguing that the quality differences fully explain the price dispersion for inputs and outputs observed, for instance, in Figures 1 and 2. In Section 8 below we consider a number of other plausible models, some involving market-power differences across plants, that may be responsible for part of the price dispersion. We make the more limited argument that

---

[14] Note that men's socks sector also includes producers of non-cotton socks; for that reason the number of plant-year observations in Figure 2.A is smaller than in Figure 2.B.

these alternative models cannot be the complete explanation for the price dispersion, because they are inconsistent with key empirical patterns, and hence that it is difficult to explain the set of cross-sectional correlations we observe without reference to quality differences among inputs and outputs. In this paper, we do not construct quantitative estimates of product quality, because such estimates would necessarily be specific to the particular functional forms of our theoretical model, for which we have little *a priori* justification. Instead, we focus here on what we consider to be the most robust theoretical predictions and empirical patterns to draw the arguably non-model-specific inference that a complementarity between input quality and plant productivity in generating output quality is likely to be playing some (as-yet-unquantified) role.[15]

4 Theory

This section develops a model of heterogeneous inputs, heterogeneous plant productivity, and heterogeneous outputs, extending the Melitz (2003) framework to accommodate a complementarity between input quality and plant productivity in generating output quality. A number of details are relegated to Appendix A.1.

4.1 Set-up

There are two symmetric countries, and in each country there are two sectors, a differentiated sector and a competitive, non-differentiated, constant-returns-to-scale sector producing a numéraire good. Without loss of generality, we focus on one of the two countries. A representative consumer has the following standard asymmetric CES utility function:

$$U = \left\{ \left[\int_{\lambda \in \Lambda} \left(q(\lambda) x(\lambda) \right)^{\frac{\sigma-1}{\sigma}} d\lambda \right]^{\frac{\sigma}{\sigma-1}} \right\}^{\beta} \{z\}^{1-\beta}, \text{ where } 0 \leq \beta \leq 1, \ \sigma > 1 \tag{1}$$

where z is consumption of the non-differentiated good; λ indexes varieties (and will also index plants) in the differentiated sector; Λ represents the set of all differentiated varieties available in the market (produced in either country); σ is a parameter capturing the elasticity of substitution between varieties, where we make the standard assumption that $\sigma > 1$; $q(\lambda)$ is the quality of variety λ, assumed to be observable to all; and $x(\lambda)$ is the quantity consumed. The Cobb-Douglas form implies that the representative consumer spends a constant share of income β on

---

[15] We plan in future work to implement an estimation strategy that relies more heavily on our specific theoretical structure to quantify quality differences across plants.

differentiated varieties and a share $1-\beta$ on the non-differentiated good. Consumer optimization yields the following demand for a particular differentiated variety, λ:

$$x(\lambda) = Xq(\lambda)^{\sigma-1}\left(\frac{p(\lambda)}{P}\right)^{-\sigma} \tag{2}$$

where P is an aggregate quality-adjusted price index and X is a quality-adjusted consumption aggregate of the differentiated varieties available on the market.[16]

We assume that there is just one input, as in Melitz (2003), but we allow for heterogeneity in the quality of this input. We denote quality of the input by c, and assume that there is a mass I of input suppliers, heterogeneous in c, with density $h(c)$ with positive support over $[1,\infty)$.[17] Suppliers are assumed to supply the input inelastically. Let the schedule of input price (which we also refer to as the "wage") for each level of input quality be given by $w(c)$. We solve below for the $w(c)$ schedule that obtains in equilibrium.

There is a pool of potential entrants, who can enter either the non-differentiated sector of the differentiated sector. Entry into the non-differentiated sector is costless. Let $F_N(c)$ be the output produced in the non-differentiated sector by one unit of an input with quality c, and assume that this function takes a particularly simple form: $F_N(c) = c$.

To enter the differentiated sector, entrants must pay an investment cost f_e, paid to input suppliers, in order to receive a productivity draw, λ. To avoid confusion, we refer to the value of this draw as a plant's "capability," borrowing the term from Sutton (2007). We assume that capability is drawn from a Pareto distribution with c.d.f. $G(\lambda) = 1 - \left(\frac{\lambda_m}{\lambda}\right)^k$, with $1 < \lambda_m \leq \lambda$.[18] Each plant that enters the differentiated sector will produce a distinct good; hence λ can be used to index both plants and goods.

Production in the differentiated sector is described by two functions, one describing the production of physical units of output and the other describing the production of quality. We allow the capability parameter to enter both equations: capability may reduce unit input requirements,

---

[16] Specifically,

$$P \equiv \left[\int_{\lambda\in\Lambda}\left(\frac{p(\lambda)}{q(\lambda)}\right)^{1-\sigma} d\lambda\right]^{\frac{1}{1-\sigma}} \tag{3}$$

$$X \equiv \left[\int_{\lambda\in\Lambda}\left(q(\lambda)x(\lambda)\right)^{\frac{\sigma-1}{\sigma}} d\lambda\right]^{\frac{\sigma}{\sigma-1}} \tag{4}$$

[17] The choice of 1 as lower bound is convenient but not crucial.

[18] Below we will impose a lower bound on the shape parameter, k, to ensure that the distribution of capability draws has a finite variance. Helpman, Melitz, and Yeaple (2004) impose a similar restriction.

as productivity does in the standard interpretation of the Melitz model, or it may raise output quality for a given set of inputs.[19] The production of physical units is assumed to be $F_D(n) = n\lambda^a$, where n is the number of units of inputs used and a is a parameter reflecting the extent to which capability lowers unit costs, with $a > 0$. The production of output quality is assumed to be given by a CES combination of capability, λ, and input quality, c.[20] It is convenient to parameterize the contribution of capability as λ^b where $b \geq 0$ and to parameterize the contribution of input quality as c^2, as follows:

$$q(\lambda) = \left[\frac{1}{2}\left(\lambda^b\right)^{\alpha} + \frac{1}{2}\left(c^2\right)^{\alpha}\right]^{\frac{1}{\alpha}} \tag{5}$$

The choices of multiplicative factor $\frac{1}{2}$ and the quadratic form in c are convenient but not crucial.[21] The parameter α reflects the degree of complementarity between capability and input quality; as α becomes more negative, the degree of complementarity increases. We impose the assumption that $\alpha < 0$. This ensures that not only the marginal increase in output quality but also the marginal benefit in profit terms is greater for higher-λ firms. (It also ensures that the second-order conditions for a maximum in the plants' optimization problem are satsified.) Intuitively, returning to the example of men's socks, we assume that the benefit of using higher-quality cotton yarn, both for the quality of the resulting sock and for the profitability of the sock firm, is greater for more capable entrepreneurs. We rule out the possibility that plant capability and input quality are substitutes, for instance because more capable entrepreneurs are particular able to compensate for deficiencies in yarn quality.[22]

The parameter b reflects the technological ease or difficulty in translating higher plant capability into improved product quality. If $b = 0$ then superior capability will not translate into

---

[19] Sutton (2007) uses the term "capability" to refer to a pair of parameters, one reflecting unit input requirements and the other governing quality for a given set of inputs. Here we collapse the two dimensions of plant heterogeneity to one dimension. The Hallak and Sivadasan (2008) model maintains the two dimensions of heterogeneity. (See also Brooks (2006).) In future work, we plan a structural estimation of a model like the current one, and for that purpose it will be important to allow for additional dimensions of heterogeneity. Given the more limited goals of the current paper — deriving simple reduced-form predictions for cross-sectional price-plant size correlations — it is not clear that the increased model flexibility outweighs the cost of added algebraic complication.

[20] Similar functional forms have been used by Sattinger (1979), Grossman and Maggi (2000), and Jones (2007), among others, to model complementarities among inputs.

[21] If the quality production function (5) were instead:

$$q(\lambda) = \left[\mu\left(\lambda^b\right)^{\alpha} + (1-\mu)\left(c^{\gamma}\right)^{\alpha}\right]^{\frac{1}{\alpha}}$$

then the conditions $0 < \mu < 1$ and $\gamma > 1$ would be sufficient.

[22] The assumption that entrepreneurial ability and input quality are substitutes in the production of output quality would lead to the prediction of a negative correlation between plant size and input prices. This is not the empirical pattern we observe in the empirical part of the paper below.

higher quality and outputs will be symmetric across plants. (Think of hollow bricks.) A higher b reflects a greater scope for quality differentiation. (Think of men's socks.) A high b corresponds loosely to what Khandelwal (2007) (borrowing the term from Grossman and Helpman (1991)) calls a long "quality ladder" and more closely to what Sutton (1998) calls a high "escalation parameter" in a sector with a single technological "trajectory" (i.e. one technologically related group of products).[23] To keep the model simple, we have not introduced a parameter capturing the willingness of consumers to pay for product quality, which may also vary across sectors. We would expect such differences across sectors to play a similar role as differences in the technological possibilities for quality upgrading. That is, one could interpret a higher b as indicating either greater technological ease in improving quality or greater willingness of consumers to pay for product quality improvements, or both.

As in Melitz (2003), there is a fixed cost of production, f, and an additional fixed cost of exporting, $f_x > f$, in each period.[24] There is an exogenous probability of exit, δ, in each period. We focus on steady-state equilibria in which the number of entrants equals the number of deaths and the distribution of capabilities in the market does not change. In the interests of simplicity, we assume that there are no variable costs of trade.

Plants are assumed to be price-takers in input markets, and all face the same input price-input quality schedule $w(c)$; given the production function for physical units of output, the cost of producing each additional physical unit is $\frac{w(c)}{\lambda^a}$. Plants maximize over the choices of input quality, c, and output price, p, and which markets to enter. The input quality determines the input price; input quality and λ together determine output quality, which, together with the output price, determines the number of units sold. Note that the symmetry of countries implies that plants will have no incentive to choose different input qualities or output prices depending on which markets they enter. Let $Z = 1$ if the plant enters the export market, and 0 otherwise. Each plant in the differentiated sector faces the following problem:

$$\pi(\lambda) = \max_{p,c,Z} \left\{ \left(p - \frac{w(c)}{\lambda^a} \right) x - f + Z \left[\left(p - \frac{w(c)}{\lambda^a} \right) x - f_x \right] \right\} \tag{6}$$

where $\frac{w(c)}{\lambda^a}$ is marginal cost, x is given by (2), and q (an argument of x) is given by (5).

---

[23] In the case of an industry with a single technological trajectory, Sutton's escalation parameter α varies inversely with the elasticity of required fixed and sunk investments (i.e. R&D and advertising expenditures) with respect to the resulting quality, which he labels β (Sutton, 1998, Ch. 3).

[24] As in Melitz (2003), it does not matter whether we think of f_x as a per-period fixed cost or as the amortized per-period portion of a single, large sunk cost paid when first entering the export market.

To simplify the search for equilibria, we make an additional assumption that guarantees that the availability of input suppliers at each level of quality, c, is sufficient to satisfy the demand for inputs of that quality from the differentiated sector. This will ensure that input of each quality level c are employed in both the differentiated and non-differentiated sectors, and hence are paid the same price in both sectors. Equation (A12) in Appendix A.1 sets out this "non-specialization" condition.

4.2 Equilibrium

Consider first the non-differentiated sector. In this sector, the net benefit of hiring one unit of an input of quality c is $c - w(c)$. Firms optimally choose c such that $w'(c) = 1$. Integrating, this implies $w(c) = c + D$, where D is a constant. In equilibrium, free entry in the sector drives profits to zero, hence it must be the case that $D = 0$ and input price for inputs used in the non-differentiated sector is simply equal to input quality: $w(c) = c$.[25] Note that with this wage-input quality schedule, firms in the non-differentiated sector are indifferent between inputs of different qualities. Under the non-specialization condition discussed above, there are enough inputs of each quality to satisfy demand from the differentiated sector, and hence the input price for any quality c in both sectors is simply c.

In the differentiated sector, each plant in the continuum of plants ignores the effects of its decisions on the aggregates X and P and the first-order conditions for the plant's maximization problem (6) imply the following:

$$c(\lambda) = \lambda^{\frac{b}{2}} \tag{8a}$$

$$q(\lambda) = \lambda^{b} \tag{8b}$$

$$p(\lambda) = \left(\frac{\sigma}{\sigma-1}\right)(\lambda)^{\frac{b}{2}-a} \tag{8c}$$

$$r(\lambda) = \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} XP^{\sigma}(\lambda)^{\eta} \tag{8d}$$

where $p(\lambda)$ is output price, $r(\lambda)$ is revenues, and $\eta \equiv (\sigma-1)\left(\frac{b}{2}+a\right) > 0$.[26] While capability,

---

[25] Total profit in the non-differentiated sector is given by:

$$\Pi_n = \int_{\underline{c}}^{\overline{c}} [c - w(c)]\ h_n(c)dc = D\int_{\underline{c}}^{\overline{c}} h_n(c)dc \tag{7}$$

where $h_n(c)$ is the density of input suppliers used in the sector. The only value of D for which the zero-profit condition can hold is $D = 0$.

[26] The fact that α drops out of these expressions is a consequence of the choices of the multiplicative factor $\frac{1}{2}$ and exponent 2 in (5). In general, if the exponent were γ in place of 2 (see footnote 21) then $c(\lambda)$ and hence $p(\lambda)$ would

λ, input quality, $c(\lambda)$, and output quality, $q(\lambda)$ are unobservable in the Colombian data, the above equations imply relationships between the variables that are observable — input price, $w(\lambda)$ (which equals $c(\lambda)$), output price, $p(\lambda)$, and revenues, $r(\lambda)$. Specifically, (8a)-(8d) imply:

$$\frac{d\ln w}{d\ln r} = \frac{b}{2\eta} \tag{9a}$$

$$\frac{d\ln p}{d\ln r} = \frac{b-2a}{2\eta} \tag{9b}$$

Appendix A.1 solves for the values of the remaining endogenous variables in equilibrium. To summarize briefly, three conditions — a zero-profit condition for remaining in the domestic market, a zero-profit condition for entering the export market, and a free-entry condition that the ex ante expected present discounted value of paying the investment cost to receive a capability draw is zero — pin down the cut-off values for remaining in the domestic market, λ^*, and entering the export market, λ_x^*, as in Melitz (2003). Since $f_x > f$ by assumption, the cut-off for entering the export market is to the right of the cut-off for remaining in the domestic market: $\lambda^* < \lambda_x^*$. The facts that total revenues of plants are equal to total income of input suppliers in the differentiated sector in each country and that the representative consumer dedicates a constant share of expenditures to the differentiated good pin down the scale of the economy.

Several remarks are in order. First, marginal cost is $\frac{c}{\lambda^a} = (\lambda)^{\frac{b}{2}-a}$ and price is a fixed mark-up over marginal cost, as is standard in models with Dixit-Stiglitz (1977) CES demand specifications.

Second, plant size, as measured by revenues, is unambiguously increasing in plant capability.

Third, if there is no scope for quality differentiation — that is, if $b = 0$ — then this model reduces to the Melitz (2003) model.[27] When $b = 0$, there is no complementarity between plant capability and input quality, and all plants choose the same wage. Marginal cost is then unambiguously declining in plant capability, since capability reduces unit input requirements. Because the mark-up is constant, $p(\lambda)$ also declines in λ. In the standard interpretation of the Melitz model, $p(\lambda)$ is taken to represent observed output prices per physical unit. Thus the standard interpretation predicts a *negative* correlation between output price and plant size. Because all plants choose the same wage, the model predicts zero correlation between input prices and plant size. Melitz (2003) points out that his model is consistent with quality differentiation given a suitable choice of quality units. In particular, if we interpret $p(\lambda)$ as reflecting price in quality

---

depend on α.

[27] To reproduce the Melitz (2003) model to the letter, we also need to assume that the size of the non-differentiated sector shrinks to zero and the distribution of input quality collapses to a point. See Appendix A.2.

units rather than physical units, the model can generate a zero or positive correlation between observed output price in physical units and plant size. (Appendix A.2 spells out this argument in detail.) But again, since there is no complementarity between plant capability and input quality, the model predicts zero correlation between input prices and plant size.

Fourth, if there is some scope for quality differentiation — that is, if $b > 0$ — then the complementarity between plant capability and input quality generates positive relationships between plant capability λ and both input price w and output quality q. Given that plant size, as measured by revenues, is unambiguously increasing in λ, the model thus predicts a positive correlation between input price and plant size in this case. Formally, $b > 0$ and (9a) imply $\frac{d \ln w}{d \ln r} > 0$.

Fifth, if $b > 0$, there are two offsetting influences on the relationship between marginal cost and λ and hence, given that the mark-up is fixed, on the relationship between output prices and marginal cost. On one hand, higher-λ plants choose higher quality inputs that carry a higher price. On the other hand, higher-λ plants have lower unit input requirements, which reduces marginal cost. At sufficiently low values of b, the input-requirement-reducing effect will dominate, and output prices will be *declining* in plant capability, λ, and hence in plant size. At sufficiently high values of b, the quality-complementarity effect will dominate, and output prices will be *increasing* in λ and hence in plant size. Formally, from (9b) it follows that the output price-plant size slope is negative for $b < 2a$ and positive for $b > 2a$.

Sixth, both the input price-plant size slope and the output price-plant size slope are predicted to be increasing in the parameter capturing the scope for quality differentiation, b. Formally, from (9a) and (9b) it follows that $\frac{\partial}{\partial b}\left(\frac{d \ln w}{d \ln r}\right) > 0$ and $\frac{\partial}{\partial b}\left(\frac{d \ln p}{d \ln r}\right) > 0$.

Seventh, since the cut-off for entry into the export market is to the right of the cut-off for entry into the domestic market, export status is positively correlated with λ, and we have the same predictions for the correlations of output and input prices with export status as for the correlations with plant size.[28]

Finally, we acknowledge that the prediction that output quality is positively correlated with plant size (as long as $b > 0$) may seem implausible to rich-country consumers of, for exam-

---

[28] Note that the symmetry between countries in this model implies that if plants enter the export market they will sell the same amount in the export market as in the domestic market. Thus the model does not predict a positive correlation of plant size and the export share of sales, conditional on exporting. Nonetheless, below we also use the export share of sales as an indicator of export status, partly for the purposes of comparison with existing results in the literature, and partly because it is not difficult to imagine extensions to our model in which the export share and plant capability would be positively related, for instance if capability reduced per-unit export costs as well as unit input requirements or if plants exported higher-quality goods with higher prices to richer consumers in foreign markets.

ple, French wines or Swiss watches, sectors in which the most expensive goods are produced by small producers. But while it may well be that this model fails to describe quality choices and the extreme high-quality end of many industries, it appears that it does capture an important characteristic of industrial sectors in countries at roughly Colombia's level of development. For instance, Verhoogen (2008) finds that larger plants in Mexico were more likely to have ISO 9000 certification, an international production standard commonly interpreted as a measure of product quality. We will also see below that the positive plant size-quality relationship is consistent with our findings in the Colombian data.

To summarize, our model makes the following testable predictions: (1) a negative correlation between output prices and plant size (or export status) in sectors with the least scope for quality differentiation ($b < 2a$); (2) a positive correlation between output prices and plant size (or export status) in sectors with the most scope for quality differentiation ($b > 2a$); (3) a zero or small positive correlation between input prices and plant size (or export status) in low-b sectors; and (4) a greater positive correlation between input prices and plant size (or export status) in high-b sectors. These predictions stand in contrast to the standard Melitz model, which predicts a negative output price-plant size correlation and a zero input price-plant size correlation, and to the "quality" Melitz model, which is consistent with a positive output price-plant size correlation but again predicts a zero input price-plant size correlation. We now turn to the data to test these predictions. We will return to a discussion of alternative theoretical models in Section 8 below.

5 Data

The data we use are from the *Encuesta Anual Manufacturera* (EAM) [Annual Manufacturing Survey], collected by the *Departamento Administrativo Nacional de Estadística* (DANE), the Colombian national statistical agency. The dataset can be considered a census of manufacturing plants with 10 or more workers.[29] Data including product-level information are available for the 1982-2005 period. Data on exports and imports, as well as employment and earnings of blue-collar and white-collar workers, are available on a consistent basis only for 1982-1994. We construct two separate plant-level unbalanced panels, a 1982-2005 panel and a 1982-1994 panel. We observe approximately 4,500-5,000 plants in each year.

In conjunction with this standard plant survey, DANE also collects information on the value

---

[29] As mentioned above, we do not have access to firm-level data or information on which plants belong to which firms. Details of the survey sampling design are discussed in Appendix B.2.

and physical quantity of each output and input of each plant, which is used to calculate national producer price indices. A unit value for each plant-product-year observation can then be calculated by dividing value (revenues or expenditures) by physical quantity. The unit value represents an average price paid or charged by a plant over a year; hereafter we will (somewhat loosely) use the terms unit value and price interchangeably. The product classification scheme is based on the 4-digit International Standard Industrial Classification (ISIC) revision 2; DANE then adds four Colombia-specific digits. We observe approximately 4,000 distinct product codes in the data. The first important advantage of these data is that they contain values and physical quantities of all inputs and all outputs of all plants in the dataset. The second important advantage is that DANE analysts have been extremely careful about maintaining consistent units of measurement within product categories. DANE dictates to plants which measurement units to use. If plants report using alternative units, then DANE follows up to request that the plant report using the correct units. If the plant insists that it is not possible to report using the units dictated by DANE, DANE creates a new product category for the good using the new units.[30] Thus, for example, there exist two 8-digit product categories corresponding to corrugated cardboard boxes *(cajas de carton acanalado)*: product code 34121010 refers to corrugated cardboard boxes measured in kilograms; product code 34121028 refers to corrugated cardboard boxes measured in number of boxes. Similarly, product code 35123067 refers to weed killers and herbicides measured in kilograms, and product code 35123075 refers to weed killers and herbicides measured in liters. Units of measurement are truly homogeneous within categories. This fact and the fact that inputs are included as well as outputs make the Colombian data unique, arguably better-suited to analyzing our research question than any dataset in any other country.

Although the Colombian data are of high quality relative to Colombia's level of development, the data still contain a fair amount of noise, like plant-level datasets in other developing countries.[31] We undertake an extensive procedure to clean the data and reduce the influence of outliers. An important step in this process is "winsorizing" the data, recommended by Angrist and Krueger (1999); the procedure is to set all values below a lower bound, for instance the 1^{st} percentile, to the value at the lower bound, and all values above an upper bound, for instance the 99^{th} percentile, to the value at the upper bound. In addition, we have recalculated results using a variety of different bounds for the winsorizing procedure as well as a number of different strate-

---

[30] Source: personal communication with Juan Francisco Martínez, Luis Miguel Suárez, and German Pérez of DANE.

[31] For more on shortcomings of plant-level datasets in many developing countries, see Tybout (1992).

gies for dealing with the remaining outliers, and have found the results we report to be robust. See Appendix B.1 for variable definitions and Appendix B.2 for full details on the cleaning and processing of the datasets.[32]

Table 1 presents summary statistics for the plant-level data in our two panels, the 1982-1994 unbalanced panel and the 1982-2005 unbalanced panel. Consistent with patterns for the U.S. documented by Bernard and Jensen (1995, 1999), exporting plants are larger, in terms of both sales and employment, and pay higher wages; also, a minority of plants export and conditional on exporting, plants derive a minority of their sales from the export market. Consistent with patterns for Taiwan (Aw and Batra, 1999) and Mexico (Verhoogen, 2008), exporting plants have a higher white-collar-to-blue-collar wage ratio. Exporting plants produce in a larger number of distinct output categories and purchase from a larger number of distinct input categories than non-exporters.[33]

Table 2 reports summary statistics for the product-level information in the 1982-2005 panel, by ISIC major group (3-digit). Column 1 reports the number of distinct 8-digit products in each 3-digit group. Columns 2 and 5 report the average number of distinct plants selling or purchasing each product in the group in each year, respectively. Columns 3 and 6 report the within-product standard deviation of log real prices for each product as output or input, respectively. Columns 4 and 7 report the within-product-year standard deviation of log real prices for each product as output or input, respectively.[34] We note that there is a fair amount of variation across sectors both in the number of selling or purchasing plants per product and in extent of within-product price dispersion.

---

[32] The plant-level information in the EAM has also been used in, for instance, Roberts and Tybout (1997), Clerides, Lach, and Tybout (1998), and Das, Roberts, and Tybout (2006). The product-level information has been used by Eslava, Haltiwanger, Kugler, and Kugler (2004, 2005, 2006, 2007) in studies that focus on the effects of market reforms on productivity, plant turnover, and factor adjustments, rather than on price-plant size correlations or quality differentiation.

[33] The fact that exporters produce in more distinct output categories than non-exporters is consistent with the prediction of the multi-product-firm theory of Bernard, Redding, and Schott (2006b) and the patterns documented in U.S. data by Bernard, Redding, and Schott (2006a).

[34] The within-product standard deviation is the standard deviation of the residuals from a regression of log real unit value on a full set of product dummies. The within-product-year standard deviation is the standard deviation of the residuals from a regression of log real unit value on a full set of product-year dummies.

6 Econometric Strategy

Our baseline econometric model is the following:

$$\ln p_{ijt} = \alpha_t + X_{jt}\gamma + \theta_{it} + \delta_{rt} + \xi_k + \varepsilon_{ijt} \tag{10}$$

where i, j, and t index goods, plants, and years, respectively; $\ln p_{ijt}$ is the log real unit value of a good; α_t is a year-specific intercept; X_{jt} is a measure of plant size; θ_{it} is a product-year effect; δ_{rt} and ξ_k are region-year and industry effects, respectively;[35] θ_{it} is a product-year effect; and ε_{ijt} is a mean-zero disturbance. We run regressions separately for output prices and input prices.

The coefficient of interest in these regressions is γ, which corresponds to the elasticities $\frac{d\ln w}{d\ln r}$ and $\frac{d\ln p}{d\ln r}$ from (9a) and (9b). It is worth emphasizing that the estimate of γ reflects a *correlation*, not a causal effect of plant size on plant-level average prices. Indeed, our argument is precisely that both plant size and prices are determined by unobserved heterogeneity in plant capability. Nonetheless, the estimate of γ is informative in the sense that it helps us to discriminate between competing models with contrasting predictions for the cross-sectional correlations. In some specifications, we estimate (10) with an indicator for export status or the export share of sales as the X_{jt} variable, in place of plant size. In others, we interact X_{jt} with a sector-level measure to examine how the price-plant size slope varies across sectors.

The product-year effects, θ_{it}, absorb all variation in prices of particular products that is common across plants. The coefficient γ is thus identified on the basis of within-product-year variation; that is, it is based on a comparison of prices between plants of different sizes producing or consuming *the same product* in a given year. In this way we avoid the difficulty that we have no metric with which to compare unit values across products.

A natural measure of plant size is gross output; this is in fact the standard measure of plant size used by the Colombian statistical agency, measured as total sales plus net intra-firm transfers plus net change in inventories. Measurement error in gross output is a potential concern, however. To the extent that the measurement error is classical, it may simply attenuate coefficient estimates toward zero. But non-classical measurement error, generating unpredictable biases, is also a possibility. In addition, revenues represent output quantities times output prices; a pos-

---

[35] Note that it is not redundant to have industry effects in this regression, even though product-year effects are included, because there is not a perfect mapping from product categories to industries. Plants are assigned to industries based on the relative importance of all the products they produce, and two plants producing the same product may belong to two different industries, depending on the plants' product mixes. For details, see Appendix B.2.

itive coefficient may simply reflect the presence of output prices on both sides of the equation, rather than the theoretical relationship we are trying to estimate. To address these concerns, we use employment as an alternative measure of plant size. Employment has the advantage that measurement error is likely to be less severe and, importantly, uncorrelated with reports of values and quantities of outputs and inputs. It also has the advantage that it does not mechanically incorporate output prices. We use log total employment as an instrument for log total output in an instrumental-variables (IV) procedure; under the assumption that the measurement errors in gross output and total employment are uncorrelated, the IV estimator will yield consistent estimates of the theoretical elasticities of interest.[36]

Observations at the plant-product-year level are likely not to be independent either across products within plant-years or across years within plants. For this reason, we cluster errors at the plant level. The number of observations for the purposes of calculating standard errors is thus effectively equal to the number of distinct plants. We report the number of distinct plants (i.e. clusters) as well as the number of plant-product-year observations used in each regression.

7 Results

7.1 Baseline Estimates

Panel A of Table 3 presents estimates of equation (10) with the log real output price index as the dependent variable. Columns 1 and 2 use log total output and log employment, respectively, as the measures of plant size. Column 3 reports the IV estimate of the coefficient on log total output using log employment as the instrument for log total output. The IV estimate for log total output is slightly larger than the OLS estimate, consistent with the observation above that measurement error in gross output is generating attenuation bias. The coefficient on log employment is quite close to the IV estimate for total output, consistent with the hypothesis that employment is measured with less error than gross output. The coefficient estimates are highly statistically significant, and indicate that output prices are *positively* correlated with plant size on average. The price and plant size variables are in log terms and the coefficients on plant size can be interpreted as elasticities. Column 2 suggests, for instance, that 10% greater employment

---

[36] Concerns about measurement error explain why we do not simply regress unit values on physical quantities at the product level. Unit values are calculated by dividing total value produced (or consumed) by quantity. Hence any measurement error in quantities will generate a spurious negative correlation between quantity and unit value. We return to this issue is Section 9.4 below.

is associated with .26% higher output prices.

Panel B of Table 3 presents the analogous regressions with the log real input prices as the dependent variable. In moving from Column 1 to Columns 2 and 3, the coefficient on plant size falls slightly. This suggests a non-classical measurement error bias.[37] But again, the important message of those columns is that input prices are *positively* correlated with plant size on average. The estimates suggest that 10% greater plant size is associated with .11-.12% higher prices paid for material inputs.[38]

Table 4 uses the 1982-1994 panel, in which export status is observed, to estimate correlations between export status and output and input prices. To simplify the presentation of results, hereafter we focus on the reduced-form regressions with log employment as the key co-variate; the IV estimates are similar. Results for output prices are in Panel A, and for input prices are in Panel B. For comparison purposes, Column 1 of each panel presents a regression with log employment as the key co-variate, comparable to Column 2 of Table 3; the results are similar to those for the longer 1982-2005 panel. The results in Column 2 indicate that both output and input prices are higher among exporters. On average, exporters (i.e. plants with non-zero exports) have approximately 11% greater output prices and 3.7% greater input prices than non-exporters. In Column 3, with the export share of revenues is the key co-variate, the coefficient is positive and significant for output prices and positive but not significant for input prices. The share of revenues derived from exports seems generally to have less explanatory power than the indicator for entry into the export market.[39] Caution is warranted in interpreting the results in Columns 4 and 5, since, if one believes our theoretical framework, both employment and export status reflect a single underlying capability parameter, λ, and are likely to be collinear. Subject to that caveat, the results in Columns 4 indicate that being an exporter is associated with both higher output prices and higher input prices, even conditional on plant size.[40] Again, the results for export share in Column 5 are less robust, especially for input prices.

---

[37] One possibility is the following. Suppose that a "producer" re-sells a good produced by a "supplier", reports the money paid to the supplier as input expenditure, reports sales of the good in total revenues, but does not include the number of physical units in quantity of the good produced. Then a regression of input unit value on gross output will yield a positively biased coefficient. Other measurement biases are possible.

[38] Note that the output price-plant size relationship estimated in Panel A is steeper than the input price-plant size relationship in Panel B, suggesting that profitability may be increasing in plant size as well. We will see in Table 9 below, however, that this difference is not robust.

[39] Recall that the theory literally predicts no variation in export share conditional on exporting. The important theoretical point is that export status is correlated with λ and hence plant size, not necessarily that the export share is.

[40] These results are consistent with the results of Hallak and Sivadasan (2008) in Indian data mentioned above. The fact that their theoretical model contains two dimensions of heterogeneity means that it is able to provide a coherent account of the finding of systematically higher prices among exporters conditional on plant size.

As mentioned in the introduction, the one input for which unit values are commonly observed in plant-level datasets is labor. To compare our results for material inputs to results for employee wages, Table 5 presents regressions that are similar to those in Columns 1-3, Panel B, Table 4 but with earnings of all employees, blue-collar employees, white-collar employees, and the log earnings ratio, respectively, as the dependent variables. We see clear evidence that the earnings of both blue-collar and white-collar workers, as well as the relative earnings of white-collar workers, are greater in larger plants and in plants with more exports. The positive wage-plant size relationship is a robust and familiar fact (Brown and Medoff, 1989; Oi and Idson, 1999), and the positive wage-exporting relationship is also consistent with long-established results (Bernard and Jensen, 1995, 1999). The positive relationships between wage inequality and plant size and between wage inequality and exporting in Column 4 are less well known, but are also consistent with findings from Taiwan (Aw and Batra, 1999) and Mexico (Verhoogen, 2008).

7.2 Comparison Across Industries

The results above indicate that output and input prices are positively correlated with plant size when we constrain the slope coefficient to be the same across industries, but this leaves open the possibility that there are significant differences across sectors. As discussed in Section 4 above, our model would lead us to expect negative output price-plant size and zero or low input price-plant size correlations in homogeneous industries yet strongly positive correlations in industries with more scope for quality differentiation.

Our measure of the scope for quality differentiation at the industry level is the ratio of total industry advertising and R&D expenditures to total industry sales for large U.S. firms from the 1975 U.S. Federal Trade Commission (FTC) Line of Business Survey. The Line of Business Program, which was in existence from 1974 to 1977, is unique in that it required firms to break down advertising and R&D expenditures by industry, as opposed to reporting consolidated figures at the firm level. As a consequence, the data are generally perceived to be the most accurate industry-level information on advertising and R&D expenditures, and have been used in a large number of studies, including Cohen and Klepper (1992), Brainard (1997), Sutton (1998), and Antras (2003). In the context of a model with fixed costs of improving quality, Sutton (1998) demonstrates rigorously that there is a mapping between the (unobserved) scope for quality differentiation in an industry and the (observed) extent of fixed investments in raising quality, which we measure here by the advertising and R&D/sales ratio.[41] Although we use a different model in this paper, the same intuition carries through: if incurring greater costs to raise consumer willingness to pay is ineffective, profit-maximizing firms will not incur the costs; if such costs are incurred by profit-maximizing firms, it must be that they are effective. Under the assumption of optimal behavior by firms, we can infer that the scope for raising consumers' willingness to pay — that is, the scope for quality differentiation — is greater in industries where firms invest more in advertising and R&D. The advertising/sales ratio may arguably be more closely tied to consumer willingness to pay than the R&D/sales ratio, so we also run separate regressions with the advertising ratio alone. We converted the information on advertising and R&D expenditures and sales from the FTC industry classification (which approximates the 1972 U.S. Standard Industrial Classification) to the ISIC revision 2 4-digit level using verbal industry descriptions.

A potential concern with using advertising and advertising + R&D intensity is that they may reflect *horizontal* rather than vertical differentiation. Theoretically, one might very well expect sectors with greater horizontal differentiation to have greater price-plant size correlations; in our model, if the scope for quality differentiation is sufficiently large ($b > 2a$), then a greater degree of horizontal differentiation (a lower σ) will give rise to steeper output price-plant size and input price-plant size slopes.[42] Our primary strategy for addressing this concern is to control explicitly for horizontal differentiation using the widely used Rauch (1999) differentiation measure, which is based on whether a good is traded on a commodity exchange or has a quoted price in industry trade publications; as a robustness check, we conduct a similar anlysis using an alternative measure of horizontal differentiation, from Gollop and Monahan (1991) (see Section 9.2 below). Details of the construction of the Rauch (1999) measure and conversion to the ISIC rev. 2 industry categories, which generated some fractional values, are in Appendix B.2. Table 6 reports summary statistics on advertising intensity, advertising + R&D intensity, and the Rauch (1999) index, by ISIC major group. The table also reports summary statistics on Herfindahl indices for producers and suppliers, which will be discussed below in Section 8.

We report the results using the differentiation measures in Table 7. Because of slippage in the concordance process, we are not able to calculate the differentiation measures for several ISIC industries, and number of observations is reduced. For comparison purposes, Columns 1 and 6 report specifications similar to Table 3, Column 2, Panels A and B, for the modified sample;

---

[41] See Theorem 3.3, the remark immediately following, and footnote 12 in Sutton (1998, Ch. 3).

[42] Equations (9a) and (9b) imply that $\frac{\partial}{\partial \sigma}\left(\frac{d \ln w}{d \ln r}\right) < 0$ and $\frac{\partial}{\partial \sigma}\left(\frac{d \ln p}{d \ln r}\right) < 0$ (since $\eta = (\sigma - 1)\left(\frac{b}{2} + a\right)$) and hence that the slopes are increasing in the extent of horizontal differentiation.

the point estimates are not statistically different from those in the earlier table. In Columns 2-3 and 7-8, we include the interaction of log employment with the advertising/sales ratio and the advertising + R&D/sales ratio corresponding to the output industry of each Colombian plant.[43] The results are consistent with the predictions of our model above: the output price-plant size slope and the input price-plant size slope are *significantly more positive* in industries with more scope for quality differentiation.

Columns 4-5 and 9-10 control for differences in horizontal differentiation by including an interaction of plant size with the Rauch (1999) measure.[44] Including the interaction with the Rauch measure has little effect on the coefficient estimates for advertising and advertising + R&D intensity. The estimates in Columns 4-5 and 9-10 are not statistically distinguishable from the estimates in Columns 2-3 and 7-8, respectively. Note that the coefficients on log employment without interactions in Columns 4-5 and 9-10 are estimates of the price-plant size slopes in the *most homogeneous* industries, that is, industries for which both advertising and R&D intensity and the Rauch measure are zero. In Columns 4-5, we see that the estimated output price-plant size slope in the most homogeneous industries is *negative* and statistically significant. In Columns 9-10, we see that the estimated input price-plant size slope in the most homogeneous industries, although positive, is effectively zero. These estimates are consistent with the theoretical predictions of our model. The results for the uninteracted log employment term in Columns 4-5 of Table 7 are also consistent with previous findings of negative price-plant size correlations for homogeneous goods in U.S. data by Roberts and Supina (1996, 2000), Syverson (2007), Hortaçsu and Syverson (2007), and Foster, Haltiwanger, and Syverson (2008) discussed in Section 2 above.

8 Alternative Explanations

This section considers two alternative types of explanations for the price-plant size correlations we observe: models with plant-specific demand shocks and market power in input markets, and models with perfect competition without quality differences.

---

[43] The definition of output industry differs slightly between Columns 2-3 and 7-8. When the output price is the dependent variable, we define output industry to be simply the first four digits from the 8-digit product code for each plant-product-year observation. When the input price is the dependent variable, this is not possible, because particular inputs are not associated with particular outputs. Instead, we use the ISIC 4-digit category of the corresponding plant, which is calculated as the industry in which the plant derives the largest share of its revenues. For details, see Appendix B.2.

[44] Note that any differences in horizontal differentiation that affect all plants equally are already captured by the product-year and industry effects; the key question is whether horizontal differentiation affects the price-plant size slopes, and that is what the interaction term picks up.

8.1 Plant-Specific Demand Shocks and Market Power in Input Sectors

A common approach to inferring product quality from output prices and quantities is to define quality as any factor that shifts the demand curve for a product outward. But there are many factors that may lead to greater demand for the products of a particular plant that do not correspond to increases in the valuation of the products by end-users, and hence not to our notion of product quality. One example might be favors from a well-placed government procurement official. Another might be collusive agreements between particular plants not to compete head-on in particular markets. Another might be plant- or firm-specific import licenses (see Mobarak and Purbasari (2006)). Although in Dixit-Stiglitz-type frameworks such shocks would typically not affect output prices since they would not affect marginal costs, in the context of other demand systems it is quite plausible that such idiosyncratic shocks would lead plants both to raise prices and to increase output. For example, in the framework of Foster, Haltiwanger, and Syverson (2008), which is based on a demand system similar to that of Melitz and Ottaviano (2008) with endogenous mark-ups, plant-specific demand shocks unrelated to quality can have such an effect.[45] Plant-specific demand shocks pose a challenge to the attempt to draw inferences about quality because they may generate a positive output price-plant size correlation even in the absence of heterogeneity in product quality. In other words, under some parameter values, the implications of the plant-specific demand shocks story and the quality story for the output price-plant size correlation are observationally equivalent.[46]

The implications of standard non-quality-related demand shocks stories and our quality model diverge when it comes to input prices, however. In Foster, Haltiwanger, and Syverson (2008), for instance, plants are subject to idiosyncratic input price shocks, but high input prices are unambiguously bad for plants: higher input costs lead to high output prices (at the same quality level) and reduced output. This mechanism generates a *negative* correlation between input prices and plant size. While it is possible that a positive shock to plant-specific demand could coincide

---

[45] In the model of Foster, Haltiwanger, and Syverson (2008), there are two offsetting effects. On one hand, demand shocks tend to induce a positive output price-plant size correlation. On the other, greater productivity tends to induce a negative price-plant size correlation, as in the standard Melitz (2003) model. Which effect will dominate is not clear a priori. Empirically, the authors find a negative, insignificant correlation between price and several measures of total output in the homogeneous sectors they focus on (Foster, Haltiwanger, and Syverson, 2008, Table 1).

[46] In Melitz and Ottaviano (2008), mark-ups are larger in more productive plants, but do not completely offset the lower production costs in such plants, and output price is predicted to be negatively related to plant size. One might imagine a "quality" interpretation of the Melitz and Ottaviano (2008) model, similar to the quality Melitz model discussed above, in which output prices are positively correlated with plant size. Our response in this case is similar to our response to the quality Melitz and Foster, Haltiwanger, and Syverson (2008) cases: such a model would be difficult to reconcile with the input price-plant size correlations we observe.

with a positive shock to plant-specific input prices, in the Foster et al. framework there is no explicit mechanism that would lead this to happen systematically.[47]

An extension of the Foster et al. framework could generate a systematically positive input price-plant size correlation. Consider the possibility that plants have monopsony power in input markets and face upward-sloping supply curves for inputs. In this case, a plant-specific demand shock will generate an increase in derived demand for inputs, which will in turn tend to lead plants to pay a higher input price. This effect could offset the effect of shocks to input prices discussed in the previous paragraph, and generate a positive input price-plant size correlation overall, even in the absence of quality differentiation. We have two responses to this objection. First, this mechanism would not lead us to expect either the output price-plant size correlation or the input price-plant size correlation to increase with the scope for quality differentiation; it is not clear, in other words, how the story would account for the results in Table 7. Second, we note that the mechanism would predict a *zero* input price-plant size correlation in sectors in which purchasers have no monopsony power, where plants presumably face flat, or very nearly flat, input supply curves. We test this implication directly below.

A related alternative explanation is that plants are subject to plant-specific demand shocks but that *suppliers* have market power in input markets. Within industries, plants facing positive demand shocks for their output may face lower elasticities of output demand, which may in turn lead them to be less sensitive to the prices of inputs. If suppliers have market power, they may optimally charge higher prices to these less price-sensitive producers. Halpern and Koren (2007) present a model with this feature, which they call "pricing-to-firm." This mechanism is consistent with both a positive correlation of output prices and plant size and a positive correlation of input prices and plant size. Note again, however, that the mechanism would not lead us to expect either the output price-plant size correlation or the input price-plant size correlation to increase with the scope for quality differentiation, as we saw in Table 7. Note also that the mechanism again predicts a zero input price-plant size correlation in the most competitive input sectors.

To test the predictions of these alternative models, we construct three different measures of market power in input markets. First, we construct a standard Herfindahl index for suppliers

---

[47] A related but distinct hypothesis is that some plants pay higher input prices not because of an idiosyncratic shock for a given input, but because they buy inputs that have gone through more stages of processing, and hence require fewer stages of processing when be transformed into outputs. We have two responses. First, in this story, if outputs differ primarily in how many processing stages are performed by a plant itself vs. by its input suppliers, then it is not clear why output prices would vary systematically with input prices. Second, it is not clear why more-processed inputs would be systematically associated with larger plants.

of each 8-digit input, defined as the sum of squared market shares of producers of the input. Second, we construct a Herfindahl index for *purchasers* of each 8-digit input, defined as the sum of squared shares of expenditures on the input by different plants. Note that it is not possible to create such a Herfindahl purchaser index in standard datasets in which material input purchases are not observed.[48] Third, we take a plant's share of total expenditures on a given input as a measure of monopsony power for the plant.[49] This measure varies within sector, unlike the Herfindahl indices. It is likely to be correlated with plant size but not perfectly so, since input mixes vary across plants. The mean values of the Herfindahl indices by ISIC major group are reported in Columns 4-5 of Table 6.

Table 8 presents results for input prices using various combinations of these market power measures. The number of plants for which all three market power measures can be constructed is smaller than the number used in Table 3 above, because some plants only use non-manufacturing inputs for which the Herfindahl supplier index cannot be constructed. For comparison purposes, Column 1 replicates Column 2 in Table 3, Panel B for the reduced sample. The coefficient on the interaction of the Herfindahl supplier index and log employment (in Columns 2, 5, 6 and 8) is negative and significant.[50] That is, the input price-plant size slope is *less steep* in input sectors in which suppliers have more market power. This result contradicts the pricing-to-firm hypothesis. The coefficient on the plant purchaser share is uniformly positive and significant. Although the coefficient on the interaction of the Herfindahl purchaser index is not robust, the point estimates are positive when the purchaser share is not also included. These two results are consistent with the purchaser monopsony power hypothesis: it does appear that larger purchasers, even controlling for overall plant size, pay higher input prices. But the important point of this table is that the coefficient on uninteracted log employment remains positive and highly significant throughout. That is, even in input sectors that approach zero supplier and purchaser concentration, and even controlling for the expenditure share of purchasers in particular markets, the positive input price-plant size correlation is robust. While producer monopsony power may well be part of the explanation of the positive input price-plant size correlation, it appears that it cannot be the whole story.[51]

---

[48] Unfortunately, we do not observe which plants purchase from which input suppliers. Clearly, such information would allow one to construct more precise measures of market power.

[49] We are grateful to Andrew Foster for suggesting this measure.

[50] Note that the Herfindahl indices without interactions are collinear with the product-year effects and therefore omitted.

[51] An additional piece of evidence on the supplier market power story was reported in Table 5 above. One input in which there is scope for quality differentiation and for which suppliers arguably have little market power is

8.2 Perfect-Competition Models

One might also ask whether the price-plant size correlations we observe could be explained by models of perfect competition without appealing to differences in product quality. Perfect competition can be reconciled with a non-degenerate distribution of plant sizes if plants have heterogeneous costs but marginal costs increase with output, as in the span-of-control model of Lucas (1978). If plants are producing the same good in such models, however, then they expand output until price equals marginal cost and in equilibrium all plants producing the good have the *same* price and costs. Thus to explain the price-plant size correlations in the context of perfect competition, one must assume that plants in the same "sector" are actually producing different goods. But if plants with different prices are assumed to be producing different goods, then it is not clear why there should be a systematic relationship between price and plant size. One could assume that, because of exogenous technological factors, the optimal plant size for producing goods with high prices happens to be larger than the optimal size for producing goods with low prices, but this is assuming what needs to be proved. In short, it seems difficult to explain the price-plant size correlations we observe without reference to differences in product quality.

This is not an argument against perfect competition *per se*. Even under perfect competition, the correlations we observe may obtain if, for instance, producing higher-quality outputs requires both (a) higher-quality inputs and (b) more recent-vintage technologies that have higher fixed costs and that particularly talented entrepreneurs are able to use at lower marginal cost.[52] Such a model, however, would retain two of the key features of our model: the fact that higher-quality inputs are used to produce higher-quality outputs, and the reduced-form complementarity between input quality and entrepreneurial capability (here mediated by technology choices).[53] Our main argument is not for or against a particular market structure or set of functional forms; it is that the empirical patterns we observe are difficult to explain without these two features.

---

unskilled labor. Union density in Colombia is low by Latin American standards; the unionization rate in 2002 was 5.2% overall, and 4.7% in the private sector (Farné, 2004). In Table 5 we saw that there is a strong positive correlation between plant size and the wage of unskilled (as well as skilled) workers. Given the low unionization rate, it does not seem likely that individual, non-union workers have the power to set higher wages at plants they perceive to be facing less elastic demand.

[52] In many cases, models based on monopolistic competition can be shown to be isomorphic to models based on perfect competition, and this case is not an exception. See Atkeson and Kehoe (2005) for a related discussion.

[53] Note that the second of these features is the crucial one. On the basis of a wide variety of models, one would expect that high-quality inputs are used in the production of high-quality outputs, and hence that input prices are positively correlated with output prices. The more distinctive implication of our model, attributable to the hypothesized complementarity between input quality and plant capability, is that both input and output prices are positively correlated with plant capability and hence with plant size.

9 Robustness Checks

This section undertakes four additional robustness checks, the first using an alternative two-step estimation strategy, the second using an alternative measure of horizontal differentiation, a modified version of the Gollop-Monahan index, the third focusing on the subset of non-exporters, and the fourth examining the relationship between prices and physical quantities at the product level.

9.1 Two-Step Method

To explore the robustness of our baseline results, we generate analogous results using a two-step method, estimating plant-average output and input prices in a first stage and then estimating the relationship between the plant-average prices and plant size in a second stage. Econometrically, the model is:

$$\ln p_{ijt} = \alpha_t + \theta_{it} + \mu_{jt} + u_{ijt} \tag{11}$$

$$\widehat{\mu}_{jt}^{OLS} = X_{jt}\gamma + \delta_{rt} + \xi_k + v_{jt} \tag{12}$$

where μ_{jt} is a plant-year effect, and u_{ijt} and v_{jt} are mean-zero disturbances. The first-stage estimates, $\widehat{\mu}_{jt}^{OLS}$, can be interpreted as plant-average prices, controlling for product-year, region-year and industry effects. Note again that these plant averages are identified by differences between the unit values of a given plant and unit values of other plants producing (or consuming) the same products in the same year.[54] All else equal, if the disturbances, u_{ijt} and v_{jt}, are uncorrelated with the co-variates, then the estimate of γ from (12) and the one-step estimate of γ from (10) will converge asymptotically to the same estimate. (See Baker and Fortin (2001, pp. 358-359) for a useful discussion of the relationship between such one-step and two-step estimators.) In our case, the two-step model differs from the one-step model in that the two-step model weights each plant-year observation equally in (12) whereas the one-step method weights each plant-product-year observation equally, effectively placing more weight on plant-years with a greater number of

---

[54] An important technical caveat is that identification of the plant-year and product-year effects in this model is not assured. Intuitively, the issue is that if in a particular year a plant only produces one product, and in that year the product is only produced by that plant, then it is not possible to identify the plant-year effect for that plant separately from the product-year effect for that product. A similar issue arises in the literature using employer-employee data to identify plant and person effects (Abowd, Kramarz, and Margolis, 1999). Generally speaking, the plant-year effects can only be uniquely identified for plants that are in a connected "network" of plants, where a plant is connected if it produces a good that is also produced by another plant in the network. To ensure this, we find the largest such network and drop the plants not in that connected set. This leads us to drop fewer than 5% of plant-year observations in the sample.

plant-product-year observations. Table 9 reports the two-step estimates corresponding to (11)-(12). The estimates for the plant-average output price are smaller than those in Panel A of Table 3, but the preferred estimates in Columns 2-3 are nonetheless positive and significant.[55] The estimates for the plant-average input price are nearly identical to those in Panel B of Table 3. Overall, it is reassuring that the one-step and two-step methods are broadly consistent.[56]

9.2 Gollop-Monahan Index

For robustness, we have also estimated the model of Table 7 using an alternative measure of horizontal differentiation derived from the Gollop and Monahan (1991) index.[57] The Gollop-Monahan index was originally designed to measure diversification across establishments of multi-establishment firms, but it has also been used to measure horizontal differentiation across firms by Syverson (2004b). Following Bernard and Jensen (2007), we use one component of full three-component index in Gollop and Monahan (1991). Details of the construction of this measure appear in Appendix B.1. Intuitively, the measure is using the dissimilarity of *input mixes* across plants within an industry to proxy for the horizontal differentiation of outputs across plants.

Table 10 presents regressions similar to Columns 1, 4, 5, 6, 9 and 10 of Table 7 with the Gollop-Monahan index as the measure of horizontal differentiation. Although the point estimates in Columns 2-3 and 5-6 are smaller than the corresponding estimates in Table 7, and the estimates for advertising + R&D intensity are only marginally significant, the results are largely consistent with those using the Rauch measure. Indeed, we cannot reject the hypothesis that the estimates on the interaction terms are the same using the Gollop-Monahan and Rauch indices. Note also that the estimates for the log employment term without interactions in Columns 2-3 are again negative and significant, consistent with the corresponding estimates in Table 7 as well as with our theoretical predictions.[58]

---

[55] Given the difference in weighting between the one-step and two-step methods described above, it appears that the differences in estimates for outputs between Tables 3 and 9 are due to the fact that the output price-plant size relationship is steeper among larger plants that produce more distinct products.

[56] In unreported results (available from the authors), we undertake an alternative estimation strategy, in which we construct a Törnqvist output price index at the plant level, comparing unit values of each plant to industry-average values, weighting each output by the share of plant revenues. Results are consistent with those reported here.

[57] We are grateful to Chad Syverson for suggesting this alternative measure.

[58] The point estimates for the log employment term without interactions in the input-price regressions in Columns 5-6 are negative but not statistically different from zero, the theoretical prediction.

9.3 Results for Non-Exporters

Another possible objection is that larger plants may tend to have higher output and input prices because they are exporters and because per-unit transport costs lead plants to "ship the good apples out", that is, export higher-quality varieties than they sell in the domestic market (Hummels and Skiba, 2004). In that case, the positive price-plant size correlations would be due to the transport costs, not to the complementarity between input quality and plant productivity in producing output quality.

To address this objection, we re-estimate our baseline model using only data from non-exporting plants. Table 11 reports the results. Comparing to Table 3, we see that the point estimates for output prices are slightly smaller and for input prices are slightly larger than for the entire sample, but the overall message is that the positive price-plant size correlations are robust and highly significant, even among non-exporters. It does not appear that the shipping-the-good-apples-out hypothesis can be the entire explanation for the price-plant size correlations we observe.

9.4 Results Using Physical Quantities at Product Level

As a final robustness check, we examine the relationship between prices and physical quantities at the product level. As mentioned in Section 6 above, care must be exercised in drawing inferences from the relationship between unit values and physical quantities: because unit values are calculated as revenues or expenditures divided by physical quantities, any measurement error in physical quantities will generate a spurious negative correlation with unit values. Column 1 of Table 12 reports regressions of the form of (10), but where log number of physical units is included in place of plant size on the right-hand size. We indeed see that the coefficient on log physical quantity is negative and highly significant both for outputs (Panel A) and for inputs (Panel B). Although we do not have an instrument for physical quantities at the product level, log employment is again available as an instrument, to reduce the influence of measurement error. When we use log employment as an instrument for log physical quantities in Column 3, we find that the estimated coefficient on log physical quantities becomes *positive* and significant. The coefficients are not statistically distinguishable from the estimates using log revenues as the measure of plant size (in Column 3 of Table 3) nor, indeed, from the reduced-form estimates in Column 2 of this table (which are the same as in Column 2 of Table 3). It appears, in other words, that the negative coefficients in Column 1 are due entirely to the mechanical negative bias induced by measurement error; once that bias is eliminated, the estimates using the component of physical quantities that is correlated with plant size are similar to the estimates using the alternative methods above.

10 Conclusion

This paper has used uniquely rich and representative data from Colombia to test the *quality-complementarity hypothesis*, the hypothesis that input quality and plant productivity are complementary in producing output quality. We have provided a tractable, general-equilibrium formalization of the hypothesis, extending the Melitz (2003) model. We have three main findings. First, output prices and plant size or export status are positively correlated within narrow industries on average. Second, input prices and plant size or export status are positively correlated within narrow industries on average. Third, both patterns are stronger in industries that have more scope for quality differentiation as measured by the advertising and R&D intensity of U.S. industries. We have also shown that the input price-plant size correlation cannot be fully explained by market power of either suppliers or purchasers in input markets. These results are consistent with the predictions of our model and difficult to reconcile with models that impose homogeneity or symmetry of either inputs or outputs and that lack the complementarity between input quality and plant capability in producing output quality.

The quality-complementarity hypothesis carries a number of broader implications. In the introduction we briefly discussed two types, the first concerning new channels through which output and input markets may interact in response to trade liberalization, the second concerning the generalization of the "employer size-wage effect" to material inputs. Here we add an additional implication: the quality-complementarity hypothesis points to shortcomings of widely used methods of productivity estimation. A standard approach is to deflate both output revenues and input expenditures by sector-level price indices, and to estimate productivity as the residual in a regression of log deflated output revenues on log deflated input expenditures. Katayama, Lu, and Tybout (2006) have argued that even if the coefficients of this regression can be estimated consistently, the resulting productivity estimates confound (at least) four distinct dimensions of heterogeneity across plants: (1) productive efficiency, (2) mark-ups, (3) output quality, and (4) input prices, which in part reflect input quality. They also note that the mere availability of data on physical units of inputs and outputs is not sufficient to identify productive efficiency separately from the other factors without further homogeneity assumptions. While a number of techniques have recently been developed to separate technical efficiency and mark-ups (see e.g. Melitz (2000), Bernard, Eaton, Jensen, and Kortum (2003)), comparatively little attention has been paid to the quality dimensions, especially to the heterogeneity in input quality.[59] Our results provide empirical reinforcement for the argument that ignoring heterogeneity in output and input quality may yield misleading inferences. We leave the further exploration of the broader implications of the quality-complementarity hypothesis — in particular, the investigation of the effect of import-tariff reductions on input quality and output quality choices, and the development of additional methods of productivity estimation that are consistent with quality-differentiated inputs and outputs — for future work.

References

Abowd, J., F. Kramarz, and D. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67(2), 251–333.

Anderson, S. P., A. de Palma, and J.-F. Thisse (1992): *Discrete-Choice Theory of Product Differentiation*. MIT Press, Cambridge, MA.

Andrabi, T., M. Ghatak, and A. I. Khwaja (2006): "Subcontractors for Tractors: Theory and Evidence on Flexible Specialization, Supplier Selection, and Contracting," *Journal of Development Economics*, 79, 273–302.

Angrist, J. D., and A. B. Krueger (1999): "Empirical Strategies in Labor Economics," in *Handbook of Labor Economics*, ed. by O. C. Ashenfelter, and D. Card, vol. 3A. Elsevier Science.

Antras, P. (2003): "Firms, Contracts, and Trade Structure," *Quarterly Journal of Economics*, 118(4), 1375 – 1418.

Atkeson, A., and P. J. Kehoe (2005): "Modeling and Measuring Organization Capital," *Journal of Political Economy*, 113(5), 1026 – 1053.

Aw, B. Y., and G. Batra (1999): "Wages, Firm Size, and Wage Inequality: How Much Do Exports Matter?," in *Innovation, Industry Evolution, and Employment*, ed. by D. B. Audretsch, and A. R. Thurik, pp. 13–56. Cambridge University Press, Cambridge, New York and Melbourne.

Aw, B. Y., G. Batra, and M. J. Roberts (2001): "Firm Heterogeneity and Export-Domestic Price Differentials: A Study of Taiwanese Electronics Products," *Journal of International Economics*, 54(1), 149–169.

Baker, M., and N. M. Fortin (2001): "Occupational Gender Composition and Wages in Canada, 1987-1988," *Canadian Journal of Economics*, 34, 345–376.

Baldwin, R., and J. Harrigan (2007): "Zeros, Quality and Space: Trade Theory and Trade Evidence," NBER Working Paper No. 13214.

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., and J. B. Jensen (1995): "Exporters, Jobs, and Wages in U.S. Manufacturing: 1976-1987," *Brookings Papers on Economic Activity*, pp. 67–112.

---

[59] Exceptions include de Loecker (2007) and the Katayama, Lu, and Tybout (2006) paper itself, both of which structurally estimate demand systems to help distinguish the contributions of mark-ups, demand shocks (i.e. output quality) and productive efficiency. A valid alternative approach has been to focus on homogeneous industries where quality differentiation is likely to be limited (Foster, Haltiwanger, and Syverson, 2008; Syverson, 2004a).

——— (1999): "Exceptional Exporter Performance: Cause, Effect, or Both?," *Journal of International Economics*, 47, 1–25.

——— (2007): "Firm Structure, Multinationals, and Manufacturing Plant Deaths," *Review of Economics and Statistics*, 89(2), 193 – 204.

Bernard, A. B., J. B. Jensen, S. J. Redding, and P. K. Schott (2007): "Firms in International Trade," *Journal of Economic Perspectives*, 21(3), 105–130.

Bernard, A. B., S. J. Redding, and P. K. Schott (2006a): "Multi-Product Firms and Product Switching," Unpub. paper, Tuck School of Business.

——— (2006b): "Multi-Product Firms and Trade Liberalization," Unpub. paper, Tuck School of Business, Dec.

Brainard, S. L. (1997): "An Empirical Assessment of the Proximity-Concentration Trade-off between Multinational Sales and Trade," *American Economic Review*, 87(4), 520 – 544.

Brooks, E. L. (2006): "Why Don't Firms Export More? Product Quality and Colombian Plants," *Journal of Development Economics*, 80(1), 160–178.

Brown, C., and J. Medoff (1989): "The Employer Size-Wage Effect," *Journal of Political Economy*, 97(5), 1027–1059.

Choi, Y., D. Hummels, and C. Xiang (2006): "Explaining Import Variety and Quality: The Role of Income Distribution," NBER Working Paper No. 12531.

Clerides, S., S. Lach, and J. Tybout (1998): "Is Learning by Exporting Important? Micro-Dynamic Evidence from Colombia, Mexico and Morocco," *Quarterly Journal of Economics*, 113, 903–947.

Cohen, W. M., and S. Klepper (1992): "The Anatomy of Industry R&D Intensity Distributions," *American Economic Review*, 82(4), 773 – 799.

Costantini, J. A., and M. J. Melitz (2007): "The Dynamics of Firm-Level Adjustment to Trade Liberalization," Unpub. paper, INSEAD and Princeton University, June.

Crozet, M., K. Head, and T. Mayer (2007): "Quality Sorting and Trade: Firm-Level Evidence for French Wine," Unpub. paper, Université de Paris I.

DANE (2004): "Metodología de la Encuesta Anual Manufacturera," Technical document, *Departamento Administrativo Nacional de Estadística*, Dec.

Das, S., M. Roberts, and J. Tybout (2006): "Market Entry Costs, Producer Heterogeneity, and Export Dynamics," Forthcoming in *Econometrica*.

Davis, S. J., C. Grim, J. Haltiwanger, and M. Streitwieser (2006): "Electricity Pricing to U.S. Manufacturing Plants, 1963-2000," Unpub. paper, Chicago GSB, May.

de Loecker, J. (2007): "Product Differentiation, Multi-Product Firms, and Estimating the Impact of Trade Liberalization on Productivity," NBER Working Paper No. 13155.

Dixit, A., and J. E. Stiglitz (1977): "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, 67, 297–308.

Dunne, T., and M. Roberts (1992): "Costs, Demand, and Imperfect Competition as Determinants of Plant-Level Output Prices," in *Empirical Studies in Industrial Organization: Essays in Honor of Leonard W. Weiss*, ed. by D. Audretsch, and J. Siegfried, pp. 13–34. Kluwer Academic Publishers.

Eaton, J., M. Eslava, M. Kugler, and J. Tybout (2007): "The Margins of Entry into Export Markets: Evidence from Colombia," Unpub. paper, NYU.

Eslava, M., J. Haltiwanger, A. Kugler, and M. Kugler (2004): "The Effects of Structural Reforms on Productivity and Profitability Enhancing Reallocation: Evidence from Colombia," *Journal of Development Economics*, 75, 333–371.

——— (2005): "Factor Adjustments After Deregulation: Panel Evidence from Colombian Plants," NBER Working Paper No. 11656.

——— (2006): "Plant Turnover and Structural Reforms in Colombia," *IMF Staff Papers*, 53, 58 – 75.

——— (2007): “Trade Reforms and Market Selection: Evidence from Manufacturing Plants in Colombia,” Unpub. paper, University of Maryland.

Farné, S. (2004): “Boletin del Observatorio del Mercado de Trabajo y la Seguridad Social No. 7,” Universidad Externado de Colombia.

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.

Goldberg, P. K., and F. Verboven (2001): “The Evolution of Price Dispersion in the European Car Market,” *Review of Economic Studies*, 68(4), 811 – 848.

Gollop, F. M., and J. L. Monahan (1991): “A Generalized Index of Diversification: Trends in U.S. Manufacturing,” *Review of Economics and Statistics*, 73(2), 318 – 330.

Grossman, G. M., and E. Helpman (1991): “Quality Ladders in the Theory of Growth,” *Review of Economic Studies*, 58, 43–61.

Grossman, G. M., and G. Maggi (2000): “Diversity and Trade,” *American Economic Review*, 90(5), 1255 – 1275.

Hallak, J. C. (2006): “Product Quality and the Direction of Trade,” *Journal of International Economics*, 68, 238–265.

Hallak, J. C., and P. Schott (2005): “Estimating Cross-Country Differences in Product Quality,” Unpub. Paper, Yale University.

Hallak, J. C., and J. Sivadasan (2008): “Productivity, Quality and Exporting Behavior under Minimum Quality Requirements,” Unpub. paper, University of Michigan.

Halpern, L., and M. Koren (2007): “Pricing to Firm: An Analysis of Firm- and Product-Level Import Prices,” *Review of International Economics*, 15(3), 574–591.

Helpman, E., M. J. Melitz, and S. R. Yeaple (2004): “Export versus FDI with Heterogenous Firms,” *American Economic Review*, 94(1), 300–316.

Hortaçsu, A., and C. Syverson (2007): “Cementing Relationships: Vertical Integration, Foreclosure, Productivity, and Prices,” *Journal of Political Economy*, 115(2), 250–301.

Hummels, D., and P. J. Klenow (2005): “The Variety and Quality of a Nation’s Exports,” *American Economic Review*, 95(3), 704–723.

Hummels, D., and A. Skiba (2004): “Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture,” *Journal of Political Economy*, 112(6), 1384–1402.

Iacovone, L., and B. Javorcik (2008): “Getting Ready: Preparing to Export,” Unpub. paper, Oxford University.

Johnson, R. C. (2007): “Trade and Prices with Heterogeneous Firms,” Unpub. paper, UC Berkeley.

Jones, C. I. (2007): “Intermediate Goods, Weak Links, and Superstars: A Theory of Economic Development,” Unpub. paper, UC Berkeley.

Katayama, H., S. Lu, and J. R. Tybout (2006): “Firm-Level Productivity Studies: Illusions and a Solution,” Unpub. paper, The University of Sydney, Charles River Associates and Pennsylvania State University.

Khandelwal, A. (2007): “The Long and Short (of) Quality Ladders,” Unpub. paper, Yale University.

Kremer, M. (1993): “The O-Ring Theory of Economic Development,” *Quarterly Journal of Economics*, 108(3), 551–575.

Lucas, R. (1978): “On the Size Distribution of Business Firms,” *Bell Journal of Economics*, 9(2), 508–523.

Melitz, M. (2000): “Estimating Firm-level Productivity in Differentiated Product Industries,” Unpub. paper, Harvard University.

——— (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.

Milgrom, P., and J. Roberts (1990): "The Economics of Modern Manufacturing: Technology, Strategy and Organization," *American Economic Review*, 80(3), 511–528.

Mobarak, A. M., and D. P. Purbasari (2006): "Corrupt Protection for Sale to Firms: Evidence from Indonesia," Unpub. paper, University of Colorado.

Oi, W. Y., and T. L. Idson (1999): "Firm Size and Wages," in *Handbook of Labor Economics. Volume 3B*, ed. by O. Ashenfelter, and D. Card, pp. 2165 – 2214. Elsevier, U Rochester.

Rauch, J. E. (1999): "Networks versus Markets in International Trade," *Journal of International Economics*, 48, 7–35.

Roberts, M. J., and D. Supina (1996): "Output Price, Markups, and Producer Size," *European Economic Review*, 40(3-5), 909–921.

——— (2000): "Output Price and Markup Dispersion in Micro Data: The Roles of Producer Heterogeneity and Noise," in *Advances in Applied Microeconomics, vol. 9*, ed. by M. R. Baye, pp. 1–36. Elsevier Science, JAI, Amsterdam, New York and Tokyo.

Roberts, M. J., and J. R. Tybout (1997): "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs," *American Economic Review*, 87(4), 545–564.

Sattinger, M. (1979): "Differential Rents and the Distribution of Earnings," *Oxford Economic Papers*, 31(1), 60 – 71.

Schott, P. (2004): "Across-Product versus Within-Product Specialization in International Trade," *Quarterly Journal of Economics*, 119(2), 647–678.

Shaked, A., and J. Sutton (1982): "Relaxing Price Competition through Product Differentiation," *Review of Economic Studies*, 49, 3–13.

Shaked, A., and J. Sutton (1987): "Product Differentiation and Industrial Structure," *Journal of Industrial Economics*, 36(2), 131–46.

Sutton, J. (1991): *Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration.* MIT Press, Cambridge, Mass.

——— (1998): *Technology and Market Structure: Theory and History.* MIT Press, Cambridge Mass.

——— (2007): "Quality, Trade and the Moving Window: The Globalization Process," *Economic Journal*, 117, F469–F498.

Syverson, C. (2004a): "Market Structure and Productivity: A Concrete Example," *Journal of Political Economy*, 112(6), 1181–1233.

——— (2004b): "Product Substitutability and Productivity Dispersion," *Review of Economics and Statistics*, 86(2), 534 – 550.

——— (2007): "Prices, Spatial Competition, and Heterogeneous Producers: An Empirical Test," *Journal of Industrial Economics*, 55(2), 197–222.

Tybout, J. (2003): "Plant- and Firm-Level Evidence on the 'New' Trade Theories," in *Handbook of International Trade*, ed. by E. K. Choi, and J. Harrigan. Basil Blackwell, Oxford.

Tybout, J. R. (1992): "Making Noisy Data Sing," *Journal of Econometrics*, 53, 25–44.

Verhoogen, E. A. (2004): "Trade, Quality Upgrading and Wage Inequality in the Mexican Manufacturing Sector: Theory and Evidence from an Exchange-Rate Shock," Center for Labor Economics, UC Berkeley, Working Paper No. 67, January.

——— (2008): "Trade, Quality Upgrading and Wage Inequality in the Mexican Manufacturing Sector," *Quarterly Journal of Economics*, 123(2), 489–530.

A Theory Appendix

A.1 Characterization of Equilibrium

In this section of the appendix, we complete the characterization of equilibrium. As mentioned above, we must impose an assumption on the "shape" parameter of the Pareto distribution in order to ensure that both the distribution of capability draws and the distribution of plant revenues have finite means, in our case that $k > \max(\eta, 1)$.

The fact that profitability is monotonically increasing in λ (which follows from the fact that $r(\lambda)$ is monotonically increasing in λ) implies that in each country in equilibrium there will be a cut-off value of λ for remaining in in the domestic market, call it λ^*; plants will leave immediately after receiving their capability draw if it is below λ^*. There is also a cut-off λ_x^* for entering the export market. The values of the cut-offs (which, because of symmetry, are the same in each country) are pinned down by three conditions. First, the profit of the plant on the margin between remaining in the domestic market and stopping production is zero:

$$\pi(\lambda^*) = \frac{r(\lambda^*)}{\sigma} - f = 0 \tag{A1}$$

where $r(\cdot)$ is given by (8d). Second, the additional profit of entering the export market for the plant on the margin between entering the export market and producing only for the domestic market is also zero:

$$\pi_x(\lambda_x^*) = \frac{r(\lambda_x^*)}{\sigma} - f_x = 0 \tag{A2}$$

where $r(\cdot)$ is again given by (8d). Third, there is a free-entry condition: the ex ante expected present discounted value of receiving a capability draw must be equal to the investment cost required to receive the draw, such that ex ante expected profits are zero. Formally, given the steady-state probability of death, δ, and assuming there is no discounting, the condition is:

$$[1 - G(\lambda^*)] \sum_{t=0}^{\infty} (1-\delta)^t \left\{ \frac{E(r(\lambda))}{\sigma} - f \right\} + [1 - G(\lambda_x^*)] \sum_{t=0}^{\infty} (1-\delta)^t \left\{ \frac{E_x(r(\lambda))}{\sigma} - f_x \right\} - f_e = 0 \tag{A3}$$

where $E(r(\lambda))$ and $E_x(r(\lambda))$ are the expected per-period revenues in the domestic and export markets, respectively, conditional on being in each market. Using (A1), (A2), and the fact that $\frac{r(\lambda)}{r(\lambda^*)} = \left(\frac{\lambda}{\lambda^*}\right)^\eta$, we have that $E(r(\lambda)) = \frac{k}{k-\eta}(\sigma f)$ and $E_x(r(\lambda)) = \frac{k}{k-\eta}(\sigma f_x)$. Then using (A3), we can solve for the entry cut-offs:

$$\lambda^* = \lambda_m \left\{ \frac{f\eta}{f_e \delta (k-\eta)} \left[1 + \left(\frac{f}{f_x} \right)^{\frac{k-\eta}{\eta}} \right] \right\}^{\frac{1}{k}} \tag{A4}$$

$$\lambda_x^* = \lambda^* \left(\frac{f_x}{f} \right)^{\frac{1}{\eta}} \tag{A5}$$

A particularly convenient feature of the Melitz (2003) framework which carries over to this model is that these cut-off values do not depend on the scale of the economy.

In steady state, the mass of new entrants in each country — that is, potential entrepreneurs who pay the investment cost to receive a capability draw and who have a capability above the cur-off to remain in the market — is equal to the mass of plants that die:

$$M_e\left(1 - G(\lambda^*)\right) = \delta M \tag{A6}$$

where M_e is the mass of entrepreneurs who pay the investment cost and M is the mass of firms that remain in business. Combining this equation with the free-entry condition (A3), it is straightforward to show that the total amount spent on investment, $M_e f_e$ (which is paid to input suppliers), is equal to total profits in the differentiated sector, and hence that total plant revenues in the differentiated sector are equal to total payments to input suppliers in the differentiated sector.[60]

Because $w = c$ for all input suppliers, regardless of sector, the total revenue of input suppliers in each country is simply $I\int_{\underline{c}}^{\overline{c}} c\, h(c)\, dc = I\mu_c$, where μ_c is the mean of the input quality distribution. In each country, the representative consumer spends a fraction of income β on the differentiated good. The symmetry between countries then ensures that total revenues of differentiated plants in a country are equal to total expenditures on varieties of the differentiated good in that country:

$$\beta I\mu_c = ME(r(\lambda)) + M_x E_x(r(\lambda)) \tag{A8}$$

Using the fact that $\frac{M_x}{M} = \frac{1-G(\lambda_x^*)}{1-G(\lambda^*)}$, we can then solve for the mass of plants:

$$M = \frac{\beta I\mu_c(k-\eta)}{k\sigma f\left[1 + \left(\frac{f}{f_x}\right)^{\frac{k-\eta}{\eta}}\right]} \tag{A9}$$

It remains to specify the condition that guarantees that the supply of each input quality if sufficient to satisfy the demand from the differentiated sector. The supply will be sufficient if the following holds for every range of plants $(\underline{\lambda}, \overline{\lambda})$:

$$M\int_{\underline{\lambda}}^{\overline{\lambda}} n(\lambda) g\left(\lambda|\lambda > \lambda^*\right) d\lambda < I\int_{c(\underline{\lambda})}^{c(\overline{\lambda})} h(c)dc \tag{A10}$$

where $n(\lambda) = \frac{x(\lambda)}{\lambda^a} = \frac{r(\lambda)}{\lambda^a p(\lambda)}$ is the number of input units used by plant λ; $g\left(\lambda|\lambda > \lambda^*\right)$ is the density of plant capability conditional on being in operation (in the domestic market); $c(\underline{\lambda})$ and $c(\overline{\lambda})$ are the input qualities chosen by $\underline{\lambda}$ and $\overline{\lambda}$, which is to say $\underline{\lambda}^{\frac{b}{2}}$ and $\overline{\lambda}^{\frac{b}{2}}$. Note that total revenues of plants in the differentiated sector can be expressed as $R \equiv XP = \beta I\mu_c$ and that P (using (3) and (A5)) can be rewritten:

$$\begin{aligned} P &= \left(\frac{\sigma}{\sigma-1}\right)\left\{M\int_{\lambda^*}^{\infty}\lambda^\eta g\left(\lambda|\lambda > \lambda^*\right)d\lambda + M_x\int_{\lambda_x^*}^{\infty}\lambda^\eta g\left(\lambda|\lambda > \lambda_x^*\right)d\lambda\right\}^{\frac{1}{1-\sigma}} \\ &= \left(\frac{\sigma}{\sigma-1}\right)\left\{\left(\frac{k}{k-\eta}\right)M\left(\lambda^*\right)^\eta\left[1 + \left(\frac{f}{f_x}\right)^{\frac{k-\eta}{\eta}}\right]\right\}^{\frac{1}{1-\sigma}} \end{aligned} \tag{A11}$$

---

[60] Total profit in the differentiated sector is given by the mass of plants in production times the expected profit conditional on being in the market:

$$\Pi_D = M\left\{\left[\frac{E(r(\lambda))}{\sigma} - f\right] + \frac{1-G(\lambda_x^*)}{1-G(\lambda^*)}\left[\frac{E_x(r(\lambda))}{\sigma} - f_x\right]\right\} \tag{A7}$$

where $E(r(\lambda))$ and $E_x(r(\lambda))$ are defined as above. Combining (A3), (A6), and (A7), we have $\Pi = M_e f_e$.

Using these facts, as well as (A4) and the fact that that $\frac{M_x}{M} = \frac{1-G(\lambda_x^*)}{1-G(\lambda^*)}$, and noting the change of variables in the integrals of (A10), condition (A10) is equivalent to the condition:

$$h(c) > \beta \left(\frac{\sigma}{\sigma-1}\right)\left(\frac{\lambda_m f \eta}{f_e \delta (k-\eta)}\right)^{\frac{k-\eta}{k}} \left[1+\left(\frac{f}{f_x}\right)^{\frac{k-\eta}{\eta}}\right]^{-\frac{\eta}{k}} \frac{\widetilde{k}}{c^{\widetilde{k}+1}} \quad \forall\, c \in [1,\infty) \tag{A12}$$

where $\widetilde{k} \equiv \frac{2}{b}(k-\eta)+1$. We assume that this condition is satisfied, and hence that the supply of inputs of each quality c is sufficient to meet the demand from the differentiated sector.[61]

A particular special case is instructive. If c has a Pareto distribution with shape parameter $\widetilde{k}$, then $h(c) = \frac{\widetilde{k}}{c^{\widetilde{k}+1}}$ and condition (A12) simplifies to:

$$\beta < \frac{\left[1+\left(\frac{f}{f_x}\right)^{\frac{k-\eta}{\eta}}\right]^{\frac{\eta}{k}}}{\left(\frac{\sigma}{\sigma-1}\right)\left(\frac{\lambda_m f \eta}{f_e \delta (k-\eta)}\right)^{\frac{k-\eta}{k}}} \tag{A13}$$

In other words, the condition places an upper bound on the expenditure share of the representative consumer on differentiated goods. Intuitively, given that the distribution of c has support over the relevant range, a sufficiently low β will guarantee that (A12) holds even in the general case.

A.2 "Quality" Melitz Model

This section spells out a "quality" interpretation of the Melitz (2003) model, which is alluded to but not made explicit in Melitz's original paper. As mentioned in the text, our model reduces to the Melitz (2003) model when $b = 0$.[62] We also set the size of the non-differentiated sector to zero (that is, we let $\beta \to 1$) and assume that the distribution of input qualities collapses to a point, with $c = 1$.[63] Let $\varphi \equiv \lambda^a$ and express other variables in terms of φ. In this case, (1)-(4) and (8a)-(8d) become:

$$U = X = \left[\int_{\varphi\in\Phi} x(\varphi)^{\frac{\sigma-1}{\sigma}} d\varphi\right]^{\frac{\sigma}{\sigma-1}} \tag{A14a}$$

$$P \equiv \left[\int_{\varphi\in\Phi} p(\varphi)^{1-\sigma} d\varphi\right]^{\frac{1}{1-\sigma}} \tag{A14b}$$

---

[61] Note that we have assumed that $\lambda_m > 1$, hence $\lambda_m^{\frac{b}{2}} > 1$, and hence there is positive support in the distribution of c for all choices of input quality in equilibrium.

[62] Note that Melitz (2003) considers a general distribution of productivity draws, rather than the Pareto distribution of this paper and Helpman, Melitz, and Yeaple (2004). This is of little consequence for the discussion here.

[63] An alternative interpretation of the collapsing of the input quality distribution would be that the heterogeneity in c does not disappear but rather becomes irrelevant, since when $b = 0$ higher-λ have no differential incentive to use input quality above the minimum and all plants choose the same level of c, which can be normalized to 1.

$$w(\varphi) = q(\varphi) = 1 \tag{A14c}$$

$$p(\varphi) = \left(\frac{\sigma}{\sigma-1}\right)\frac{1}{\varphi} \tag{A14d}$$

$$r(\varphi) = \left(\frac{\sigma-1}{\sigma}\right)^{\sigma-1} XP^{\sigma}\varphi^{\sigma-1} \tag{A14e}$$

$$x(\varphi) = \left(\frac{\sigma}{\sigma-1}\right)^{-\sigma} XP^{\sigma}\varphi^{\sigma} \tag{A14f}$$

which correspond exactly to Melitz (2003).

Now consider the following thought experiment, which generates the quality interpretation of the Melitz model. Suppose that the above equations refer to goods measured in *quality* units, which we will call "utils". Suppose further that higher-φ plants, in addition to requiring fewer units of inputs to produce one util of output, also produce goods with more utils per physical unit, where utils per physical unit are given by:

$$\tilde{q}(\varphi) = \varphi^{\epsilon} \tag{A15}$$

The existence of a relationship of this kind is alluded to in Melitz (2003, p. 1699) but not explicitly specified. Given (A15), price and quantity in *physical* units are given by:

$$\tilde{p}(\varphi) = p(\varphi)\,\tilde{q}(\varphi) = \left(\frac{\sigma}{\sigma-1}\right)\varphi^{\epsilon-1} \tag{A16}$$

$$\tilde{x}(\varphi) = \frac{x(\varphi)}{\tilde{q}(\varphi)} = \left(\frac{\sigma}{\sigma-1}\right)^{-\sigma} XP^{\sigma}\varphi^{\sigma-\epsilon} \tag{A17}$$

The expression for revenues is unchanged by the redefinition of units.[64]

Several remarks are in order. First, if $\epsilon > 1$, then both output price in physical units and revenues are increasing in φ and hence are positively correlated with one another. Note also that setting $\epsilon = 1$ yields a model in which higher φ corresponds to higher quality but marginal cost and hence output price in physical units are constant, as alluded to by Melitz (2003, p. 1699).

Second, this "quality" Melitz model is isomorphic to the quality model of Baldwin and Harrigan (2007, Section 4) if one abstracts from the differences in distance between countries. Baldwin and Harrigan's parameter a represents marginal cost, which here corresponds to $\varphi^{\epsilon-1}$, and their θ corresponds to $\frac{1}{\epsilon-1}$. Their assumption that $\theta > 0$ here corresponds to the condition that $\epsilon > 1$, which guarantees that output price is increasing in φ. The value-added of the Baldwin-Harrigan model over this quality Melitz model is that it explicitly considers distance and the differential selection of higher-productivity firms into more-distant markets.

Third, the key difference between this quality Melitz model (with $\epsilon > 1$) and the quality model we present in this paper lies in the role of inputs. Here output price and marginal cost per physical unit are increasing in φ because plants are using *more units of inputs of homogeneous quality* to produce each physical unit, rather than inputs of higher quality as in our model. (That

---

[64] The aggregates P and X can be rewritten as:

$$X = \left[\int_{\varphi\in\Phi} \left(\tilde{x}(\varphi)\tilde{q}(\varphi)\right)^{\frac{\sigma-1}{\sigma}} d\varphi\right]^{\frac{\sigma}{\sigma-1}} \tag{A18}$$

$$P \equiv \left[\int_{\varphi\in\Phi} \left(\frac{\tilde{p}(\varphi)}{\tilde{q}(\varphi)}\right)^{1-\sigma} d\varphi\right]^{\frac{1}{1-\sigma}} \tag{A19}$$

is, higher-φ plants use fewer units of inputs *per util*, but since the number of utils per physical unit increases in φ faster than input requirements decline, they use more physical units of inputs per physical unit of output.)

Fourth, even if one were to introduce heterogeneity of inputs in this quality Melitz framework, in the absence of the complementarity between plant capability and input quality, there would be no systematic reason for higher-φ plants to use higher-quality inputs.

Fifth, a shortcoming of the quality Melitz/Baldwin-Harrigan framework for addressing issues of quality differentiation is that quality is a deterministic function of a plant's capability draw. There is no quality choice. Hence quality does not depend on factors such as the technological possibilities for upgrading quality or consumers' willingness to pay for such improvements, and there is no endogenous variation in the extent of quality differentiation across sectors.

B Data Appendix

B.1 Variable Definitions

Output unit value: Value of output of 8-digit product, divided by number of physical units of product produced. Output is sales plus net intra-firm transfers plus net increase in inventories. We also refer to the output unit value (somewhat loosely, since it represents an average) as the *output price*. In 1998 Colombian pesos.

Input unit value: Value consumed of 8-digit product, divided by number of physical units of product consumed. Consumption is purchases minus net intra-firm transfers minus net increase in inventories. We also refer to the input unit value (somewhat loosely, since it represents an average) as the *input price*. In 1998 Colombian pesos.

Total output: Total value of output of all products, valued at factory price. Total output is sales plus net transfers to other plants in same firm plus net increases in inventories. In billions of 1998 Colombian pesos.

Employment: The number of permanent, paid employees.

Exporter: Indicator variable taking the value 1 if plant has export sales > 0, and 0 otherwise.

Export share: Export sales as a fraction of total sales.

Average earnings: Total annual wage bill of permanent, remunerated workers, in millions of 1998 Colombian pesos, divided by total number of permanent, remunerated workers on Nov. 15 of corresponding year.

Average white-collar earnings: Annual wage bill of permanent, remunerated white-collar workers, in millions of 1998 Colombian pesos, divided by number of permanent, remunerated white-collar workers on Nov. 15 of corresponding year. White-collar workers defined as managers (*directivos*), non-production salaried workers (*empleados*), and technical employees (*técnicos*). The white-collar/blue-collar distinction is available on a consistent basis only for 1982-1994.

Average blue-collar earnings: Annual wage bill of permanent, remunerated blue-collar workers, in millions of 1998 Colombian pesos, divided by number of permanent, remunerated blue-collar workers on Nov. 15 of corresponding year. Blue-collar workers are defined as operators (*obreros* and *operarios*) and apprentices (*aprendices*). The white-collar/blue-collar distinction is available on a consistent basis only for 1982-1994.

Advertising/sales ratio: Ratio of advertising expenditures to total sales at sector level, from the U.S. Federal Trade Commission (FTC) 1975 Line of Business Survey. Converted from FTC 4-digit industry classification to ISIC 4-digit rev. 2 classification using verbal industry descriptions.

Advertising + R&D/sales ratio: Ratio of advertising plus research and development (R&D) expenditures to total sales, from the U.S. Federal Trade Commission (FTC) 1975 Line of Business Survey. Converted from FTC 4-digit industry classification to ISIC 4-digit rev. 2 classification using verbal industry descriptions.

Rauch (1999) measure of horizontal differentiation: SITC 4-digit sectors classified by Rauch's "liberal" classification as "homogeneous" or "reference-priced" are assigned 0, others are assigned 1. SITC 4-digit industries were then converted to ISIC rev. 2 4-digit using concordance from OECD, which generated some fractional values.

Modified Gollop-Monahan measure of differentiation: Following Bernard and Jensen (2007), we use just the "dissimilarity" component of the full Gollop and Monahan (1991) index. We define the index as follows:

$$GM_k = \left(\sum_{j,t} \frac{s_{ijkt} - \overline{s}_{ik}}{2} \right)^{\frac{1}{2}}$$

where i, j, k, and t index products, plants, 5-digit industries and years; s_{ijkt} is the expenditure share of plant j in industry k on input i in year t; and $\overline{s}_{ik}$ is the average expenditure share on input i by all plants in industry k in all years.[65]

Herfindahl index (of purchasers): Sum of squares of expenditure shares of purchasers of the corresponding 8-digit input, where the expenditure share is the expenditure by a given purchaser as a share of total expenditures on the good.

Herfindahl index (of suppliers): Sum of squares of market shares of producers of the corresponding 8-digit input.

Purchaser share: Expenditures on product by plant as a share of total expenditures on product by all plants in a given year.

All monetary variables have been deflated to constant 1998 values using the national producer price index. Average 1998 exchange rate: 1,546 pesos/US$1.

B.2 Data Processing

The data we use in this paper are from the *Encuesta Anual Manufacturera* (EAM) [Annual Manufacturing Survey]. Plant-level data are available over the 1977-2005 period, but product-level data are available only for 1982-2005. The EAM is a census of all manufacturing plants in Colombia with 10 or more workers, with the following qualification. Prior to 1992, the sole

---

[65] Note that Gollop and Monahan (1991) construct a dissimilarity measure at the level of products, rather than plants. But this requires information on the input mix for each product. To recover this information, Gollop and Monahan (1991) focus on plants that produce only the particular product for which they need information on the input mix. This solution seems unattractive, since plants producing only a particular good are a selected subsample of the set of plants producing that good. Instead of following this solution, we calculate the index of dissimilarity at the plant level, where the input mix is observed for all plants.

criterion for initial inclusion of a plant in the census was that the plant have a total of 10 or more employees.[66] Beginning in 1992, an additional criterion was added: a plant would be included if it had 10 or more workers *or* nominal value of total output (defined as in Appendix B.1) in excess of 65 million Colombian pesos (approx. US$95,000) (DANE, 2004, p. 8). The monetary limit has been raised in nominal terms over time. There are two exceptions to these rules. First, once a plant is included in the sample it is followed over time until it goes out of business, regardless of whether the criteria for inclusion continue to be satisfied. Second, multi-plant firms are included, even if not all plants satisfy one of the above criteria. To maintain consistency of the sample over time, we removed all plants with fewer than 10 employees.[67]

The longitudinal links between plant-level observations we use are those that are reported directly by DANE. In 1991 and again in 1992, plant identification numbers were changed, with the result that it was no longer possible to follow some plants over time, despite the fact that they remained in the dataset.[68]

From 1982-2000, the product-level data were reported using an 8-digit classification system with four digits from the International Standard Industrial Classification (ISIC) revision 2 and four Colombia-specific digits (one of which is only used for verification purposes).[69] In 2001, a new classification was constructed, with the first five digits based on the U.N. Central Product Classification (CPC) version 1.0 and two Colombia-specific digits. We used a concordance provided by DANE to convert back to the earlier product classification. There are approximately 6,000 distinct product categories.

To construct a plant's 5 digit industry, we aggregated revenues within plants across all years from the 8-digit to the 5-digit level, then chose the 5-digit category with the greatest share of total revenues. Our industry categories thus do not change over time within plant. Note, however, that plants continually changes their output and input mixes, and these changes are accommodated by our econometric procedure.

To reduce the influence of measurement error and outliers, we carried out the following additional cleaning procedures:

1. In the plant-level file, we dropped any plant-year observation for which a key variable — total output, employment, white-collar wage, blue-collar wage or average wage — differed by more than a factor of 5 from adjacent periods.[70]

2. In the plant-level file, we dropped plants that were reported to be cooperatives, publicly owned, or owned by a religious organization.

---

[66] This was the sole criterion over the 1970-1992 period. Prior to 1970, an additional output criterion had been in place.

[67] In implementing this criterion, we followed DANE's definition and counted all employees, including those that are unpaid or temporary.

[68] Eslava, Haltiwanger, Kugler, and Kugler (2004) construct some links probabilistically (see the data appendix of that paper); we use only the links constructed on the basis of name, address and telephone information.

[69] The Spanish acronym for this classification system is CIIU2AC, for *Clasificación Internacional Industrial Uniforme revisión 2 adaptada para Colombia* [ISIC revision 2 adapted for Colombia].

[70] To be precise, an observation was dropped if one of the following criteria was met: (a) the plant-year observation differed by more than a factor of 5 from both the previous and the subsequent year; (b) the observation differed by more than a factor of 5 from the previous year and data for the subsequent year was missing; (c) the observation differed by more than a factor of 5 from the subsequent year and data for the previous year was missing; (d) the observation differed by more than a factor of 5 from the subsequent year but not the previous year and the subsequent year did not differ by more than a factor of 5 from the following year; or (e) the observation differed by more than a factor of 5 from the previous year but not the subsequent year and the previous year did not differ by more than a factor of 5 from the preceding year.

3. In the plant-level file, , we "winsorized" the data within each year (Angrist and Krueger, 1999) for the variables total output, employment, white-collar wage, blue-collar wage or average wage , setting all values below the 1^{st} percentile to the value at the 1^{st} percentile, and all values above the 99^{th} percentile to the 99^{th} percentile.

4. In the product-level data, we dropped product-year observations that were not assigned to any 8-digit product code (i.e. that were in a "not elsewhere classified" category with no information on industry).

5. In the product-level data, we dropped information on unit values for subcontracted outputs or inputs, since the reported value typically does not reflect the market price. (The product-level data contain an identifier to indicate whether the good is produced or purchased under a sub-contracting arrangement.) Goods produced under subcontract are included in total output, however.

6. In the product-level file, we dropped product-year observations reporting values of revenues or expenditures or physical quantities equal to the integers 1, 2 or 3. These observations were responsible for many of the most severe outliers in the raw data. The integer values 1, 2, and 3 appear to be reporting or transcription errors.

7. In the product level file, we winsorized real output and input unit values within product, separately for outputs and inputs. Because of the small number of observations for many product-years and the noise in the unit value information, we winsorized within product for all years together and at the 5^{th} and 95^{th} percentiles, rather than 1^{st} and 99^{th} as above.

8. In the product-level file, we carried out an additional winsorizing procedure, winsorizing observations on log real unit values that differed from the mean by 5 times the standard deviation for log real unit values within product, separately for outputs and inputs.

9. In the plant-level and product-level files, we dropped observations corresponding to any plant that did not have complete information on key variables: total output, employment, white-collar wage, blue-collar wage, average wage, output prices and quantities and input prices and quantities.

As discussed in footnote 54 above, in order to carry out the estimation of plant-year effects in the two-step method in Section 9.1, plants must be in a connected "network" of plants, where a plant is connected if it produces a good that is also produced by another plant in the network. More than 95% of plants are in the largest such chain. In order to maintain as consistent a sample as possible across different specifications, we use only the "connected" plants also when using the one-step procedure described by equation (10).

We refer to the unbalanced panel consisting of all plant-year observations that survive the cleaning procedure as the 1982-2005 panel. We refer to the subset of observations of that panel that contain complete information on exports, white-collar and blue-collar earnings (which are only available on a consistent basis for the period 1982-1994) as the 1982-2005 panel.[71]

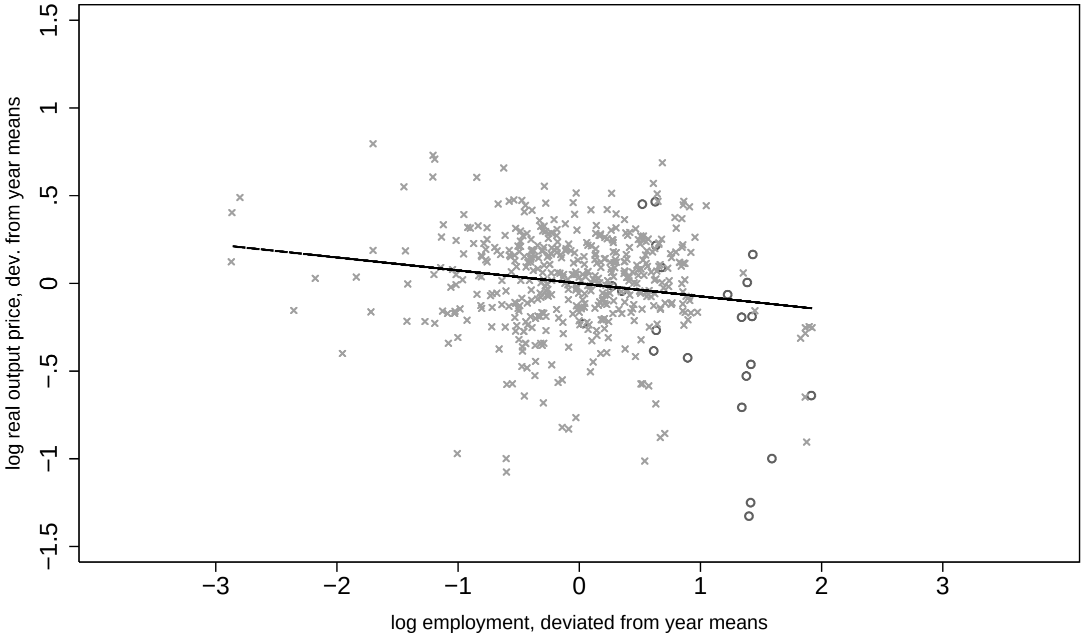
The primary sub-national administrative region in Colombia is the *departamento*, of which there are 32 plus the federal district of Bogotá. Four *departamentos* have zero plants in our sample. Another eight little-populated *departamentos* — Amazonas, Arauca, Caqueta, Casanaré, Chocó, La Guajira, Putumayo, and San Andres — together have just 184 plant-year observations in the entire 1982-2005 panel. We aggregated these eight *departamentos* into a single region.

---

[71] Information on exports and imported inputs is also available in 2000-2005, but the information is collected in a different way and there appear to be incomparabilities between the 1982-1994 and 2000-2005 values.

Figure 1:
Illustrative example: output and input prices, hollow brick, 1982-1994 data

A. Output prices, hollow brick (ladrillo hueco)
x=non−exporter, o=exporter; slope=−0.074, se=0.047



B. Input prices, common clay, for producers of hollow brick
x=non−exporter, o=exporter; slope=−0.247, se=0.103

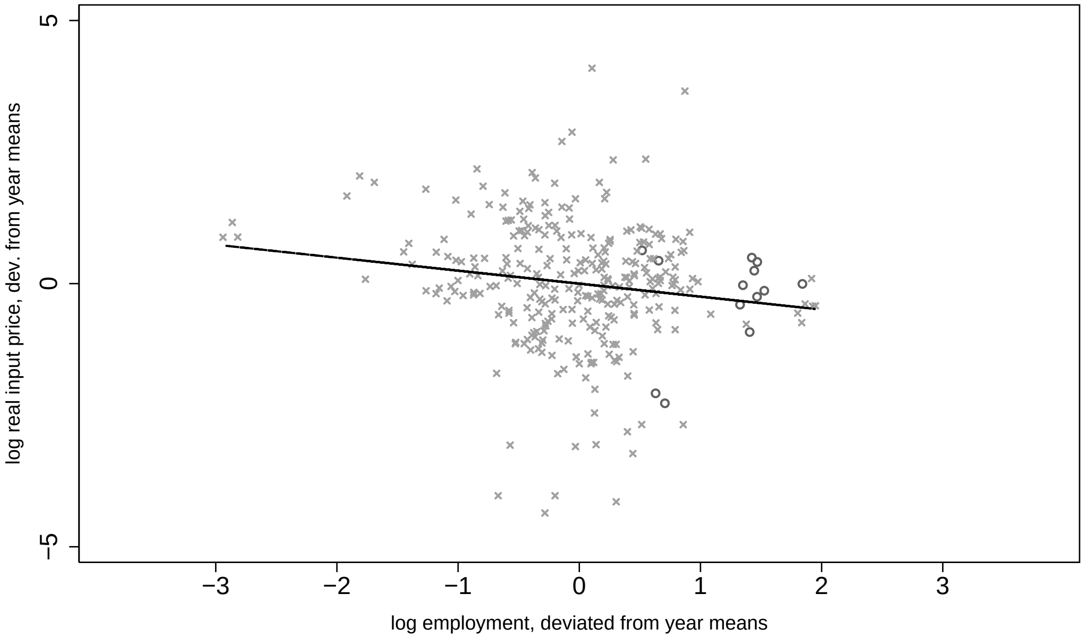
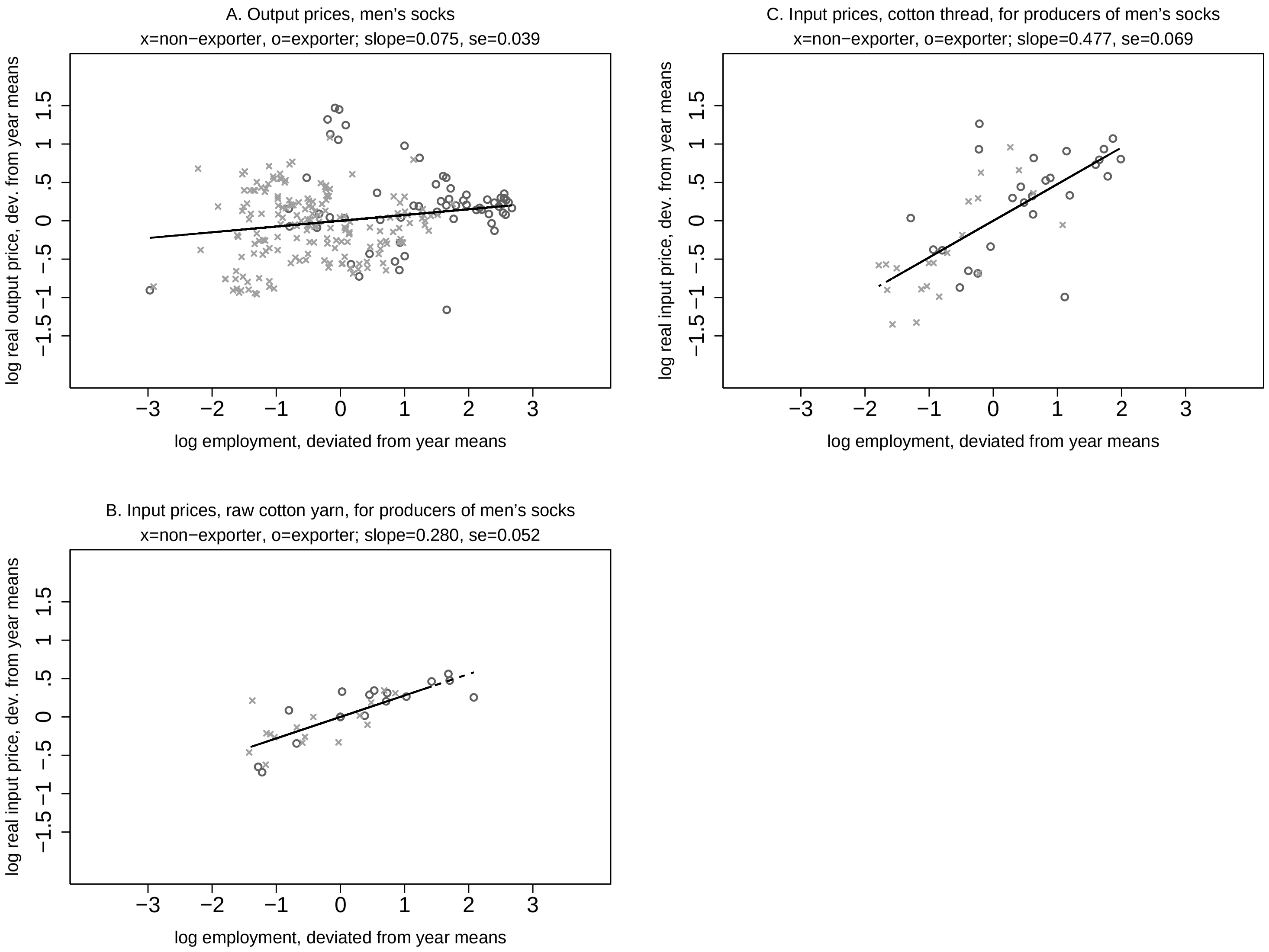


Figure 2:
Illustrative example: output and input prices, men's socks, 1982-1994 data


A. Output prices, men's socks
x=non−exporter, o=exporter; slope=0.075, se=0.039

C. Input prices, cotton thread, for producers of men's socks
x=non−exporter, o=exporter; slope=0.477, se=0.069

B. Input prices, raw cotton yarn, for producers of men's socks
x=non−exporter, o=exporter; slope=0.280, se=0.052

Table 1:
Summary statistics, plant-level data

	1982-1994 panel			1982-2005 panel
	non-exporters (1)	exporters (2)	all plants (3)	all plants (4)
Output	2.77	11.98	4.35	5.47
	(0.04)	(0.19)	(0.05)	(0.04)
Employment	56.65	193.16	79.98	70.40
	(0.40)	(2.06)	(0.53)	(0.34)
Avg. earnings	3.26	4.66	3.50	4.39
	(0.01)	(0.02)	(0.01)	(0.01)
White-collar earnings	4.36	6.62	4.75	
	(0.01)	(0.03)	(0.01)	
Blue-collar earnings	2.77	3.47	2.89	
	(0.00)	(0.01)	(0.00)	
White-collar/blue-collar earnings ratio	1.62	1.97	1.68	
	(0.00)	(0.01)	(0.00)	
White-collar employment share	0.29	0.33	0.30	
	(0.00)	(0.00)	(0.00)	
Number of output categories	3.44	4.49	3.62	3.61
	(0.01)	(0.04)	(0.01)	(0.01)
Number of input categories	10.29	17.10	11.46	11.69
	(0.03)	(0.15)	(0.04)	(0.03)
Export share of sales		0.17		
		(0.00)		
Import share of input expenditures	0.06	0.23	0.09	
	(0.00)	(0.00)	(0.00)	
N (plant-year obs.)	49546	10216	59762	114500
N (distinct plants)	9352	2308	10106	13582

Notes: Standard errors of means in parentheses. Exporter defined as export sales > 0. Export share is fraction of total sales derived from exports. Output is annual sales, measured in billions of 1998 Colombian pesos. Annual earnings measured in millions of 1998 pesos. Average 1998 exchange rate: 1,546 pesos/US$1. Number of output or input categories refers to number of distinct categories in which non-zero revenues or expenditures are reported. See Appendix B.1 for more detailed variable descriptions and Appendix B.2 for details of data processing.

Table 2:
Summary statistics, product-level data, by ISIC 3-digit industry, 1982-2005 panel

		product as output			product as input		
ISIC rev. 2 major group	# products (1)	avg. # selling plants per year (2)	within-product std. dev. log price (3)	within-prod.-year std. dev. log price (4)	avg. # purchasing plants per year (5)	within-product std. dev. log price (6)	within-prod.-year std. dev. log price (7)
Food	446	43.82	0.51	0.46	124.60	0.55	0.51
Beverages	32	34.15	0.50	0.44	73.64	0.57	0.49
Tobacco	5	3.16	0.35	0.29	2.31	0.77	0.60
Textiles	227	10.60	0.72	0.64	240.99	0.80	0.78
Apparel, exc. footwear	171	38.08	0.58	0.55	27.85	0.71	0.67
Leather prod., exc. footwear/apparel	71	13.35	0.86	0.70	124.41	0.83	0.61
Footwear, exc. rubber/plastic	28	43.89	0.49	0.46	39.39	0.94	0.90
Wood products, exc. furniture	77	21.54	1.07	0.95	121.04	0.87	0.81
Furniture, exc. metal	79	54.25	0.89	0.85	3.86	0.88	0.61
Paper products	138	22.36	0.98	0.84	363.01	0.91	0.89
Printing and publishing	83	79.90	1.22	1.15	505.76	1.10	1.08
Industrial chemicals	277	5.17	0.78	0.67	102.86	0.85	0.81
Other chemical products	220	15.05	0.83	0.78	198.99	0.86	0.82
Petroleum refineries	29	1.38	0.89	0.28	70.66	0.87	0.83
Misc. petroleum/coal products	16	8.12	0.80	0.71	154.99	0.68	0.66
Rubber products	82	7.35	0.74	0.64	105.06	0.94	0.91
Plastic products	232	19.03	1.00	0.87	331.10	0.95	0.91
Pottery, china, earthenware	26	3.03	0.75	0.52	10.07	1.25	1.06
Glass products	85	4.47	0.86	0.71	51.44	0.89	0.85
Other non-metallic mineral products	110	13.94	0.71	0.62	48.30	0.92	0.85
Iron and steel basic industries	61	12.66	0.93	0.81	143.57	0.77	0.75
Non-ferrous metal basic industries	97	4.51	0.78	0.61	44.56	0.75	0.70
Metal prod., exc. machinery/equip.	406	13.72	1.05	0.97	210.26	1.00	0.95
Machinery, exc. electrical	285	7.12	1.33	1.18	27.02	1.37	1.28
Electrical machinery	168	6.40	1.41	1.26	161.88	1.30	1.22
Transport equipment	180	5.87	0.98	0.79	5.18	1.20	0.96
Professional equipment, n.e.c.	79	3.36	1.23	0.92	11.51	1.29	1.12
Other manufactures	172	7.05	1.14	0.99	137.81	0.95	0.89
All sectors	3882	30.06	0.87	0.79	193.30	0.87	0.83

Notes: Number of products is number of distinct products with non-zero sales in any year. Average number of selling or purchasing plants is average across products of number of distinct plants selling or purchasing product in each year. Within-product and within-product-year standard deviations are calculated as standard deviations of residuals from regression of log real prices on sets of product effects or product-year effects, respectively. Averages assign equal weight to each plant-product-year observation in product-level data on outputs (Columns 2-4) or inputs (Columns 5-7). See Appendix B.2 for details of data processing.

Table 3:
Product-level prices vs. plant size, 1982-2005 panel

	OLS (1)	Reduced form (2)	IV (3)
A. Dependent variable: log real output unit value			
log total output	0.021***		0.025***
	(0.005)		(0.006)
log employment		0.026***	
		(0.007)	
product-year effects	Y	Y	Y
industry effects	Y	Y	Y
region-year effects	Y	Y	Y
R^2	0.90	0.90	
N (obs.)	413789	413789	413789
N (plants)	13582	13582	13582
B. Dependent variable: log real input unit value			
log total output	0.015***		0.011***
	(0.002)		(0.003)
log employment		0.012***	
		(0.003)	
product-year effects	Y	Y	Y
industry effects	Y	Y	Y
region-year effects	Y	Y	Y
R^2	0.78	0.78	
N (obs.)	1338921	1338921	1338921
N (plants)	13582	13582	13582

Notes: Total output is total value of production, defined as sales plus net transfers plus net change in inventories. In Column 3, log employment is instrument for log total output; the coefficient on log employment, its robust standard error and the R^2 in the first stage are 1.058, 0.011 and 0.733 in Panel A and 1.082, 0.010 and 0.782 in Panel B, respectively. Product-year and industry effects are not perfectly collinear because industry is defined as the industry category with the greatest share of plant sales, and two plants producing the same product may be in different industries. Errors clustered at plant level. N (plants) reports number of clusters (i.e. distinct plants that appear in any year). Robust standard errors in brackets. *10% level, **5% level, ***1% level. See Appendix B.1 for more detailed variable descriptions and Appendix B.2 for details of data processing.

Table 4:
Product-level prices vs. plant size and exporting variables, 1982-1994 panel

	(1)	(2)	(3)	(4)	(5)
A. Dependent variable: log real output price					
log employment	0.025***			0.009	0.020**
	(0.008)			(0.008)	(0.008)
exporter		0.114***		0.104***	
		(0.022)		(0.023)	
export share			0.288**		0.251*
			(0.137)		(0.142)
product-year effects	Y	Y	Y	Y	Y
industry effects	Y	Y	Y	Y	Y
region-year effects	Y	Y	Y	Y	Y
R^2	0.90	0.90	0.90	0.90	0.90
N (obs.)	216155	216155	216155	216155	216155
N (plants)	10106	10106	10106	10106	10106
B. Dependent variable: log real input price					
log employment	0.013***			0.008**	0.013***
	(0.004)			(0.004)	(0.004)
exporter		0.037***		0.028***	
		(0.009)		(0.009)	
export share			0.021		-0.002
			(0.027)		(0.027)
product-year effects	Y	Y	Y	Y	Y
industry effects	Y	Y	Y	Y	Y
region-year effects	Y	Y	Y	Y	Y
R^2	0.80	0.80	0.80	0.80	0.80
N (obs.)	684746	684746	684746	684746	684746
N (plants)	10106	10106	10106	10106	10106

Notes: Exporter equals 1 if plant has exports>0, and 0 otherwise. Export share is fraction of total sales derived from exports. Product-year and industry effects are not perfectly collinear because industry is defined as the industry category with the greatest share of plant sales, and two plants producing the same product may be in different industries. Errors clustered at plant level. N (plants) reports number of clusters (i.e. distinct plants that appear in any year). Robust standard errors in parentheses. *10% level, **5% level, ***1% level. See Appendix B.1 for more detailed variable descriptions and Appendix B.2 for details of data processing.

Table 5:
Wage variables vs. plant size and exporting variables, 1982-1994 panel

dependent var.:	log avg. earnings			log blue-collar earnings			log white-collar earnings			log earnings ratio		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
log employment	0.140***			0.100***			0.198***			0.098***		
	(0.003)			(0.003)			(0.004)			(0.003)		
exporter		0.269***			0.181***			0.326***			0.145***	
		(0.009)			(0.007)			(0.011)			(0.008)	
export share			0.318***			0.212***			0.478***			0.266***
			(0.026)			(0.022)			(0.032)			(0.024)
industry effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
region-year effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2	0.51	0.46	0.42	0.40	0.36	0.33	0.42	0.34	0.30	0.16	0.12	0.11
N (obs.)	59762	59762	59762	59762	59762	59762	59762	59762	59762	59762	59762	59762
N (plants)	10106	10106	10106	10106	10106	10106	10106	10106	10106	10106	10106	10106

Notes: Exporter equals 1 if plant has exports > 0, and 0 otherwise. Export share is fraction of total sales derived from exports. Regressions analogous to columns (1)-(3) with the white-collar employment share as dependent variable (omitted from table to conserve space) have coefficients 0.007, 0.036, and -0.012 with standard errors 0.002, 0.004, and 0.010 respectively. Errors clustered at plant level. N (plants) reports number of clusters (i.e. distinct plants that appear in any year). Robust standard errors in brackets. *10% level, **5% level, ***1% level. See Appendix B.1 for more detailed variable descriptions and Appendix B.2 for details of data processing.

Table 6:
Measures of differentiation and concentration, by ISIC 3-digit industry, 1982-2005 panel

ISIC rev. 2 major group		advertising intensity (1)	R&D + advertising intensity (2)	Rauch (1999) index (3)	Herfindahl index (suppliers) (4)	Herfindahl index (purchasers) (5)
311-312	Food	0.026	0.029	0.35	0.24	0.45
313	Beverages	0.045	0.046	0.68	0.20	0.70
314	Tobacco	0.076	0.082	0.25	0.62	0.74
321	Textiles	0.014	0.019	0.88	0.30	0.27
322	Apparel, exc. footwear	0.015	0.018	1.00	0.17	0.93
323	Leather prod., exc. footwear/apparel	0.000	0.002	0.67	0.36	0.24
324	Footwear, exc. rubber/plastic	0.015	0.017	1.00	0.22	0.24
331	Wood products, exc. furniture	0.002	0.005	0.58	0.29	0.50
332	Furniture, exc. metal	0.014	0.019	1.00	0.13	0.83
341	Paper products	0.002	0.006	0.30	0.33	0.13
342	Printing and publishing	0.028	0.041	0.86	0.18	0.50
351	Industrial chemicals	0.005	0.029	0.18	0.57	0.35
352	Other chemical products	0.083	0.107	0.95	0.36	0.46
353	Petroleum refineries	0.002	0.004	0.09	0.88	0.38
355	Rubber products	0.012	0.026	1.00	0.43	0.40
356	Plastic products	0.008	0.031	0.79	0.33	0.28
361	Pottery, china, earthenware	0.007	0.020	1.00	0.56	0.92
362	Glass products	0.008	0.046	1.00	0.51	0.38
369	Other non-metallic mineral products	0.006	0.017	0.68	0.32	0.54
371	Iron and steel basic industries	0.001	0.006	0.25	0.41	0.22
372	Non-ferrous metal basic industries	0.002	0.011	0.02	0.60	0.33
381	Metal prod., exc. machinery/equip.	0.011	0.018	0.79	0.46	0.34
382	Machinery, exc. electrical	0.007	0.028	1.00	0.49	0.55
383	Electrical machinery	0.009	0.031	0.98	0.49	0.57
384	Transport equipment	0.008	0.033	1.00	0.51	0.75
385	Professional equipment, n.e.c.	0.013	0.052	0.99	0.66	0.70
390	Other manufactures	0.040	0.052	0.90	0.45	0.89
All sectors		0.020	0.029	0.74	0.28	0.43

Notes: Advertising intensity defined as ratio of advertising expenditures to total industry sales and advertising + R&D intensity defined as ratio of advertising and R&D expenditures to total industry sales, using data from the U.S. Federal Trade Commission (FTC) 1975 Line of Business Survey, converted from FTC 4-digit industry classification to ISIC 4-digit rev. 2 classification using verbal industry descriptions. At SITC 4-digit level, Rauch (1999) measure set to 0 if good is "homogeneous" or "reference-priced" according to the Rauch "liberal" definition, to 1 if reported not to be in either category, and then concorded to ISIC rev. 2 4-digit categories. Herfindahl index of suppliers is sum of squared market shares of producers of product, by 5-digit industry. Herfindahl index of purchasers is sum of squared expenditure shares of purchasers of product, by 5-digit industry. Averages assign equal weight to each plant-product-year in product-level observation data on outputs. See Appendix B.1 for more detailed variable descriptions and Appendix B.2 for details of data processing.

Table 7:
Product-level prices vs. plant size interacted with advertising + R&D intensity, 1982-2005 panel

	dep. var.: log real output price					dep. var.: log real input price				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
log employment	0.030***	0.009	0.003	-0.025**	-0.029**	0.012***	0.003	0.002	0.006	0.005
	(0.007)	(0.009)	(0.011)	(0.012)	(0.013)	(0.003)	(0.005)	(0.005)	(0.008)	(0.008)
log emp.*advertising ratio		1.042***		1.004***			0.374**		0.380**	
		(0.351)		(0.350)			(0.165)		(0.164)	
log emp.*(adv. + R&D) ratio			0.920***		0.876***			0.271**		0.277**
			(0.307)		(0.308)			(0.136)		(0.136)
log emp.*Rauch measure				0.045***	0.043***				-0.004	-0.004
				(0.015)	(0.015)				(0.009)	(0.009)
product-year effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
industry effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
region-year effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
R^2	0.90	0.90	0.90	0.90	0.90	0.79	0.79	0.79	0.79	0.79
N (obs.)	320618	320618	320618	320618	320618	1039673	1039673	1039673	1039673	1039673
N (plants)	11971	11971	11971	11971	11971	10718	10718	10718	10718	10718

Notes: Data on advertising and R&D expenditures as a share of total industry sales from the U.S. Federal Trade Commission (FTC) 1975 Line of Business Survey, converted from FTC 4-digit industry classification to ISIC 4-digit rev. 2 classification using verbal industry descriptions. At SITC 4-digit level, Rauch (1999) measure set to 0 if good is "homogeneous" or "reference-priced" according to the Rauch "liberal" definition, to 1 if reported not to be in either category, and then concorded to ISIC rev. 2 4-digit categories. Columns 1 and 6 correspond to Column 2 of Table 9 but use reduced sample. Errors clustered at plant level. N (plants) reports number of cluster (i.e. distinct plants that appear in any year). Four-digit industry for outputs (Columns 1-5) defined as first four digits from 8-digit product code. Four-digit industry for inputs (Columns 6-10) defined as industry of corresponding plant. Sample includes sectors with complete data on advertising and R&D intensity and Rauch measures. Number of plants differ between Columns 1-5 and 6-10 because of difference in industry definitions. Robust standard errors in brackets. *10% level, **5% level, ***1% level. See Appendix B.1 for more detailed variable descriptions and Appendix B.2 for details of data processing.

Table 8:
Controlling for market power in input markets

	dependent variable: log real input unit value							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log employment	0.012***	0.019***	0.010***	0.009***	0.017***	0.018***	0.010***	0.018***
	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)
log emp.*Herf. suppliers index		-0.014**			-0.018***	-0.018***		-0.018***
		(0.006)			(0.006)	(0.006)		(0.006)
log emp.*Herf. purchasers index			0.017		0.026**		-0.010	-0.001
			(0.011)		(0.011)		(0.011)	(0.011)
purchaser share				0.230***		0.237***	0.238***	0.238***
				(0.037)		(0.036)	(0.037)	(0.037)
product-year effects	Y	Y	Y	Y	Y	Y	Y	Y
industry effects	Y	Y	Y	Y	Y	Y	Y	Y
region-year effects	Y	Y	Y	Y	Y	Y	Y	Y
R^2	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76
N (obs.)	1067789	1067789	1067789	1067789	1067789	1067789	1067789	1067789
N (plants)	13294	13294	13294	13294	13294	13294	13294	13294

Notes: Herfindahl index of suppliers is sum of squared market shares of producers of input, at 5-digit industry level. Herfindahl index of purchasers is sum of squared expenditure shares of purchasers of input, at 5-digit industry level. Purchaser share defined as expenditures on product by plant as a share of total expenditures on product by all plants in a given year. Sample includes plant-product-year observations for which both Herfindahl indices could be constructed; sample is smaller than in baseline model (Table 3) because some plants use only non-manufacturing inputs for which Herfindahl purchaser index could not be constructed. Errors clustered at plant level. N (plants) reports number of clusters (i.e. distinct plants that appear in any year). Robust standard errors in parentheses. *10% level, **5% level, ***1% level. See Appendix B.1 for more detailed variable descriptions and Appendix B.2 for details of data processing.

Table 9:
Plant-average prices vs. plant size, 1982-2005 panel

	OLS (1)	Reduced form (2)	IV (3)
A. Dependent variable: plant-average output price			
log total output	0.010*		0.012**
	(0.005)		(0.006)
log employment		0.013**	
		(0.006)	
industry effects	Y	Y	Y
region-year effects	Y	Y	Y
R^2	0.44	0.44	
N (obs.)	114500	114500	114500
N (plants)	13582	13582	13582
B. Dependent variable: plant-average input price			
log total output	0.017***		0.012***
	(0.002)		(0.003)
log employment		0.013***	
		(0.003)	
industry effects	Y	Y	Y
region-year effects	Y	Y	Y
R^2	0.33	0.33	
N (obs.)	114500	114500	114500
N (plants)	13582	13582	13582

Notes: Output (input) price index defined as coefficient on plant-year effect from product-level regression of log real output (input) unit values on full sets of plant-year and product-year effects. (Refer to equations (11)-(12) in Section 9.1 of text.) Total output is total value of production, defined as sales plus net transfers plus net change in inventories. In Column 3, log employment is instrument for log total output; the coefficient on log employment, its robust standard error and the R^2 in the first stage are 1.067, 0.008 and 0.664, respectively. Errors clustered at plant level. N (plants) reports number of clusters (i.e. distinct plants that appear in any year). Robust standard errors in parentheses. *10% level, **5% level, ***1% level. See Appendix B.1 for more detailed variable descriptions and Appendix B.2 for details of data processing.

Table 10:
Gollop-Monahan Index as measure of horizontal differentiation, 1982-2005 panel

	dep. var.: log real output price			dep. var.: log real input price		
	(1)	(2)	(3)	(4)	(5)	(6)
log employment	0.030***	-0.067***	-0.068***	0.012***	-0.020	-0.019
	(0.007)	(0.022)	(0.022)	(0.003)	(0.014)	(0.014)
log emp.*advertising ratio		0.742**			0.359**	
		(0.376)			(0.164)	
log emp.*(adv. + R&D) ratio			0.637*			0.254*
			(0.329)			(0.135)
log emp.*Gollop-Monahan index		0.147***	0.141***		0.042*	0.041*
		(0.038)	(0.038)		(0.025)	(0.025)
product-year effects	Y	Y	Y	Y	Y	Y
industry effects	Y	Y	Y	Y	Y	Y
region-year effects	Y	Y	Y	Y	Y	Y
R^2	0.90	0.90	0.90	0.79	0.79	0.79
N (obs.)	322044	322044	322044	1039673	1039673	1039673
N (plants)	10718	10718	10718	10718	10718	10718

Notes: Data on advertising and R&D expenditures as a share of total industry sales are from the U.S. Federal Trade Commission (FTC) 1975 Line of Business Survey, converted from FTC 4-digit industry classification to ISIC 4-digit rev. 2 classification using verbal industry descriptions. Gollop-Monahan index is a measure of dissimilarity of input mixes among plants in a 5-digit industry; see Appendix B.1 for a detailed description of the construction of this variable. Sample includes sectors for which advertising and R&D intensity, Rauch and Gollop-Monahan measure could be constructed. Columns 1 and 6 correspond to Column 2 of Table 9 but use reduced sample. Errors clustered at plant level. N (plants) reports number of clusters (i.e. distinct plants that appear in any year). Robust standard errors in brackets. *10% level, **5% level, ***1% level. See Appendix B.1 for further variable descriptions and Appendix B.2 for details of data processing.

Table 11:
Product-level prices vs. plant size, non-exporters only, 1982-1994 panel

	OLS (1)	Reduced form (2)	IV (3)
A. Dependent variable: log real output unit value			
log total output	0.013*		0.020**
	(0.007)		(0.008)
log employment		0.023**	
		(0.009)	
product-year effects	Y	Y	Y
industry effects	Y	Y	Y
region-year effects	Y	Y	Y
R^2	0.91	0.91	
N (obs.)	170261	170261	170261
N (plants)	9352	9352	9352
B. Dependent variable: log real input unit value			
log total output	0.023***		0.017***
	(0.003)		(0.003)
log employment		0.020***	
		(0.004)	
product-year effects	Y	Y	Y
industry effects	Y	Y	Y
region-year effects	Y	Y	Y
R^2	0.81	0.81	
N (obs.)	510011	510011	510011
N (plants)	9352	9352	9352

Notes: Specifications are the same as in Table 3, but only include non-exporting plants (i.e. plants with zero exports). Total output is total value of production, defined as sales plus net transfers plus net change in inventories. In Column 3, log employment is instrument for log total output; the coefficient on log employment, its robust standard error and the R^2 in the first stage are 1.136, 0.010 and 0.777 in Panel A and 1.165, 0.009 and 0.832 in Panel B, respectively. Product-year and industry effects are not perfectly collinear because industry is defined as the industry category with the greatest share of plant sales, and two plants producing the same product may be in different industries. Errors clustered at plant level. N (plants) reports number of clusters (i.e. distinct plants that appear in any year). Robust standard errors in brackets. *10% level, **5% level, ***1% level. See Appendix B.1 for more detailed variable descriptions and Appendix B.2 for details of data processing.

Table 12:
Product-level prices vs. physical quantities, 1982-2005 panel

	OLS (1)	Reduced form (2)	IV (3)
A. Dependent variable: log real output unit value			
log physical quantity	-0.171***		0.032***
	(0.004)		(0.009)
log employment		0.026***	
		(0.007)	
product-year effects	Y	Y	Y
industry effects	Y	Y	Y
region-year effects	Y	Y	Y
R^2	0.91	0.90	
N (obs.)	413789	413789	413789
N (plants)	13582	13582	13582
B. Dependent variable: log real input unit value			
log physical quantity	-0.137***		0.016**
	(0.001)		(0.005)
log employment		0.012***	
		(0.003)	
product-year effects	Y	Y	Y
industry effects	Y	Y	Y
region-year effects	Y	Y	Y
R^2	0.80	0.78	
N (obs.)	1338921	1338921	1338921
N (plants)	13582	13582	13582

Notes: Physical quantity is number of physical units reported. In Column 3, log employment is instrument for log physical quantity; the coefficient on log employment, its robust standard error and the R^2 in the first stage are 0.789, 0.013 and 0.247 in Panel A and 0.744, 0.011 and 0.25 in Panel B, respectively. Product-year and industry effects are not perfectly collinear because industry is defined as the industry category with the greatest share of plant sales, and two plants producing the same product may be in different industries. Errors clustered at plant level. N (plants) reports number of clusters (i.e. distinct plants that appear in any year). Robust standard errors in brackets. *10% level, **5% level, ***1% level. See Appendix B.1 for more detailed variable descriptions and Appendix B.2 for details of data processing.